

附件 3

资本市场金融科技创新试点（上海） 项目公示表

填报时间：2022 年 1 月 19 日

■ **试点公示** （对于通过试点申请的项目，《公示表》将在项目公示阶段对社会公开）

□ **辅导公示** （对于通过辅导申请的项目，《公示表》将在项目公示阶段对社会公开，
标*项目可酌情填写，或填“暂无”、“不适用”）

一、 项目 基本 信息	1.1 申报单位 （以重要性为序 逐行列明单位营 业执照上的全称）	1.1.1 牵头申报单位： 上海期货交易所
		1.1.2 联合申报单位： 国泰君安期货有限公司、海通期货股份有限公司、上海东证 期货有限公司及新潮期货股份有限公司
	1.2 项目名称	基于可信隐私计算技术的期货行业智能风控辅助平台
	1.3 项目类型 （可多选）	☐ 金融服务 ☐ 科技产品 ☐ 业务辅助 ☐ 合规科技 ☒ 监管科技 ☒ 行业平台 ☐ 行业基础设施 ☐ 其他(需补充说明)：
	1.4 应用场景	近年来，人工智能、大数据等新兴技术被广泛应用在商品期货市场，一方面对活跃市场起到了积极推动作用，但同时也给市场监管工作带来新的挑战。由于商品期货业务与大宗商品现货交易存在紧密关联，相关监管工作不仅仅覆盖交易、结算、交割等核心流程，内容还涉及质押、货权转移等关联环节。在此背景下，商品期货行业的风控工作也需在传统风控范围基础上不断向外延伸，导致传统风险识别手段的能力短板日益凸显。例如，传统风控方法的视角广度相对不足，缺乏对监管对象风险外部关联数据，尤其是舆情、工商司法等外部数据的有效分析利用，难以及时精准的识别动态多变的各类风险隐患。因此，如何在保护数据隐私和安全的前提下，破解“数据孤岛”困境，开展跨机构的数据协同，成为提升商品期货市场风控能力亟需解决的问题。 针对上述问题，本项目围绕核心机构和经营机构关心的

		“延伸风险隐患”，例如，行业监管工作密切相关的运营流程风险、产业链风险、会员及客户信用风险等内容，探索利用市场外部相关数据资源融合分析的方法，建设企业级的基于可信隐私计算技术的期货行业智能风控辅助平台，提升风险识别、预警及全视角评估等科技监管能力。 项目一方面通过融合舆情数据、工商数据、司法数据、以及其他可供使用的有效外部数据等，建立适用特定风控场景的人工智能模型，对大宗商品期货市场潜在风险进行多角度、全方位的分析，实现风险发现、风险动态识别、风险传导、风险评估及风险处置的业务闭环，助力期货市场核心机构及经营机构向风控智能化转型。另一方面引入隐私计算及区块链技术，使得各方数据可用不可见，在保障各方数据安全的前提下，打破“数据孤岛”困境。根据期货市场潜在风险的类型，本项目将试点应用场景分为运营流程风险防控、产业链风险防控及信用风险防控，具体如下： 1、运营流程风险防控 大宗商品期现流转包含交割、质押、结算、货权转移等环节，涉及到多家机构，任一环节或机构出现问题，都可能引发运营流程风险，从而影响期货市场的稳定运行。以交割仓库为例，交割仓库的管理水平及资信状况是影响交割环节正常运行的重要因素之一，因此其经营状况的变化极易成为期货市场运营流程风险爆发的诱因。目前，交易所主要通过现场查库、年审等方式对交割仓库的管理水平及资信状况进行监管，缺乏智能化的实时风险把控能力。此外，期货公司可获取的仓库信息较少，更加难以全面认知仓库的运营状况，无法及时进行风险防控。因此，期货行业智能风控辅助平台创新性地结合交易所及期货公司的外部舆情数据、工商数据、司法数据等，利用机器学习、知识图谱等技术，建立风险画像体系，进行智能化分析，进而实现风险实时预警，助力期货市场核心机构及经营机构更加全面地了解仓库经营状况，提升期货市场整体运营流程风险管理水平。 2、产业链风险防控 期货品种产业链风险是影响期货价格的重要因素之一，其潜在风险可能引发期货价格在短期内剧烈波动，从而影响市场平稳运行。期货行业智能风控辅助平台通过融合交易所及期货公司现有的外部舆情、工商、司法等数据，利用人工智能技术对市场形势进行多角度、全方位的分析。同时，引入隐私计算、区块链等技术，在保护数据隐私的同时，打破
--	--	---

	“数据孤岛”困境，弥补现有数据的不足，强化对产业链上下游企业风险状况监控能力。一方面，可以辅助交易所准确把握期货市场产业风险，制定相关交易政策，提升产业客户参与度的同时，也为及时采取相关风险防范和化解措施提供辅助决策的依据。另一方面，有助于经营机构及时向客户提示产业链相关风险，规避因期货价格剧烈波动而造成的财产损失。 3、信用风险防控 会员及客户信用风险防控能力逐渐成为制约期货市场核心机构及经营机构扩展业务的重要因素之一。目前，交易所正持续强化对会员的穿透式监管，但相关数据利用有限，效果有待进一步提高。与此同时，尽管部分期货公司已在客户风险识别领域进行了一定的探索，但受限于自身资源投入，致使其在数据、技术等方面储备不足，难以满足大数据时代下的客户风险识别预警需求。期货行业智能风控辅助平台融合海量多源数据，一方面可借助积累的人工智能技术基础，实现对会员、客户的精准画像，及早发现其中的信用风险。另一方面引入隐私计算、区块链技术，规避各方数据泄露风险的同时，充分整合各类其他社会信用相关数据，在保证数据安全的前提下充分发挥不同主体的数据进行交叉协同验证和融合带来的新价值。
*1.5 数据应用	1、数据来源 项目使用的外部数据主要包括慧科讯业的舆情数据，中证数据的工商司法数据，以及隆众资讯、上海钢联、新华财经、卓创资讯、百川数据、wind等数据公司的行情数据、资讯数据，同时按需补充内/外部合规渠道可获得的其他外部数据。 2、采集方式 考虑到数据来源、数据内容、实现复杂度等角度，本项目通过数据接口调用的方式获取数据。此外，权属明确的数据由权属方负责数据的存储，公开数据存储于独立的项目应用数据库。 3、数据规模 本项目近两年汇集的舆情数据规模约为 100GB，平均日增量约为 100M，工商司法数据规模约为 300GB，平均日增量约为 500MB，金融资讯数据以及各类价格行情数据规模约为 3GB，平均日增量约为 10M，同时按需辅以其他合规渠道可获得的外部数据，相关数据可从数十 GB 至数百 GB 规模不等，

