

中期协联合研究计划（第十七期）项目

人工智能在期货行业智能风控的应用

Application of Artificial Intelligence in Intelligent Risk Control
for the Futures Industry

时间

2024 年 10 月

摘要:

本文详细概述了人工智能在智能风控领域的应用，特别是在期货行业基于大语言模型的舆情分析和风险预警系统的构建与应用。具体来看，本文先是通过案例分析等方法着重回顾了目前在金融领域较为成功的本文通过引入大语言模型技术，实现了对海量非结构化数据的自动化处理和情绪分析，构建了能够实时分析财经新闻、社交媒体内容和市场动态的智能风控体系。该系统在“数据采集-分析处理-风险识别”各阶段实现了自动化，能够实时捕捉和分析市场变化与潜在风险，并通过情感分析和数据筛选，为客户和风控团队提供风险预警和舆情分析支持。这种创新设计使期货公司能够在快速变化的市场环境中提升决策支持能力和响应速度，显著提高了风险管理的准确性与效率。

另外在文章第五部分提出了基于 AGI 的关联交易识别框架。关联交易识别是反洗钱检测中的核心环节。成交数据因其量大、内容复杂，传统规则化的检测手段很难处理；基于 AGI 的多算法协同分析则能精准识别关联交易网络中的高风险账户。我们的框架以多层次模块化设计为基础，涵盖数据收集、储存、分析、可视化四大模块。框架借助图数据库和图分析算法，高效存储、识别关联交易。我们相信这一设计能在识别团体协作洗钱风险起到重要作用。

项目成果显示，系统在数据收集、处理、风险识别到可视化展示等各方面实现了创新突破，构建了一个自动化、实时、高效的智能风控体系。该系统成功提高了期货公司在复杂市场环境下的风险应对能力，展示了 AI 技术在舆情分析和智能风控中的广泛应用潜力。通过多源数据实时整合与智能处理，系统显著提升了市场信息捕捉与风险管理的精准度，同时实现了合规性和隐私保护，为期货公司提供了强有力的支持。此外，项目为其他金融机构在智能风控和反洗钱监控方面提供了宝贵的参考经验，展示了大语言模型和图分析技术在金融科技中的广阔前景。

关键词：人工智能、大语言模型、智能风控、舆情分析、关联交易识别

ABSTRACT:

The article provides a detailed overview of the application of artificial intelligence in the field of intelligent risk control, with a particular focus on the construction and application of sentiment analysis and risk warning systems based on large language models in the futures industry. Specifically, the paper begins by reviewing successful cases in the financial sector through case studies and similar methods. By introducing large language model technology, the paper achieves automated processing and sentiment analysis of massive amounts of unstructured data, building an intelligent risk control system capable of real-time analysis of financial news, social media content, and market dynamics. The system achieves automation in each phase of “data collection–analysis processing–risk identification,” enabling real-time capture and analysis of market changes and potential risks. Through sentiment analysis and data filtering, it provides risk warnings and sentiment analysis support to clients and risk control teams. This innovative design allows futures companies to enhance decision support capabilities and response speeds in rapidly changing market environments, significantly improving the accuracy and efficiency of risk management.

In addition, the paper’s fifth section proposes an AGI-based framework for identifying related transactions, a core element in anti-money laundering (AML) detection. Due to the large volume and complexity of transaction data, traditional rule-based detection methods struggle to handle it. The AGI-based multi-algorithm collaborative analysis can precisely identify high-risk accounts within related transaction networks. The framework is based on a multi-layer modular design that includes four main modules: data collection, storage, analysis, and visualization. Leveraging graph databases and graph analysis algorithms, the framework efficiently stores and identifies related transactions, and is expected to play a crucial role in identifying coordinated money laundering risks.

The project results indicate that the system has achieved innovative breakthroughs across various aspects, from data collection and processing to risk identification and visualization, building an automated, real-time, and efficient intelligent risk control system. This system significantly improves futures companies' ability to respond to risks in complex market environments, showcasing the broad application potential of AI technology in sentiment analysis and intelligent risk control. Through real-time integration and intelligent processing of multi-source data, the system greatly enhances the precision of market information capture and risk management, while also achieving compliance and privacy protection, providing robust support for futures companies. Additionally, the project offers valuable insights for other financial institutions in intelligent risk control and AML monitoring, highlighting the expansive prospects of large language models and graph analysis technology in financial technology.

