

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01月18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：中国金融期货交易所股份有限公司
		1.1.2 联合申报单位：
	1.2 项目名称	基于知识图谱的关联账户分析系统
	1.3 项目类型 (可多选)	<input type="checkbox"/> 金融服务 <input type="checkbox"/> 科技产品 <input checked="" type="checkbox"/> 业务辅助 <input type="checkbox"/> 合规科技 <input checked="" type="checkbox"/> 监管科技 <input type="checkbox"/> 行业平台 <input type="checkbox"/> 行业基础设施 <input type="checkbox"/> 其他(需补充说明)：_____
	1.4 应用场景	<p>(试点项目应用业务领域、主要功能、提供的服务、解决的问题等。)</p> <p>在证监会印发《证券期货业科技发展“十四五”规划》中，明确要求持续探索人工智能、机器学习、认知推理等前沿科技手段的运用，发挥海量、多元、异构大数据的互联互通价值，不断丰富科技监管和金融科技创新应用。而在期货衍生品交易过程中，部分客户为规避监管，采用了分仓分时段交易、策略混用和多账户协同等情况，给市场监管带来了困难和挑战。为更加准确有效地识别疑似关联客户群体，中金所应用大数据、知识图谱和人工智能等技术，自主研发了关联账户分析系统，对账户关联情况进行有效的识别，提高了监管效率和智能化水平。</p> <p>项目以大数据为基础，广泛采集获取了各类可以应用于关联分析的数据，包括内部客户资料数据、监管共享数据、工商数据、穿透式监管数据、私募产品数据、舆情数据等。</p> <p>在设计实施过程方面，项目对各类数据中潜在关联分析点进行了深度加工和价值挖掘，形成了自然人、企业、产品的实体信息、投资线索、终端线索、上市公司股东、亲属关系等高质量的数据模型。</p>

		<p>在算法应用方面，项目利用路径搜索、连通区域、社区发现的知识图谱技术用于关联账户的分析和识别；使用机器学习算法判定交易行为的相似；应用自然语言处理技术，对舆情数据进行挖掘。实现了共 21 类关联线索的自动发现和疑似关联组的自动识别，自动关联组与已排查关联组覆盖率超过 93%。</p> <p>提供易用、好用的功能是系统设计的初衷，项目利用了图谱的可视化技术提升关联分析的交互性和操作性；并通过与 BI 分析平台、订单分析平台等业务分析系统无缝衔接；实现图谱技术与关联分析操作流程的融合，大幅提升了监管效能。</p>
	*1.5 数据应用	<p>（试点项目使用的数据来源，应区分内/外部数据，区分公开/私有数据，明确数据主体、采集方式、数据规模、数据分类、安全级别、数据共享和融合应用安排等。）</p> <p>1. 数据来源</p> <p>本项目以交易所的大数据平台为基础，充分利用了交易所内部的客户资料数据、监管数据、工商数据、IP 数据和私募产品数据，以及外部采购和采集的数据等。</p> <p>从数据来源看，可以分以下三类：</p> <p>(1) 内部数据：客户信息、委托、成交、持仓、行情数据等。</p> <p>(2) 监管共享数据：企业/个人工商信息、穿透式监管数据、证券投资者信息、金融产品出资信息等。</p> <p>(3) 外部数据：高精度 IP、股东资料、私募产品、舆情数据等。</p> <p>2. 数据规模</p> <p>基于以上数据，项目在图谱中共构建实体 24 类，共计 1.3 亿；形成构建关联关系的边 31 类，共计 1.5 亿；形成关联线索 21 类，共计 1800 万条，形成客户的关联关系 30 万条。</p> <p>3. 数据安全</p> <p>关联账户分析系统的数据属于敏感数据。该系统只能在交易所专网环境内使用，使用人员必须在受控环境使用专用访问终端，确保数据安全。</p> <p>4. 数据共享</p> <p>按照证监会发布的《中国证监会数据采集管理细则（试行）》的规定，进行监管共享数据的申请，相关数据在受控环境下进行存储和使用。</p>
	*1.6 实施计划	<p>（项目研发、测试、上线等各主要阶段时间节点及进度安排。试点申报项目应已完成研发、测试等主要工作，已经在生产环境实际运行或具备在被允许试点之日起一年内上线运行条件。针对分期建设开发的项目，应注明各期或版本的主要</p>