	用于期货市场风险完整、准确、实时的分析，进一步提升期货行业智能风控能力。 4、数据分类及安全级别 本项目参照《证券期货业数据分类分级指引》（JR/T 0158-2018）所述数据归类及级别判定的原则及方法，按照数据级别的重要性和敏感程度，对项目使用的数据进行分类分级，并根据数据的不同类型和级别进行差异化管理。例如，行情数据、外部舆情数据等市场公开数据为1级数据，工商司法数据为2级数据。由于本项目不使用生产数据等内部数据，因此不涉及3级或4级的数据。 5、数据共享和融合应用安排 本项目以保障数据的可用性、完整性、安全性为前提，原始舆情数据、工商司法数据、行情数据、资讯数据等汇聚后，借助特定模型、算法进行处理、分析后对服务对象输出和共享。目前已构建集“盘、归、治、用、智”于一体的完整数据服务体系，应用于期货行业智能舆情、智能监管、智能文本识别等业务场景。
*1.6 实施计划	项目分为五个阶段： 1、第一阶段：夯实智能风控的数据资源基础，提升数据治理水平。 日程安排：2019年1月-2019年12月 主要内容：搭建大数据基础平台，整合所内核心业务数据，所外舆情数据、工商数据、司法数据等，围绕一套平台架构，构建一套管理体系，搭建一个数据生态圈，实现简单的业务使用查询及前端可视化界面。 2、第二阶段：搭建底层大数据基础平台，启动期货行业智能风控辅助平台建设。 日程安排：2020年1月-2020年12月 主要内容：对OCR、NLP、知识图谱技术进行预研，引入并行计算，提升数据处理分析能力，初探人工智能、大数据等新技术在期货行业监管的应用价值。 3、第三阶段，提升复杂数据关键处理水平，启动典型应用场景落地。 日程安排：2021年1月-2021年12月 主要内容：通过自主研发适合业务场景且国内领先的实体识别、OCR、NLP等人工智能、大数据技术，探索非结构化数据的处理和知识图谱技术应用的可行性，并基于前期的研究成果，完成期货行业智能风控辅助平台的落地。

		4、第四阶段：丰富智能风控应用场景，初步完成面向行业的智能风控辅助平台落地。 日程安排：2022 年 1 月-2022 年 12 月 主要内容：挖掘期货公司等经营机构的风险防控需求，分析其在风控领域的业务及技术痛点，针对具体业务场景，不断优化模型，提升模型精度，迭代现有数据及技术，构建期货行业场景化智能风控应用体系，完成面向行业的智能风控辅助平台的落地，进一步加强 AI 赋能期货行业风险研判和防范。 与此同时，研究可信隐私计算技术，初步实现行业风控能力多方共建共享。强化对联邦学习、安全多方计算等隐私计算技术的研究，借助区块链作为“可信引擎”，打破“数据孤岛”困境，避免数据隐私泄露，确保人工智能安全、可靠、可信。 5、未来迭代优化：持续提升平台服务能力，实现行业协同数智赋能。 根据平台试运行情况，持续优化完善风控模型，提升模型精度，强化平台服务能力，维护期货市场稳定发展。同时，基于隐私计算技术及区块链技术，增强 AI 可信性，确保人工智能安全、可靠、可信，实现行业风控能力多方共建共享。此外，进行信创探索，验证该平台在 CPU、操作系统、技术架构等方面的兼容性。
	1.7 面临的困难及解决思路	1、面临的问题 调研结果显示，期货行业的核心机构及经营机构在智能风控领域存在以下的共性问题：一是面临数据资源缺乏的困境，缺乏体系化数据治理手段；二是现有风控指标难以适应日趋复杂的市场环境，缺乏数据驱动的系统化风控指标体系；三是技术储备不足，缺乏风险识别预警分析能力；四是技术融合能力不足，缺乏针对海量复杂数据的有效违规交易识别分析建模方法；五是面临数据孤岛问题，缺乏适配多源数据的高效安全隐私计算方法；六是数据存在泄露隐患，缺乏可信的数据传输通道；七是信创探索不足，缺乏国产化软硬件的替代能力。 此外，相较于交易所等核心机构，期货公司受自身资源约束等因素限制，不同机构之间的技术水平差异较大，在提高智能风控整体技术水平方面的需求更加迫切。 2、解决思路 针对上述面临的困难，本项目从业务、技术及合作形式

	三个方面进行创新，提出相应的解决思路。 业务创新方面：聚焦“数据让监管更加智慧”的愿景，针对特色业务场景，创造性地将智能舆情分析技术与期货市场风险分析相关联，持续优化和探索智能风控手段，借助人工智能、大数据技术，强化对海量舆情数据的实时分析能力，借助隐私计算技术，规避数据隐私泄露风险，借助区块链技术，为数据传输保驾护航。一方面提升期货市场核心机构业务部门对期货市场风险的管控能力，另一方面弥补经营机构在风控方面的数据、技术等资源储备的不足，从而实现期货市场智能风险识别、预警及评估，有效维护期货市场的稳定发展。 技术创新方面：技术创新主要包括数据、平台、方法及信创四个方面。 （1）数据层面 汇集期货市场行情数据、资讯数据、工商数据、司法数据、外部舆情信息等海量数据，打破了数据源单一对应用的限制，充分满足交易所及期货公司的业务需求。此外，积极探索运营商通讯数据、银行数据等其他外部数据在期货行业智能风控的应用前景。 （2）平台层面 在原有平台的基础上，按阶段分期建设人工智能、大数据、隐私计算及区块链组件，最终完成一套集“人工智能”、“大数据”、“隐私计算”、“区块链”四位一体的期货行业智能风控辅助平台，实现海量数据的计算与存储。 （3）方法层面 1) 隐私计算 探索联邦学习、安全多方计算等隐私计算技术在期货行业的应用，将数据可见的具体信息部分和不可见的计算价值部分进行分离，实现“数据可用（可计算）不可见（不可获取）”，进而消除各个数据协同方之间对于数据安全和隐私泄漏的顾虑，从而以技术手段有效的破解“数据孤岛”的困境。 2) 可信传输 利用区块链技术作为可信传输通道，利用其不可篡改的特性，实现数据确权，保障信息传输过程的可信性。 3) 基于 AI 技术的智能化分析 利用自研 OCR 技术对图片、视频数据进行处理，识别关键文本信息；利用 NLP 技术对舆情语义进行分析，挖掘情感、观点等特征，实现智能化风险识别；利用知识图谱技术进行
--	---

		风险传导分析，实现风险预警及评估。 4) 基于大数据技术的海量多源数据计算及存储 利用大数据平台，融合所内外海量多源数据，实现高效稳定的数据计算及存储。 (4) 信创层面 进行信创探索，验证该平台在国产 CPU、操作系统、技术架构等方面的兼容性。 3、合作形式创新 总体上，采用交易所牵头，联合期货公司、行业外部技术/数据服务商共同参与的合作模式，进行智能风控的技术研究及应用探索。根据 MVP 原则，项目合作分阶段推进： (1) 核心机构与经营机构共同打造应用场景，核心机构外部技术服务商构建底层架构，初步实现平台落地； (2) 核心机构与经营机构、外部技术服务商共建应用体系，实现行业风控能力共享共建； 项目合作方式充分考虑各参与方的需求及资源现状，实现技术优势互补，资源共享，服务能力及平台建设迭代推进。既可弥补经营机构在资源储备方面的短板，又可全面发挥核心机构的技术优势，助力期货行业整体提升数字化、智能化水平，进而更好的赋能行业发展，为市场提供更好的服务。
	1.8 专利、认证或奖项	(1) 奖项名称：2020 年度金融科技发展奖 获奖时间：2021 年 12 月 获奖等级：二等奖 颁发单位：中国人民银行 (2) 奖项名称：2019-2020 年度优秀研究成果 获奖时间：2021 年 8 月 获奖等级：二等奖 颁发单位：上海期货交易所
二、依法合规原则评估	*2.1 涉及的业务场景是否由持牌机构提供	2.1.1 申报机构已取得的证券期货相关法定业务资格名称 (本表所称证券期货相关业务指受到中国证监会及其派出机构或相关自律组织认可并进行监管的业务，业务资格取得方式不限于行政审批、备案、登记等)： 不适用
		2.1.2 本次申报项目业务场景涉及的业务资格： 不适用