Keywords: Artificial intelligence, Large language model, Intelligent risk control, Public opinion analysis, Related party transaction identification

目录

- 一、引言.....1
- 二、人工智能在期货公司智能风控研究的背景与意义.....3
 - （一）人工智能过去的发展和新的方向.....3
 - 1. AI 浪潮的历史进程3
 - 2. 生成式 AI 应用发展现状及未来展望.....4
 - （二）期货公司对于人工智能以及智能风控的新建设背景.....8
 - 1. 政策驱动.....8
 - 2. 行业发展需求.....8
 - 3. 技术革新带来的探索.....9
- 三、国内外金融机构在智能风控上成功的案例.....11
 - 1. 摩根大通运用 AI 在金融科技和智能风控上的实践.....11
 - 2. 平安银行在智能风控领域的应用.....12
- 四、基于大语言模型的舆情分析与智能风控层面框架与应用14
 - （一）基于大模型的期货行业智能风控多场景系统整体框架介绍14
 - （二）期货行业智能风控系统搭建难点及解决方案.....15
 - （三）大语言模型在舆情分析与智能风控层面应用案例介绍18
 - （四）基于大语言模型生成信息的收益与波动率分析21
 - （五）大语言模型在舆情分析与智能风控层面总结与展望23
- 五、基于 AGI 的关联交易识别框架.....26
 - （一）基于 AGI 的关联交易识别框架的应用背景.....26
 - （二）基于 AGI 的关联交易识别框架的具体结构.....26
 - 1. 基于 AGI 的关联交易识别框架——数据输入层27
 - 2. 基于 AGI 的关联交易识别框架——数据储存层27
 - 3. 基于 AGI 的关联交易识别框架——图分析算法模块.....31
 - 4. 基于 AGI 的关联交易识别框架——可视化与用户界面模块33
 - （三）基于 AGI 的关联交易识别框架落地挑战和解决方案.....36
- 六、总结与未来展望.....37
 - （一）总结.....37
 - （二）未来展望.....37
- 参考文献.....40

一、引言

近年来，金融科技的迅猛发展已深刻改变了金融行业的诸多面貌，而人工智能（AI）作为这一转型的重要推动力量，其在金融服务领域的广泛应用尤其是在风险管理方面展现出强大的潜力。期货市场作为金融市场的关键组成部分，其高杠杆和高风险的特性使得风险管理需求尤为迫切。传统的风险控制模型往往依赖于人工判断和历史数据分析，这不仅限于执行速度慢，还在海量数据处理能力上显得捉襟见肘。尤其在当前经济环境变化愈发迅速的时代，传统的风控模式难以应对市场的瞬息万变。因此，整合先进的人工智能技术为期货公司的风险控制提供了新的机遇和挑战。

人工智能技术的迅速进步，尤其是机器学习、自然语言处理和深度学习的应用使得金融机构在风险管理方面能够获得更高的效率和更强的预测能力。期货市场产生的数据量巨大，作为一个动态变化的金融市场，如何迅速而精准地识别和处理数据中隐藏的风险信息，成为了期货公司必须面对的课题。AI 的强大数据处理能力能够快速分析和挖掘这些海量信息，提取出关键风险信号。相比传统的人工监控和人工数据分析，AI 的引入显著提高了工作效率与准确性。

值得关注的是，虽然传统的智能风控体系依赖专业人士的参数设定，而这一过程往往漫长而复杂。然而，眼下备受关注的的大语言模型（如 GPT 等）凭借其自适应和自我学习的能力，能够更快、更高效地搭建起 AI 风控系统，极大地增强了传统风险控制的灵活性。此外，这些大语言模型的文本理解能力和信息处理能力，使其能够在更高层面上进行复杂的数据分析和情绪识别，为风险控制提供了全新的技术路径。

本文将进一步探讨人工智能在期货行业智能风控中的多种应用场景和技术路径，尤其重点分析 AI 在舆情实时监控预警系统和关联交易监控系统建设中的具体应用及其优势。通过对实际项目的探索与尝试，我们期望推动这些应用场景的落地，实现技术创新与实践的深度结合。

在智能风控系统的建设中，舆情实时监控预警系统无疑是一个重要应用领域。舆情的变化往往与市场情绪密切相关，能够直接影响交易决策和市场走向。因此，通过对新闻报道、社交媒体动态和其他公共信息源进行实时监控与分析，期货公司可以及早识别潜在的市场风险。这种监控系统利用 AI 技术中的情绪分析算法，能够对大量文本数据进行快速处理，评估市场参与者的情绪变化，并及时捕捉可能引发市场波动的关键事件。