		<p>内容和日程安排，远期目标可作为单独项目后续另行申报。)</p> <p>本项目将分五阶段建设实施：</p> <ol style="list-style-type: none"> 1. 需求分析阶段（2020年3月至2020年7月） 探查可用于关联分析的数据范围，总结关联账户分析的业务实操经验，形成关联规则；从关联账户分析的业务场景出发，设计功能需求，形成关联分析业务操作的闭环。 2. 系统设计阶段（2020年8月至2020年12月） 根据基础数据的特点和关联分析的需求设计图谱模型；设计关联线索、关联组识别的算法和技术方案；设计交易行为相似的算法和技术方案；设计功能界面和交互逻辑；设计关联账户分析系统与交易所现有大数据平台及BI系统的整合方案。 3. 编码开发阶段（2021年1月至2021年9月） 完成代码开发、功能测试、算法优化。 4. 上线试运行阶段（2021年9月至2021年11月） 完成生产环境部署和试运行。 5. 持续优化提升阶段（2021年11月至2022年10月） 根据业务经验核实关联账户自动识别的结果，优化关联规则和识别算法，提高关联识别的精准度。
	<p>1.7 面临的困难及解决思路</p>	<p>（试点项目研发过程中可能或已经面临的各类困难，包括但不限于技术、业务、人力、资金、合规、风控等方面，以及后续解决的方向和思路。）</p> <ol style="list-style-type: none"> 1. 关联关系强度的量化和智能化 难点：在传统的关联分析过程中，对关联关系的强弱只有定性的描述。在此次项目中，我们需要使用算法将原来定性的关联规则量化。在这个过程中，算法产生的关联关系系数有时会与业务人员对关联关系强度的感知存在偏差。 解决思路：加强技术与业务的融合，通过对系统的使用反馈和算法的不断优化，使关联分析的自动化结果与业务判断形成一致。 2. 算法逻辑在业务上的解释性 难点：本项目在很多环节使用了各类算法解决问题，在此过程中，我们经常需要面对算法逻辑在业务上的是否可以解释的问题。例如，在高频交易行为相似性的研究中，传统交易特征向量相似算法除了提供一个相似度数值外无法说明高频交易行为的特点，

		<p>也无法从交易策略、交易偏好的角度说明不同客户间行为相似的特点是什么。</p> <p>解决思路：在项目中，我们在使用机器学习算法拟合客户交易行为后，将交易行为的偏好特点以决策树的形式进行展现和说明，决策树的分支策略即为客户在特定交易信号出现时的交易策略。并且，通过比较不同客户决策树分支策略的相似和差异程度，计算客户交易行为的相似性。</p>
	1.8 专利、认证或奖项	(项目所获得的专利、认证或奖项的名称、时间及颁发单位等主要信息。)
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）： 不涉及
		2.1.2 本次申报项目业务场景涉及的业务资格： 不涉及
	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）： 据《期货交易所管理办法》和《中国证监会监管科技总体建设方案》，不存在违反禁止性规定的情形
		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）： 根据《中国金融期货交易所交易规则》、《中国金融期货交易所异常交易管理办法》、《中国金融期货交易所风险控制管理办法》，不存在违反禁止性规定的情形
		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）： 根据《中华人民共和国数据安全法》和《期货交易管理条例》，不存在违反禁止性规定的情形
	*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）： 中国金融期货交易所
2.3.2 出具时间（如包含有效期的请注明）： 2022年1月17日		