	2.2 现行法律法规和监管规定符合情况 (对与项目应用场景相关的业务法规和技术规范符合情况进行梳理分析,是否存在违反禁止性规定的情形)	2.2.1 证券监管部门的相关法规及符合情况 (不存在违反禁止性规定的情况,包括但不限于账户实名、资金安全、公平交易、个人信息保护、可控数据跨境流动、反洗钱、网络安全等): 符合证券监管部门的相关法规
		2.2.2 行业协会、交易所等自律组织的相关规范及符合情况 (要求同上): 符合行业协会、交易所等自律组织的相关规范
		2.2.3 国家或其他管理部门的相关法规及符合情况 (要求同上): 符合国家或其他管理部门的相关法规
	*2.3 出具合规评估意见的机构、评估时间及评估结论	2.3.1 评估机构名称 (公司合规部门或第三方专业机构): 上海期货信息技术有限公司质量控制部
		2.3.2 出具时间 (如包含有效期的请注明): 2022 年 1 月 12 日
		2.3.3 评估结论 (最终结论): 基于可信隐私计算的智能风控辅助平台属于市场监管科技类平台项目,与当前有效的法律以及行政法规的强制性规定不冲突,具备合法性和可行性。
三、有序创新原则评估	3.1 技术创新情况	项目采用核心机构牵头,经营机构参与,外部技术/数据服务机构配合的方式联合共建,融合大数据、人工智能、隐私计算及区块链技术,搭建期货行业智能风控辅助平台,具体技术创新情况如下。 1、大数据技术 融合多方现有的舆情数据、工商数据、司法数据、期现价格行情数据、资讯等数据,探索从交易所到期货公司自上到下的数据共享路径。建设期货行业数据体系标准,包括不同类型数据的标准、数据质量、标签体系、指标体系、接口规范等行业共享平台标准规范,集“盘、归、治、用、智”于一体,具体如下: 1) 盘:汇集行情数据、资讯数据、工商数据、司法数据、外部舆情信息等海量多源数据; 2) 归:基于 Hadoop 大数据技术栈,运算性能提升 10 倍,保障对海量数据的存储及处理高效、稳定; 3) 治:利用数据脱敏、权限管控、事后审计等措施对数据进行治理,保障数据的安全性; 4) 用:打造敏捷分析工具,构建自助数据服务生态,

		针对具体业务场景，开展专题研究，建立可靠、高效的科技监管体系； 5) 智：基于大数据及人工智能算法，挖掘业务痛点问题，创新数据产品，满足差异化的客户服务需求。 2、人工智能技术 1) 混合文本识别模型 随着社交媒体的发展，舆情不再局限于以文本的形式传播，图片等非结构化数据同样可能蕴含着重要的信息。非结构化文本数据的自动化字符识别有利于助推期货行业客户服务能力及效率的提升。电子化文档可以根据格式分为已知模板和未知模板的文档。对于已知模板的电子化文档而言，有效文本未知相对固定，传统文本识别方法缺乏对已知信息充分利用的能力，导致识别效率低下；对于未知模板的电子化文档而言，由于文档背景相对复杂，文本位置不固定，需要尽可能保留所有文本信息。 因此，本项目自主研发了一种全新的混合文本识别模型。在文本定位阶段，对于已知模板的电子化文档，首先利用人工标注锚点在模板库中自动匹配模板，并获取该模板下预标注的有效文本区域；若文档中印有红章且覆盖文本区域，则采用基于颜色通道的快速行进算法进行图像修复，再执行文本定位工作，进而获取标准化的文本行区域。对于未知模板的电子化文档，为了尽可能捕获所有文本特征，利用CTPN网络扫描整个文件，将提取的特征集成到双向LSTM中以捕获长距离依赖关系，并通过全连接网络定位框获得最终所有文本结果。在文本识别阶段，全部采用由CNN，双向LSTM和CTC-loss组成的CRNN模型，实现变长文本识别并输出最终的文本识别结果。混合文本识别模型的示意图如下所示。
--	--	--

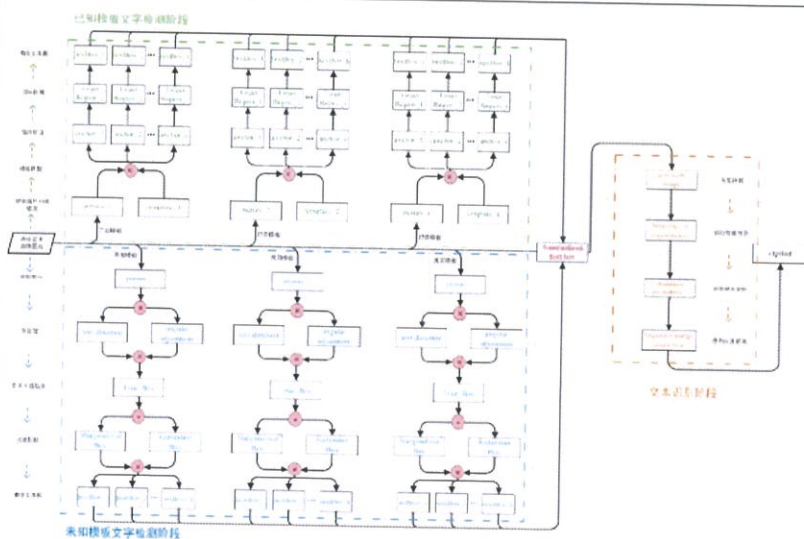


图 1 混合文本识别模型示意图

2) 复杂场景单证关键信息识别模型

在票据电子化归档的趋势下，代替人工的自动化关键文本识别有助于期货行业各机构提升服务效能。由于部分票据在电子化归档的过程中图像质量不高，现有的文本识别方法仍存在一些缺陷。一方面，票据在电子化的过程中，可能受到光照、倾斜、缩放等因素的影响，导致电子票据关键信息失真，增加了文本识别的难度。另一方面，电子票据大多存在固定模板，而现有的方法鲜有对先验知识充分利用，从而增加了计算成本。