例如，通过对社交媒体中的舆情分析，期货公司可以了解投资者对某一事件或市场动向的反应，进而做出相应的风险防控决策。此外，结合地理位置和社交网络的分析，舆情监控系统能够捕捉到某一特定区域或行业的情绪变化，为期货公司提供更加细致的市场洞察。

在构建舆情监控框架的过程中，关键是如何将大语言模型的优势与期货行业

的实际需求相结合。大语言模型在处理自然语言理解和生成方面的能力使其能够高效地从非结构化数据中提取有价值的信息。因此，本文将尝试设计一个全面的舆情监控系统，包括数据获取、数据清洗、情绪分析、事件识别和风险预警等多个模块。通过自动化的数据处理流程实现实时监测，保障期货公司在市场波动中更具灵活性和应变能力。

除了舆情监控以外，另一重要应用场景则是关联交易的监控系统。关联交易常常涉及大量资金的频繁流动，若未能妥善管理，可能导致洗钱、内幕交易及其他违法行为，带来严重的合规风险。受限于传统人工监测手段在复杂交易模式识别上的局限，期货公司在监管合规与风险防控方面的要求愈加高涨。AI 技术的引入为此提供了新的解决方法。

本文将从期货公司稽核部门的实际需求出发，构建一个 AI 驱动的关联交易监控框架。该监控系统能够通过对交易行为进行深度学习与分析，识别潜在的异常交易和操控行为。一方面，该系统利用 AI 算法对海量交易数据进行自动化处理，从中挖掘出隐藏的关联特征；另一方面，通过规则与模式的匹配，系统能够及时发现可能涉及市场操控、内幕交易等风险行为。

结合实际案例，我们将分析 AI 驱动的关联交易监控系统的运行效果和应用前景。以某期货公司为例，其在部署 AI 风险监控系统中后，通过对交易数据的实时分析，不仅有效识别出了多个违规交易，还大幅提高了合规审查效率，降低了因合规风险而导致的损失。此外，随着技术的不断进步，该系统的学习能力也在不断提升，能够随着交易模式的变化不断进行自我优化，确保有效应对新的风险挑战。

总的来说，人工智能在智能风控系统的建设与应用上，表现出了巨大的潜力和价值。通过搭建 AI 驱动的舆情监控与关联交易监控系统，期货公司能够获得更高效、精准的风控能力，实现对市场波动风险的及时预警与应对。AI 智能风控系统不仅提升了风控的实时性与准确性，同时还推动了风险管理流程的智能化和自动化，为期货公司在变换莫测的市场环境中提供更全面、更具有前瞻性的风险控制手段。

在未来的发展中，期货行业的智能风控将持续与 AI 技术深度融合，形成一套更加灵活、高效、智能的风险控制体系。这一体系不仅需要持续创新与完善，还需要与特定法规环境与市场条件相适应。随着技术的发展与应用的深入，未来的金融市场将在 AI 技术的助力下向着更加安全、透明、公正的方向迈进，实现真正意义上的风险管理智能化的愿景。通过不断的技术创新与实践探索，期货行业将迎来更加辉煌的未来。

二、人工智能在期货公司智能风控研究的背景与意义

（一）人工智能过去的发展和新的方向

1. AI 浪潮的历史进程

人工智能技术发展经历了四个重要的阶段，每个阶段都有其特定的技术特点和应用领域，以下是对于这四次浪潮的描述：

第一次浪潮（1956-1976 年）是人工智能的起步发展期，这一时期的核心是逻辑主义。逻辑主义主要关注于使用机器证明的方法去证明和推理知识，例如用机器证明数学定理。这一时期的研究集中在逻辑抽象、逻辑运算和逻辑表达等方面。数学定理证明是这一时期实现效果最好的领域，出现了许多与逻辑证明相关的逻辑程序语言，如 Prolog。然而，这些成果虽然能够解决拼图或简单游戏，但几乎无法解决任何实用问题。

第二次浪潮（1976—2006 年）是联结主义盛行的时期。70 年代末，神经元联结网络和模型取得了显著进步，尤其是 BP 前馈神经网络的出现，解决了不少问题，并在模式识别、手写文字识别、字符识别、简单人脸识别等领域得到应用。然而，随着研究的深入，人们发现神经网络在解决单一问题上表现良好，但在处理复杂问题时却力不从心，尤其是在训练学习时数据量巨大，以当时的算力难以达到要求，且结果在一定程度上难以进一步提升。