		<p>2.3.3 评估结论（最终结论）：</p> <p>“基于知识图谱的关联账户分析系统”项目业务上、技术上均遵照相关法律法规规定，依法合规开展各项业务。在业务范围上，中金所按照《期货交易管理条例》依法履行自律监管职责，落实《中国证监会监管科技总体建设方案》的相关要求运用知识图谱和人工智能技术建立关联账户分析系统，可以有效的分析识别疑似关联客户群体，为更好发现和处置市场违法违规行，提供了有效的数据支撑。在业务流程和操作上，依据《中华人民共和国数据安全法》、《期货交易所管理办法》等，制定了《中国金融期货交易所业务数据管理办法》，对数据安全管理和数据密级进行了定义，严格规定了数据物理隔离机制和操作授权机制，确保数据安全和业务合规。</p>
<p>三、有序 创新原 则评估</p>	<p>3.1 技术创新情况</p>	<p>（试点项目所使用的新兴技术及为业务赋能的基本原理，与传统技术方案相比的价值体现。涉及多项技术应用的，可逐条列明，同时注明多项技术的融合应用原理与价值。）</p> <p>本项目中应用了大数据、知识图谱和人工智能等技术，自主研发了关联账户分析系统，综合应用各类数据，进行模型的融合和深加工，应用知识图谱和人工智能算法等，实现对全量市场数据的筛查，筛查过程自动化，具有较高的覆盖率和识别的准确率，与BI分析平台进行可视化集成，有效提高了监管效率和智能化水平。</p> <p>具体包括：</p> <p>1. 大数据融合，广泛收集和整合了内部客户资料数据、监管共享数据、工商数据、终端登录数据、私募产品数据、舆情数据。并在此基础上进行深度加工，构建了高质量的数据模型。包括：整合内部客户资料、客户资料变化历史、证券账户信息、工商企业信息、产品信息，按照内部客户的类型构建了自然人、企业、产品的实体信息模型；将工商投资、个人工商、企业法人、董监高、产品出资等数据融合，构建了投资线索模型；利用穿透式监管信息、datafeed 登录信息、高精度 IP 数据等，构建终端登录线索模型；利用万得股东数据、证监会处罚数据、舆情数据等，补充已知关联账户、亲属关系等模型。</p> <p>2. 应用知识图谱技术，实现了对关联分析知识的持续积累和逐步迭代优化，将客户资料、穿透式监管数据、工商数据、IP 数据等不同领域内的数据进行综合，形成了一个关系网络，最终可以以可视化的方式展现实体间的关联，能够更快、更简单的发现关联</p>

		<p>账户的线索，让关联分析工作更加简单和易于使用。</p> <p>3. 综合应用人工智能算法实现关系网络的剪裁和优化，使关联分析结果更准确，提升分析效率。具体包括使用路径搜索算法进行关联线索的发现，使用连通区域进行关联账户组识别，应用强连通区域和社区发现算法对关联组进行裁剪和优化。</p> <p>4. 应用机器学习算法进行于交易行为相似的判定，基于期货和现货行情以及客户内交易特征建立客户级的交易模型，再针对不同客户交易模型进行交易拟合分析，创新性的在客户交易模型拟合过程中分析可以交易策略的特点，使用决策树的分支策略展示交易策略的逻辑，并通过比较不同客户决策树模型的差异完成交易行为相似的判定。传统计算行为相似算法的通常将交易行为的特征直接向量化，通过计算向量的相似性判断交易行为的相似性。</p> <p>5. BI 分析平台集成。使关联分析系统与 BI 分析平台的客户画像、市场运行分析、重点客户监控等模块无缝衔接，提高关联分析便捷性和易用性。</p>
3.2 技术领先优势		<p>（项目技术应用、业务模式、工作流程等属于首创还是对同业做法有显著改进；所用技术先进性衡量指标及相对其他同业做法的主要优势，如：算法、技术路线、设备平台等方面。）</p> <p>在核心算法的先进性上，项目中应用了知识图谱平台、人工智能算法，建立了客户关联的多维度复合分析模型，相对于传统的关联分析模型，分析效率大幅提升，并支持知识的逐步累积，识别效果随着业务与技术的融合完善，识别效果会逐渐提升；</p> <p>在数据准确度方面，采用了 18 类静态数据特征，以及客户级别的交易模型拟合和相似度判定，自动产生的关联组对当前已掌握关联组的覆盖率达到 93%，并在现有数据基础上提供了更多新关联线索的发现；</p> <p>平台灵活性和用户体验方面，项目将图谱的可视化技术与关联排查的业务流程相融合，实现关联检索、图谱展示、关联排查、疑似关联组确认的业务操作闭环，用户可直接进行交互式分析，这大大提升了分析效能和用户体验。</p> <p>平台体系方面，构建了基于大数据平台、机器学习计算平台和知识图谱平台等，提供了完整全面的平台能力，有效支持了大数据分析和监管科技应用。</p>