鉴于以上事实，本项目自主研发了一种基于先验知识的复杂场景电子票据关键信息识别方法。在训练阶段，该方法通过使用随机噪声、颜色抖动、横线、旋转等图像增强技术来模拟复杂场景，以提高文本识别准确率，增强文本定位模型的鲁棒性。在建模阶段，所提出的方法充分利用先验知识来获取关键信息切片，以提升电子票据关键信息的识别效率和识别精度。最后，在实际的应用中，验证了该模型的有效性。

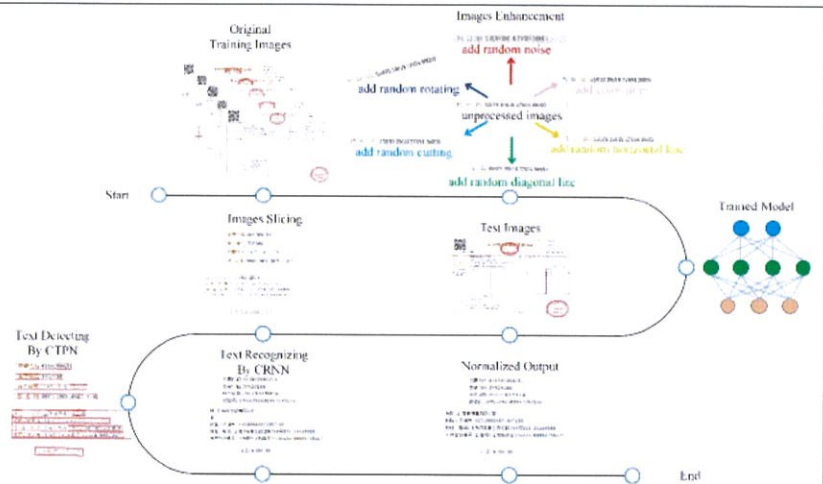


图 2 复杂场景单证关键信息识别模型

3) 基于辅助信息的虚假新闻识别模型

财经新闻是投资者进行研究和投资决策不可缺少的来源。然而，也有许多虚假的财经新闻涌入人们的日常生活。这类信息可能影响舆论，为一些犯罪分子操纵金融市场提供机会。随着直播，小视频等行业的兴起，除了传统的文本或者图片的形式外，视频、音频等多媒体中也包含海量的重要信息。此外，社交媒体中的评论、弹幕等信息同样可以为虚假新闻识别提供支撑。面对复杂、多元的新媒体新闻，传统的虚假新闻识别方法略显单一，局限性大。此外，现有的虚假新闻模型大多采用暴力挖掘的特征，对于部分难以察觉的特征无能为力。

因此，本项目自主研发了一种基于多元化数据特征决策支持的虚假新闻识别，针对不同的新闻载体采用不同的方式。首先，利用 OCR、语音文字识别等技术获取多元新闻数据的文本信息，同时结合用户评论、新闻来源及市场数据，深度挖掘多元化文本数据中的观点特征、情感特征、用户特征及传播特征，以实现多元数据下的虚假新闻检测。在构建特征时，选用提取抽象特征能力更强的神经网络模型。模型整体分为嵌入层、编码层、交互层、解码层以及输出层。嵌入层负责将文本输入转化为机器可以处理的稠密向量和字符级别的信息，编码层使用深度神经网络提取高维度特征，交互层使用注意力机制帮助模型学习各输入之间的关系，解码层使用深度神经网络将特征解码成输出字符的概率分布，输出层则直接输出对应的预测结果。多元化数据特征决策支持的虚假新闻识别的特征提取示意图如下图所示。



在程序化交易的背景下，虚假新闻的传播可能扰乱金融市场秩序。传播渠道多样化、数据类型多元化的新媒体新闻给传统的虚假新闻检测方法带来一定的挑战。一方面虚假新闻的分类特征更加复杂，这提升了虚假新闻检测的难度。另一方面，一些辅助信息，例如评论者的观点，可以为识别虚假新闻提供有力支撑，而这部分信息在现有工作中并没有被充分利用。

鉴于上述事实，本项目自主研发了一种基于多元数据特征挖掘的多层虚假新闻识别框架。底层采集不同渠道的多元化新闻数据并统一编码转化为文本形式，同时对新闻评论的主题相关性和情感强度进行分析，抽取其中的有效评论。中间层负责挖掘文本中的观点特征、情感特征、用户特征及传播特征。具体而言，观点特征包括新闻间及新闻与评论间观点的相互支持或对立特征。情感特征包括情感类别、情感强度及情感表达三个方面。用户特征包括基础特征和统计特征。传播特征包括转发量、评论量、可靠用户的转发量、评论量等。顶层基于主流的 BERT 模型，获得虚假新闻的识别结果。

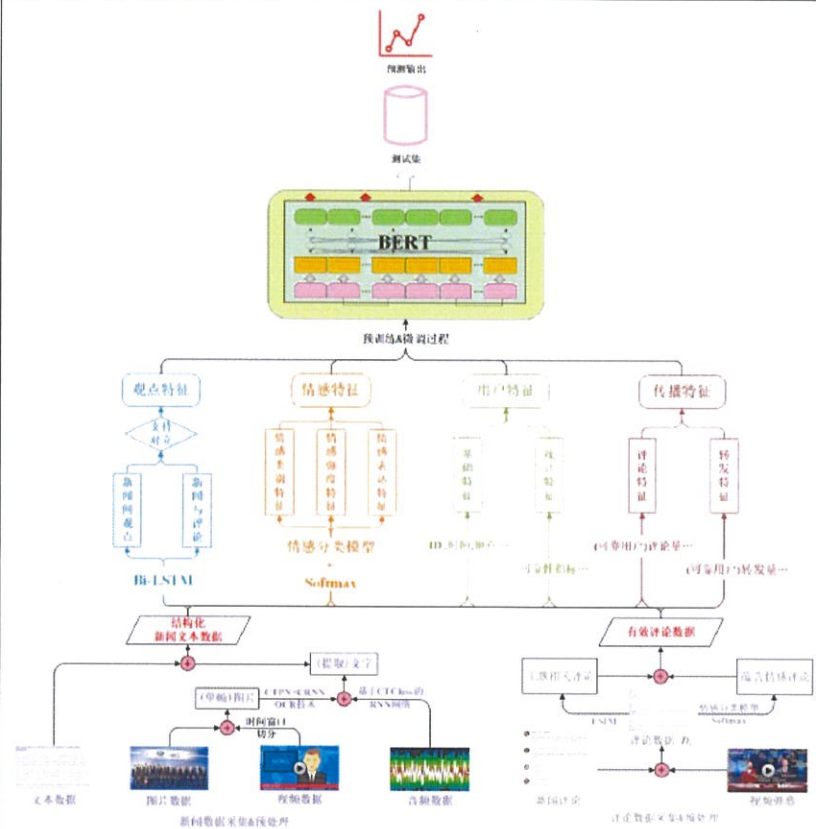


图 4 基于多元化数据特征决策支持的虚假新闻识别模型

5) 实体抽取模型

实体抽取有助于舆情的深度分析。目前主流的实体抽取模型主要包括 Bi-LSTM+CRF、Lattice-LSTM+CRF、Join Model 和 Matching the Blanks (MTB)。Bi-LSTM+CRF 是一种经典的运用深度学习完成序列标注任务的模型，在英文语料中取得了较好的效果。Lattice-LSTM+CRF 模型增加了对字粒度、词粒度的考虑，相对于只使用字粒度的模型，增加了词信息，丰富了语义表达，相对于仅使用词粒度的模型，可以避免分词错误带来的影响。Join Model 基于联合抽取的思想，通过一个模型实现端到端的实体识别和关系确认。尽管三种模型在实体抽取中取得了一定的成就，但是它们都需要预定义实体关系类型，再识别实体之间是否存在该种关系，因此无法挖掘新的实体关系类型。Matching the Blanks (MTB) 设计一个通用的关系抽取器，提高了泛化能力，无需事先定义好关系，可以对任意关系进行抽取建模。