第三次浪潮（2000s-2020s）是人工智能发展的关键时期，以深度学习技术的突破和应用为标志。这一时期，深度学习算法，尤其是卷积神经网络（CNN）和递归神经网络（RNN），在图像和语音识别等领域取得了显著进展。大数据的增长和计算能力的提升，尤其是 GPU 和 TPU 等专用硬件的发展，为深度学习提供了强大的支持。深度学习的进步使得人工智能在多个领域实现了商业化应用，如 Apple Siri 的语音识别、Microsoft Bing 的图像搜索等。

第四次浪潮（2023-至今）被认为是通用人工智能（AGI）时代的到来。这一阶段的人工智能技术正朝着能够执行任何智能任务的方向发展，而不仅仅是特定任务。2017 年，Vaswani 等人在《Attention is All You Need》中率先提出 Transformer 模型，它的核心创新在于自注意力机制，使得模型能够在处理输入序列时动态关注不同部分，从而捕捉到长距离依赖关系。Transformer 模型由编码器和解码器两部分组成，编码器负责输入序列的特征提取，而解码器生成输出序列。这个模型经过了改进之后，就成为了划时代产品-ChatGPT 的原型。这一代的人工智能技术，以 ChatGPT 等生成式 AI 模型为代表，展示了在自然语言处理、内容生成等方面的强大能力。开源人工智能的发展、许可辩论以及生成式 AI 模型的出现，标志着人工智能领域的深刻变革。此外，多模态学习、强化学习等技术的发展，以及对人工智能可解释性、伦理和安全性的深入研究，都是第四次浪潮的重要特征。当前，我们就处于被称之为 AIGC 最重要的奇点时刻，未来的时期将是人工

智能在社会经济活动中深度融合、广泛应用的时期。

在这四次浪潮中，人工智能可以大致分为两种技术路径，分别为决策式 AI 和生成式 AI，以下是他们的主要区别：

决策式 AI：也被称为判别式 AI，指利用人工智能技术来辅助或者自动化决策过程的一系列方法和系统，决策式 AI 通过训练集得到相应参数，识别数据中的某种规律，并解决某个业务相关问题。在应用方向上，决策式 AI 多应用于人脸识别、智能推荐、自动驾驶和智能风控等方向，在智能风控领域，决策式 AI 目前主要应用于银行的身份信息核准、信用评估等等场景。

生成式 AI：是指基于大型预训练模型等人工智能的技术方法，通过已有数据的学习和识别，以适当的泛化能力生成文本、图片、代码、音频等相关内容额的方法。早期的生成式 AI 主要是生成简单的图像和文字，但随着 Transformer 等大语言模型带来的不断突破，生成式 AI 也迎来了质变，可以处理更加泛化的问题，并随着算力的提升大大节约开发者的时间。

下表中详细对比了两种 AI 的相异性。

表 1 决策式 AI 与生成式 AI 横向对比

比较维度	决策式 AI	生成式 AI
定义	用于判断、预测和决策	用于生成新的内容
工作原理	基于数据和规则做出最优决策	通过生成模型学习数据分布生成内容
应用场景	分类、预测、优化等	文本生成、图像生成、音频生成等
输入	数据或特征	初始条件（或无输入）
输出	决策、预测结果或标签	新生成的内容（如文本、图像）
算法示例	决策树、SVM、随机森林、强化学习	GAN、VAE、Transformer
优点	决策准确、可解释性强	可以生成多样化内容
缺点	依赖规则，无法生成新数据	控制性差，可能生成不合规内容

2. 生成式 AI 应用发展现状及未来展望

生成式 AI (Generative AI) 近年来取得了快速发展，特别是在自然语言处理、图像生成、视频生成和音频合成等领域，带来了显著的技术进步和广泛的商业应用。大语言模型 (LLMs，如 GPT、BERT 等) 带来了自然语言处理领域的质变，其革命性影响主要体现在以下几个方面：

a) 通用性和任务泛化能力的提升

传统的自然语言处理模型通常是为单一任务构建的。例如，情感分析、机器翻译、命名实体识别大语言模型通过预训练阶段在海量数据上学习了通用的语言知识和结构，不局限于特定任务。通过少量数据微调，它们能够适应各种不同任务，如文本生成、情感分析、信息抽取、翻译等。这种任务泛化能力减少了模型