	3.3 服务对象与渠道	<p>(试点项目上线后的预期服务对象, 区分内/外部, 区分机构/个人; 涉及个人投资者的, 应详细描述获客渠道、服务方式、适当性要求等; 试点单位应按照风险可控原则合理确定服务投资者范围、规模和适当性要求等。)</p> <p>服务对象: 交易所内部监查业务人员 其它服务对象: 资本市场监管核心机构</p>
四、风险可控原则评估	4.1 业务风险控制	<p>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): 该系统主要是负责关联账户分析, 所产生的数据供内部监查人员进一步研究、分析使用, 不涉及相关风险。</p>
	4.1.2 风险监测机制	<p>(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): 不涉及</p>
	4.1.3 风险控制措施	<p>(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): 不涉及</p>
	4.1.4 应急预案	<p>(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): 不涉及</p>
四、风险可控原则评估	4.2 技术风险控制	<p>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):</p> <ol style="list-style-type: none"> 1. 基础设施故障风险; 2. 不可抗力的系统性风险 3. 关联分析系统运行异常风险 4. 数据安全风险 5. 算法风险, 有可能误判, 不直接应用, 通过人工方式进一步筛选确认, 风险可控
	4.3.2 风险监测机制	<p>(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):</p> <ol style="list-style-type: none"> 1. 交易所针对基础设施和软件系统运行部署了集中的监控系统, 可以及时发现基础设施和软件本身运行过程风险和问题, 以利于快速发现和处置; 2. 交易所配备了专业化运维团队结构, 严格执行值班制度, 实现7*24小时专人值守, 以保障系统的稳定运行。
	4.3.3 风险控制措施	<p>(应描述如何采取措施来防控上述技术风险, 针对各类风险分别列举):</p>

		<p>1. 交易所的 IT 系统整体采用了两地三中心的部署方式，支持数据中心的快速切换，以应对不可抗力的系统性风险；</p> <p>2. 知识图谱系统本身以及所访问的大数据平台、人工智能平台和图计算平台都采用了多节点并行运行的部署架构，数据多份冗余，部分设备运行故障对系统运行的基本无影响；</p> <p>3. 交易所制定了《中国金融期货交易所业务数据管理办法》，对于数据使用流程、数据授权等进行了严格的规定，敏感数据限制在受控环境访问；数据访问采用最小授权原则，即所授予的权限是要完成任务所必需的最小权限，权限授予采用了审批与操作分离的方式。</p> <p>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)： 交易所制定了《中国金融期货交易所突发事件应急预案》和《中国金融期货交易所业务持续和应急管理办法》，针对系统性运行风险制定了应急处置方案，并开展定期演练提升应急处置能力。</p>
	<p>*4.3 投资者保护机制</p>	<p>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)： 内部使用系统，不涉及。</p> <p>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)： 内部使用系统，不涉及。</p> <p>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)： 内部使用系统，不涉及。</p> <p>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)： 内部使用系统，不涉及。</p>

附页：

<p>牵头申报单位 承诺</p>	<p>本单位郑重承诺：</p> <ol style="list-style-type: none">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。 <p>以上承诺如有违反，愿承担相应责任与后果。</p> <p>单位（公章） </p> <p>法定代表人（签字）： </p> <p>2022年 1月 21日</p>
<p>联合申报单位 1 承诺</p>	<p>本单位郑重承诺：</p> <ol style="list-style-type: none">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。 <p>以上承诺如有违反，愿承担相应责任与后果。</p> <p>单位（公章）</p> <p>法定代表人（签字）：</p> <p>年 月 日</p>

（注：联合申报单位如多于 1 家，承诺签章栏请相应增加）