事实上一些实体关系类型显而易见，MTB 模型缺少对先验知识的充分利用。因此，本项目自主研发了一种基于混合模型的多级实体关系抽取策略，首先利用 Lattice-LSTM+CRF 和 Join Model 对可以预定义、有规则的实体关系进行抽取，

同时利用 MTB 模型挖掘新关系以进行补充,实现对实体关系类型的初步抽取,随后将抽取的实体关系类型输入到 Bert 模型中,以实现实体关系类型的深度挖掘。混合模型的多级实体关系抽取策略的实现思路如下图所示。



图 5 实体抽取模型示意图

6) 事理逻辑判断趋势模型

传统的针对金融行业趋势判断的模型,往往都是通过关键词+规则进行判断,针对各实体之间的关系,通常采用以本体概念为基础的知识图谱形式,这种形式存在一些局限性,例如实体语义不完善,容易断章取义,不能有效地传递出各实体之间的逻辑关系等。

因此,本项目自主研发了一种基于专家知识推理的事理逻辑判断趋势模型。该模型首先利用基于 TF-IDF 和 TextRank 的模型寻找原始文本关键词语实体,利用 MMR 和权重过滤相似实体并连接合并得到重要目标实体短语。其次,利用基本句法关系寻找句子之间的六种逻辑要素,包括:因果、转折、条件、顺承、并列、反因果。接着,利用基于哈工大的 PYLTP 模型,对整理的六种逻辑要素句子对进行语义句法分析,寻找重要目标实体的前后谓词性事件并连接得到谓词性事件实体。最后,利用情感分析或者专家知识库规则匹配得到谓词性事件重要实体对品种的趋势影响,并根据不同场景需要,选择是否形成图形式的关系网络图。

自主研发的基于专家知识推理的事理逻辑判断趋势模型能够解决谓词性事件实体之间、谓词性事件实体之间状态的逻辑关系,保证了重要目标实体语义完整性和谓词性事件实体状态之间的清晰性,从而更准确地判断舆情类新闻对品种实体的影响趋势。基于专家知识推理的事理逻辑判断趋势模型的示意图如下所示。

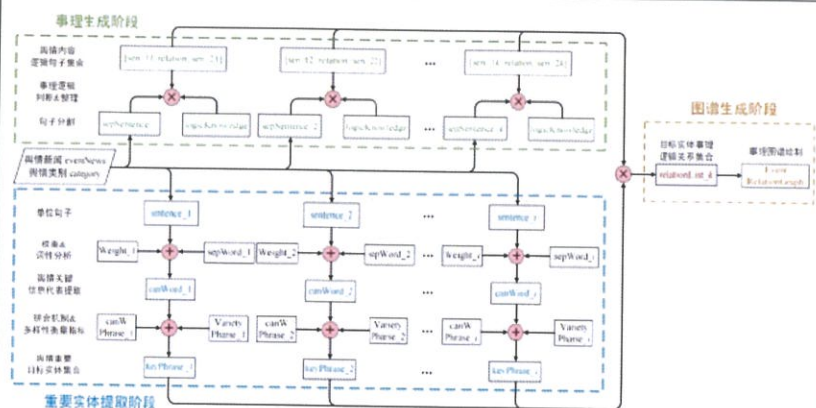


图 6 事理逻辑判断趋势模型示意图

7) 期货行业知识图谱技术

目前，对大宗商品上下游企业的宏观研究略显不足，无法对宏观金融风险做到提前感知、处置。知识图谱技术是解决这一痛点问题的重要手段，因此期货行业的核心机构及经营机构纷纷对此技术投入了大量的研发力量。在对十数家厂商的调研及 POC 后，项目组发现现有的知识图谱产品大多应用于客户营销、智能搜索、智慧医疗、互联网金融等领域，鲜有针对期货行业数据特性而定制的知识图谱产品。

因此，本项目利用期货行业核心机构及经营机构在期货行业知识方面的优势，结合知识图谱技术，率先构建期货行业知识图谱，其架构图如图 7 所示。首先通过专家协助，及人工智能等学习算法，建立包括实体库、金融词库、舆情标签、事件库的期货行业知识库，为后续知识图谱构建提供支撑。接着，将不同来源的数据进行数据抽取，转化成图谱构建所需的实体、关系和属性。其中，数据抽取主要包括交易分析场景（结构化数据）的数据抽取、事理/事件分析场景（非结构化数据）的数据抽取等。随后，通过构建自动融合策略，辅以人工校对，实现知识融合。最后，基于逻辑推理、基本关系集、专题关系集、常识计算、专家系统、图挖掘、概率模型等方法实现知识推理，主要包括：对公司实际控制人/一致行动人的推理、基于概念泛化的推理、投资/竞争关系推理等。

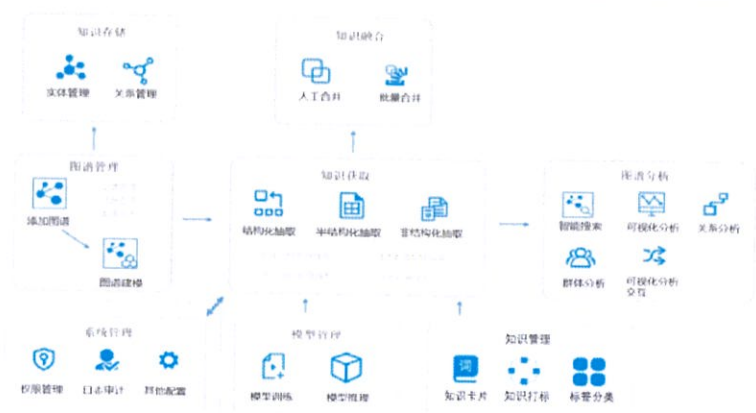


图 7 知识图谱架构图

3、隐私计算技术

为进一步扩展期货市场风控视角，项目组初步探索隐私计算技术在期货行业的应用价值，在保障各方数据安全的前提下，加强跨机构的数据协同。

1) 联邦学习

联邦机器学习(Federated machine learning/Federated Learning)，又名联邦学习，联合学习，联盟学习。联邦机器学习是一个机器学习框架，能有效帮助多个机构在满足用户隐私保护、数据安全和政府法规的要求下，进行数据使用和机器学习建模。联邦学习作为分布式的机器学习范式，可以有效解决数据孤岛问题，让参与方在不共享数据的基础上联合建模，能从技术上打破数据孤岛，实现AI 协作。根据参与各方数据源分布的情况不同，联邦学习可以被分为三类：横向联邦学习、纵向联邦学习、联邦迁移学习。