开发的重复工作，为研发人员节省了大量时间和成本。等任务需要分别设计和训练独立的模型。这不仅费时费力，还导致模型无法泛化到新任务。由于大语言模型可以灵活地应用于多种任务，它们大大拓宽了 AI 在各行业中的应用场景。例如，金融、法律、医疗等行业可以利用单一的大语言模型处理各类文本分析和生成任务，极大提高了应用效率。

b) 上下文理解和连贯性显著增强

早期的自然语言处理模型通常只关注局部的词组或短句，缺乏对长文本的整体理解。RNN（循环神经网络）和 LSTM 虽然引入了序列化处理，但由于长程依赖问题，难以有效处理长文本的上下文。大语言模型（如 GPT、BERT）基于 Transformer 架构，利用自注意力机制能够在处理长文本时保留前后文信息，捕捉文本中的长程依赖。这使得它们在理解文章主题、语义关系和逻辑连贯性方面表现更优异。再实际运用中模型上下文理解能力增强，使得大语言模型可以生成更连贯的文本，并且在长对话系统中具备“记忆”之前对话内容的能力，提升用户体验。例如，在自动写作和多轮对话任务中，模型能够保持话题连贯，并对用户的上下文意图做出更精准的响应。

c) 生成式任务的跨越性进展

过去的文本生成模型多基于规则和模板，生成的文本僵硬且缺乏灵活性。生成对抗网络（GAN）和变分自编码器（VAE）在图像生成方面表现优秀，但对于高质量的文本生成仍然存在局限。大语言模型在生成式任务中展现了突破性的能力。它们可以生成连贯、符合语法的文本，甚至模仿特定风格的写作。用户可以向模型输入一个提示（prompt），大语言模型能够根据提示生成回答、故事、代码、摘要等内容。生成能力大幅拓展了 AI 的应用范围。例如，在内容创作领域，新闻写作、广告文案、社交媒体内容生成都可以借助大语言模型。在教育领域，模型可以生成练习题、解释复杂概念等。这种生成式任务的跨越性进展使得 AI 从“理解”语言发展到“创造”语言，对媒体、教育、营销等领域产生了颠覆性影响。

d) 减少对标注数据的依赖

大多数传统的自然语言处理模型需要大量的标注数据进行监督训练，但数据标注非常耗费人力成本，尤其是对于低资源语言或小众领域的的数据。大语言模型采用了无监督或自监督的方式，通过大量未标注的数据进行预训练。这种预训练能够让模型在海量数据中学习语言模式和知识结构。之后，模型只需在少量标注数据上进行微调，即可适应具体任务。这减少了标注数据的需求，使得 AI 技术更容易应用在低资源语言或小众任务中。对于许多公司和研究机构来说，标注数据的获取成本非常高。大语言模型大大降低了训练数据的要求，特别适合在标注数据不足的场景下应用，如小众语言的翻译、特定行业的情感分析等。这为语言处理技术的普及和跨领域应用奠定了基础。

e) 人机交互体验的质变

大语言模型的出现让机器可以通过自然语言与人类进行更顺畅的交流，支持

多轮对话、上下文理解和个性化响应，接近人类的交互体验。这种质变提升了人机交互的效果，使得虚拟助手、客户服务机器人、教育辅导系统等的应用体验大幅改善，为 AI 在生活、教育和商业领域的推广带来了巨大的推动力。

大语言模型真正带来的质变是从“专用任务的工具”向“广泛适用的智能体”进化，使得 AI 系统能够更智能地理解和生成自然语言，为人类提供更强大、更自然的语言处理能力。这不仅改变了技术的能力边界，也深刻影响了各行业的应用场景和人机交互的方式。在我们进行项目实际开发过程中，我们也尝试用两种不同的方式解决实际问题，发现在一些传统 AI 相对完善的应用中，尝试利用生成式 AI 同样能得到非常不错的效果，而整体开发时间能大大缩减，虽然其正确率在进行微调以前仍然无法达到传统模型的水平，但可以预见大模型可以在不消耗大量资源和时间的情况下仍然可以有相当的拓展和推进。未来随着算力的不断提升，可以预见大模型仍然有非常好的上升空间。

从目前的全球普及率来看，截至 2024 年，生成式 AI 在全球的普及率快速上升。数据显示，约 72% 的企业已经在至少一个业务功能中采用了某种形式的 AI，其中 65% 的企业定期使用生成式 AI 技术。相比 2023 年的 33%，这个比例几乎翻倍。

在个人使用方面，生成式 AI 的普及也非常显著。据调查，在美国，18 至 64 岁的人群中，39% 已经使用过生成式 AI，其中 24% 的人在工作中至少每周使用一次，而几乎 1/9 的工人每天都在使用该技术。这些数据表明生成式 AI 在工作和日常生活中的影响力正在不断扩大，其采用速度已经超过了历史上个人电脑和互联网的普及速度。