横向联邦学习：在两个数据集的用户特征重叠较多而用户重叠较少的情况下，通过把数据集按照横向(即用户维度)切分，并取出双方用户特征相同而用户不完全相同的那部分数据进行训练。这种方法叫做横向联邦学习。

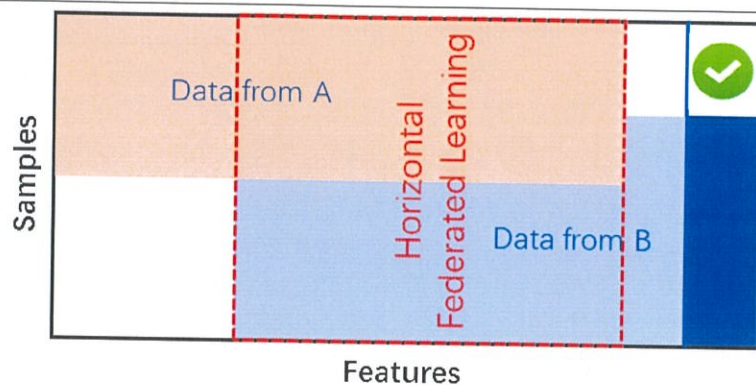


图 8 横向联邦学习

纵向联邦学习：在两个数据集的用户重叠较多而用户特征重叠较少的情况下，通过把数据集按照纵向（即特征维度）切分，并取出双方用户相同而用户特征不完全相同的那部分数据进行训练。这种方法叫做纵向联邦学习。

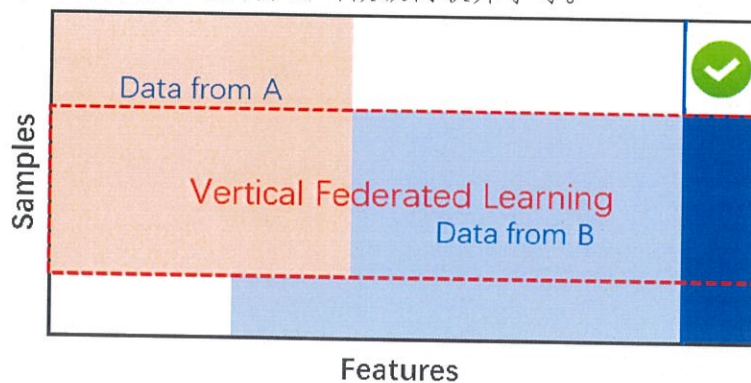


图 9 纵向联邦学习

迁移联邦学习：在两个数据集的用户与用户特征重叠都较少的情况下，不对数据进行切分，而可以利用迁移学习来克服数据或标签不足的情况。这种方法叫做联邦迁移学习。

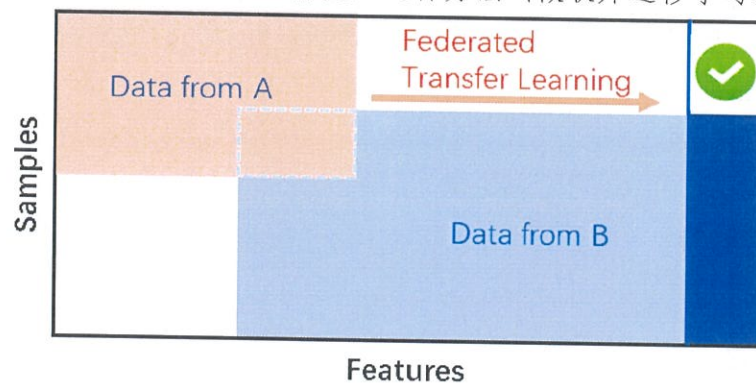


图 10 迁移联邦学习

2) 安全多方计算

安全多方计算 (Secure Multi-party Computation, 简

称 MPC，主要研究参与者间协同计算及隐私信息保护问题，其特点包括输入隐私性、计算正确性及去中心化等特性。技术架构如下图所示。

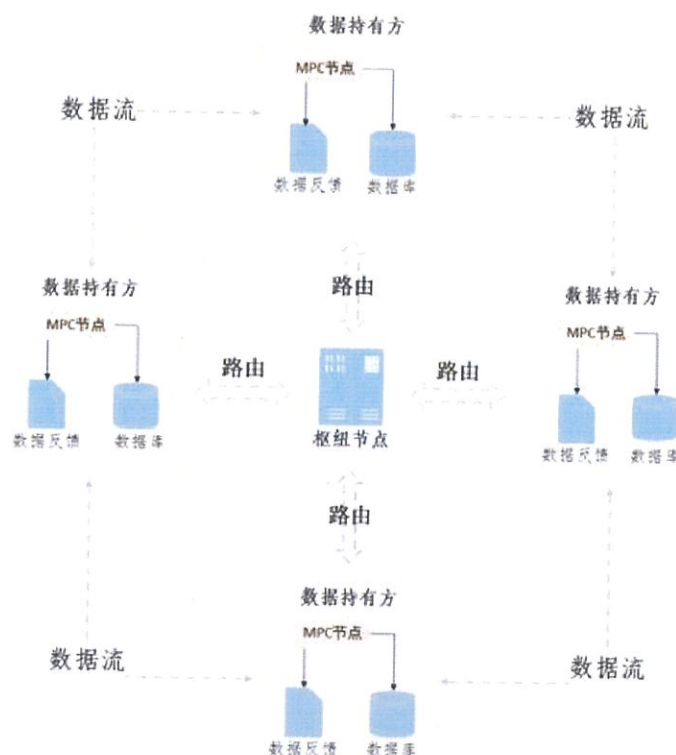


图 11 安全多方计算

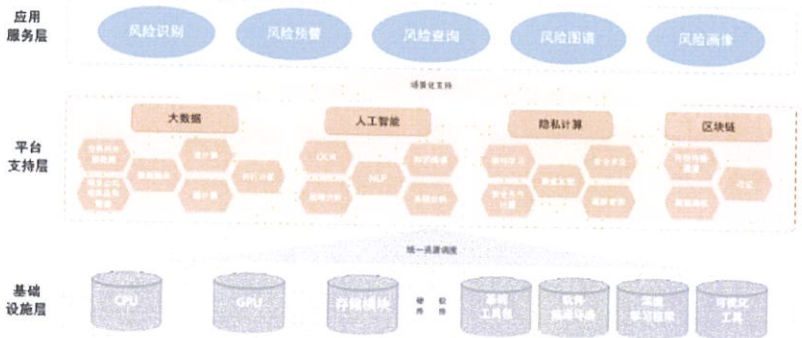
4、区块链技术

区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式，具有去中心化、开放性、自治性、信息不可篡改、匿名性等特点，可为跨机构的信息传递提供可信传输通道，助力数据确权。

去中心化：由于使用分布式核算和存储，不存在中心化的硬件或管理机构，任意节点的权利和义务都是均等的，系统中的数据块由整个系统中具有维护功能的节点来共同维护。

开放性：系统是开放的，除了交易各方的私有信息被加密外，区块链的数据对所有人公开，任何人都可以通过公开的接口查询区块链数据和开发相关应用，因此整个系统信息高度透明。

自治性：区块链采用基于协商一致的规范和协议（比如一套公开透明的算法）使得整个系统中的所有节点能够在去信任的环境自由安全的交换数据，使得对“人”的信任改成了对机器的信任，任何人为的干预不起作用。

		信息不可篡改：一旦信息经过验证并添加至区块链，就会永久的存储起来，除非能够同时控制住系统中超过 51% 的节点，否则单个节点上对数据库的修改是无效的，因此区块链的数据稳定性和可靠性极高。 匿名性：由于节点之间的交换遵循固定的算法，其数据交互是无需信任的（区块链中的程序规则会自行判断活动是否有效），因此交易对手无须通过公开身份的方式让对方自己产生信任，对信用的累积非常有帮助。 5、结合原理 大数据技术实现了对多源复杂数据的采集、计算及存储，而 AI 技术的成长离不开基于数据的深度学习，区块链技术则是数据真实性、可靠性的重要保障，隐私计算技术可打破数据孤岛壁垒。四种技术的有效融合有助于提升期货行业的整体风险防控能力，推动期货行业风险防控迈向数字化、智能化，具有积极的理论价值和实践意义。平台架构图如下所示：  图 12 期货行业智能风控辅助平台架构图 目前，国外金融行业关于大数据的应用主要基于数据仓库等基础设施建设，实现了数据的统一归集管理，在大数据赋能客户营销、资本运营、数据服务等业务上取得了一定的进展。然而，随着业务场景越来越复杂，原有基础设施愈发难以满足业务需求，一方面原有数据源通常来自企业内部产生的数据，来源相对单一，随着业务场景的变化，又是逐渐不足以支撑业务部门的需求；另一方面，目前金融市场的大数据平台大多服务于券商、商业银行等机构，目的大多集中于客户营销、市场分析等业务场景，鲜有站在风险防控角度的大数据创新应用。 国内金融行业的数字化转型也已全面开展，大数据、人工智能等技术在其中已得到了广泛应用，不仅促进了企业金融创新、帮助其控制风险、助力金融监管，而且也为金融行
3.2 技术领先优势		

		业的发展提供了新的技术支撑。但在具体应用实施过程中普遍存在下述困难：一方面由于数据呈现爆发式增长，原有系统架构难以支撑海量数据吞吐、多元异构访问等新挑战，数据处理能力有待进一步提升；另一方面，由于期货行业监管业务场景独特且复杂，如何利用人工智能赋能风险防控也是一个亟需解决的问题。 本项目技术应用属于对同业做法有显著改进。主要体现在技术合作层面、数据层面、平台层面、方法层面。 技术合作层面：项目采用核心机构牵头，经营机构参与，外部技术/数据服务机构配合的方式联合共建，共同开展期货行业智能风控的技术研究及应用探索，一方面即可弥补经营机构在技术储备方面的短板，另一方面又可全面发挥核心机构的技术优势，助力期货行业整体提升数字化、智能化水平，进而更好的赋能行业发展，为市场提供更好的服务。 数据层面：汇集行情数据、舆情数据、工商司法数据等，对结构化数据、半结构化数据、非结构化数据进行联合建模，多角度、全方位进行期货市场风险预警、识别及防控。 平台层面：在原有平台的基础上，按阶段分期建设集人工智能、大数据、隐私计算、区块链四位一体的期货行业智能风控辅助平台，融合几项关键技术，使其互通互补，相得益彰。 方法层面：利用大数据技术，融合所内外海量多源数据，实现高效稳定的数据计算及存储，为 AI 建模提供算力支持。利用自研 OCR 技术对图片、视频数据进行处理，识别关键文本信息；利用 NLP 技术对舆情语义进行分析，挖掘情感、观点等特征，实现智能化风险识别；利用知识图谱技术进行风险传导分析，实现风险预警及评估。初步尝试利用联邦学习、安全多方计算等隐私计算技术将数据可见的具体信息部分和不可见的计算价值部分进行分离，实现“数据可用（可计算）不可见（不可获取）”，进而消除各个数据协同方之间对于数据安全和隐私泄漏的顾虑，从而以技术手段有效的破解“数据孤岛”的困境。同时，探索区块链技术作为可信传输通道在期货行业应用的可行性，利用其不可篡改的特性，实现数据确权，保障联邦学习传输过程的可信性。
	3.3 服务对象与渠道	试点期间，项目的预期服务对象为交易所及期货公司，后续将在监管部门的指导下，扩展服务对象，全面提升期货行业整体的风控数字化、智能化能力。

四、风险 可控原 则评估	4.1 业务风险防 控	4.1.1 业务风险点(应结合试点项目特点,描述试点项目上线后可能面临的业务风险,包括但不限于市场风险、信用风险、流动性风险、操作风险、合规风险、舆情风险等): (1) 模型风险:项目可能存在由于基础数据质量不佳、模型设计不合理、业务规则调整等原因,导致评估结果精确度、时效性不足。 (2) 舆情风险:项目可能在以上风险的基础上进一步衍生舆情风险。
		4.1.2 风险监测机制(应描述如何采取措施及时发现和准确评估上述业务风险,针对各类风险分别列举): 项目按照“早发现、早报告、早处置”的原则,加强对模型风险、舆情风险等相关信息的收集、分析、判断与监测。 (1) 模型风险管理方面:建立多个模型进行相互验证,并在模型正式上线前对模型运行情况进行充分验证,同时利用外部数据丰富用于建模的特征,不断提升模型的精度。 (2) 舆情风险管理方面:建立完善的舆情风险管理机制,切实降低项目舆情风险。
		4.1.3 风险控制措施(应描述如何采取措施防控上述业务风险,针对各类风险分别列举): 强化业务风险意识,从源头阻断风险事件的发生。 (1) 模型风险控制方面:建立多个模型进行相互验证,并在模型正式上线前对模型运行情况进行充分验证,同时利用外部数据丰富用于建模的特征,持续优化迭代模型。 (2) 舆情风险控制方面:持续完善舆情风险管理机制,尽早发现项目舆情风险,并及时分析、预警、评估,力争把舆情风险管理控制在萌芽状态。
	4.2 技术风险防 控	4.1.4 应急预案(应描述如若上述业务风险发生将如何采取有效措施尽可能降低或消除负面影响): 项目成立应急响应小组,组织开展应急预案的制定、演练、评估与改进工作,明确当出现业务风险时的联系人、处理流程、响应机制,确保对项目业务风险迅速反应和紧急处理,尽可能消除负面影响。此外,加强应急培训与演练,提高有关人员防范各类安全生产事故的意识与应急能力。 4.3.1 技术风险点(应结合试点项目特点,描述试点项目可能存在的技术风险,包括但不限于网络安全风险、数据安全风险等): (1) 项目可能存在数据泄露或被篡改的数据安全风险。 (2) 项目上线运行后,可能面临网络攻击、业务连续

		性中断等网络安全风险。
		4.3.2 风险监测机制(应描述如何采取措施及时发现和准确评估上述技术风险,针对各类风险分别列举): 项目制定完善的数据安全、网络安全突发事件监测、识别、预警、评估机制,按照“早发现、早报告、早处置”的原则,加强对数据安全风险、网络安全风险等相关信息的收集、分析、判断与监测。此外,根据项目安全需求,明确日常巡检内容和标准,制定巡检手册,开发巡检工具,并由专员对巡检结果进行分析及事后处理,以便及时发现和评估数据安全风险、网络安全风险等。
		4.3.3 风险控制措施(应描述如何采取措施防控上述业务风险,针对各类风险分别列举): (1) 数据风险控制措施: 1) 项目按照证监会信息系统备份能力等级有关等级要求建设备份系统,制定数据备份方案,定期执行数据恢复测试,验证数据的可恢复性,确保可以快速有效地完成数据恢复任务。 2) 利用隐私计算和区块链技术保障数据安全,为信息传输提供可信通道,避免隐私泄露。 (2) 网络风险防控措施:制定完善的网络风险安全防控机制,定期巡检,加强监控,强化工作人员网络风险防控意识。
		4.3.4 应急预案(应描述如若上述技术风险发生将如何采取有效措施尽可能降低或消除负面影响): 项目成立应急响应小组,组织开展应急预案的制定、演练、评估与改进工作,明确当出现技术风险时的联系人、处理流程、响应机制,确保对项目技术风险迅速反应和紧急处理,尽可能消除负面影响。此外,加强应急培训与演练,提高有关人员防范各类安全生产事故的意识与应急能力。
	*4.3 投资者保护机制	4.3.1 客户投诉渠道(接受客户投诉的渠道信息,包括但不限于营业网点地址、通讯地址、电话、传真、电子邮箱、官方网站等): 本平台为行业平台、监管科技类系统,非面向投资者系统,不涉及。
		4.3.2 投诉处理机制(客户投诉受理与处理机制相关内容,包括但不限于受理部门、受理时间、处理流程、处理时限等信息):

		本平台为行业平台、监管科技类系统，非面向投资者系统，不涉及。
		4.3.3 风险补偿机制(应描述申报单位就本试点项目建立的风险补偿和赔付机制，确保试点项目出现意外风险时能够及时对投资者损失进行合理补偿，降低试点项目的负面影响。对于多个单位联合申报的试点项目，应明确风险补偿责任主体)： 本平台为行业平台、监管科技类系统，非面向投资者系统，不涉及客户交易业务及客户损失，不需要风险补偿。
		4.3.4 项目退出机制(应描述试点项目因发生特殊情况需终止或下线时的工作安排。项目退出应平稳有序，确保投资者资金和数据安全，最大程度减少对市场的负面影响。退出机制包括但不限于退出触发条件、业务退出安排、技术退出安排等内容)： 当项目因不可抗力因素导致无法继续，各参与方应遵循以下原则推出项目： 1、冻结当前项目成果（包括但不限于项目相关文档、程序源代码、采集到的原始数据和经平台加工的数据），由牵头单位组织各参与方协商项目成果处置方案。在达成一致之前，牵头单位应确保项目成果的安全，防止被泄露或误用。 2、各方如有信息系统对接项目平台的，应及时断开相应接口，对接系统的恢复由各方自行负责。 3、提供原始数据的参与方对项目平台里存储的原始数据有处置权，提供技术支撑单位有义务配合各参与方完成相关数据处置工作。

附页：

牵头申报单位 承诺	本单位郑重承诺： 1. 本单位在申报资本市场金融科技创新试点（上海）项目过程中，所提供的一切申报材料信息真实、准确和完整。 2. 申报项目符合依法合规、有序创新、风险可控的申报原则。 3. 申报项目不存在违反法律和行政法规情况，不包含国家秘密信息。 4. 本单位将配合监管部门完成后续评审公示、监督检查或风险处置等工作。 5. 本单位已全面开展合规性评估和内控审计，能够有效保障业务连续性和用户信息安全，保证资金安全。 以上承诺如有违反，愿承担相应责任与后果。 单位（公章）  法定代表人（签字）：  2022 年 1 月 19 日
----------------------	--

联合申报单位
承诺

本单位郑重承诺：

1. 本单位在申报资本市场金融科技创新试点（上海）项目过程中，所提供的一切申报材料信息真实、准确和完整。
2. 申报项目符合依法合规、有序创新、风险可控的申报原则。
3. 申报项目不存在违反法律和行政法规情况，不包含国家秘密信息。
4. 本单位将配合监管部门完成后续评审公示、监督检查或风险处置等工作。
5. 本单位已全面开展合规性评估和内控审计，能够有效保障业务连续性和用户信息安全，保证资金安全。

以上承诺如有违反，愿承担相应责任与后果。

单位（公章）



法定代表人（签字）：



2022 年 1 月 19 日

联合申报单位
承诺

本单位郑重承诺：

1. 本单位在申报资本市场金融科技创新试点（上海）项目过程中，所提供的一切申报材料信息真实、准确和完整。
2. 申报项目符合依法合规、有序创新、风险可控的申报原则。
3. 申报项目不存在违反法律和行政法规情况，不包含国家秘密信息。
4. 本单位将配合监管部门完成后续评审公示、监督检查或风险处置等工作。
5. 本单位已全面开展合规性评估和内控审计，能够有效保障业务连续性和用户信息安全，保证资金安全。

以上承诺如有违反，愿承担相应责任与后果。

单位（公章）

法定代表人（签字）：

吴红松

2022 年 1 月 18 日

联合申报单位
承诺

本单位郑重承诺：

1. 本单位在申报资本市场金融科技创新试点（上海）项目过程中，所提供的一切申报材料信息真实、准确和完整。
2. 申报项目符合依法合规、有序创新、风险可控的申报原则。
3. 申报项目不存在违反法律和行政法规情况，不包含国家秘密信息。
4. 本单位将配合监管部门完成后续评审公示、监督检查或风险处置等工作。
5. 本单位已全面开展合规性评估和内控审计，能够有效保障业务连续性和用户信息安全，保证资金安全。

以上承诺如有违反，愿承担相应责任与后果。

单位（公章）

法定代表人（签字）：

2022 年 1 月 17 日

联合申报单位
承诺

本单位郑重承诺：

1. 本单位在申报资本市场金融科技创新试点（上海）项目过程中，所提供的一切申报材料信息真实、准确和完整。
2. 申报项目符合依法合规、有序创新、风险可控的申报原则。
3. 申报项目不存在违反法律和行政法规情况，不包含国家秘密信息。
4. 本单位将配合监管部门完成后续评审公示、监督检查或风险处置等工作。
5. 本单位已全面开展合规性评估和内控审计，能够有效保障业务连续性和用户信息安全，保证资金安全。

以上承诺如有违反，愿承担相应责任与后果。

单位（公章）

法定代表人（签字）



2022 年 1 月 19 日