企业层面，各行业对生成式 AI 的采用情况有所不同。媒体、电信、科技和金融服务等行业中，生成式 AI 的普及率较高，这些行业报告称已经通过生成式 AI 实现了效率 and 创新的提升。然而，人才短缺和技术信任度问题仍是一些企业在大规模部署生成式 AI 时面临的主要挑战。

图 1 为目前 AI 在全球不同行业的应用占比，目前还主要集中在科技电信和金融服务领域，但随着大语言模型的不断发展，各行各业的比重都在不断增加。

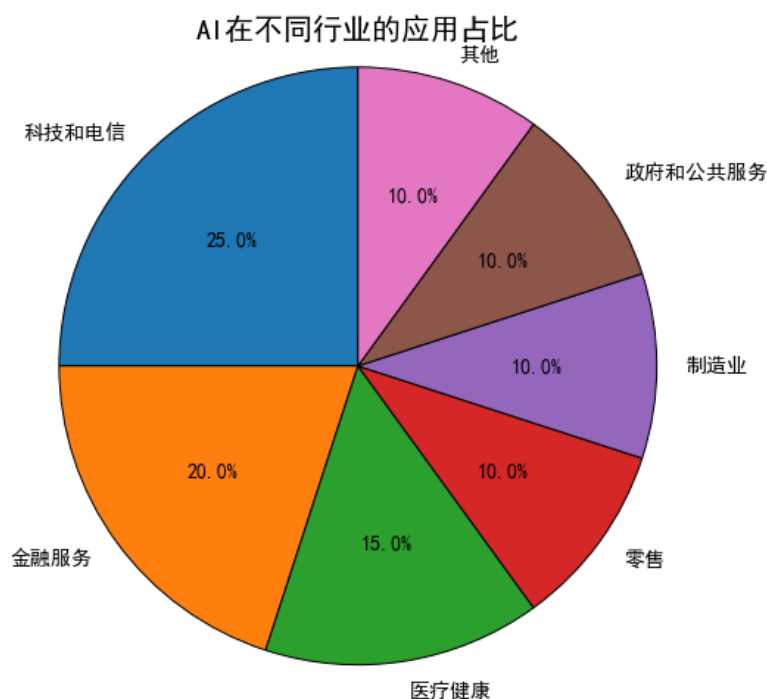


图 1 AI 在不同行业的应用占比

从投资情况来看，中国的人工智能投资正经历快速增长，尤其是在生成式 AI、大模型开发以及基础设施建设方面表现尤为突出。根据 2024 年的统计数据，中国 AI 行业的整体融资规模已经达到 5210 亿美元，同比增长了 13.9%。虽然受到全球经济环境的影响，中国的 AI 投资事件数量下降了约 18.2%，但令人瞩目的是，总融资金额却增长了 51%。这突显了国内市场对 AI 应用的持续旺盛需求。

目前，中国拥有全球 36% 的大语言模型，这些模型主要由国内知名科技巨头（如百度、阿里巴巴、腾讯）以及一些正在崛起的新兴 AI 公司（如智谱 AI、文心一言）所主导。自 2023 年以来，中国的监管机构已经批准了 180 多个大模型，使得 AI 技术在各行业中实现快速渗透。这一现象显示出中国在 AI 领域不仅在应用层面加速发展，也在政策环境上给予了支持，为创新提供了肥沃的土壤。

从国际视角来看，美国的人工智能企业投资同样表现出强劲的增长势头。尤其是在生成式 AI 以及相关硬件与软件技术的发展推动下，美国市场在这个领域呈现快速扩张的态势。据多方分析，预计到 2027 年，AI 市场规模将激增至 7800 亿至 9900 亿美元之间，年均增长率大约为 40% 到 55%。这种市场的扩展不仅限于大型模型的开发和计算基础设施的增长，更是因为大量小型企业和开源模型的出现正在不断推动行业创新。

整体而言，生成式 AI，尤其是大语言模型，将成为全球各国与企业竞争的关键领域。在不断提升算法和算力的同时，AI 技术相关的应用也在持续进行创新尝试，并日益成为资本市场关注的焦点。这一趋势不仅推动了科技的进步，也在重塑全球经济格局和各行各业的未来发展路径。未来，随着技术的逐步成熟和应用的深入，AI 的影响力将进一步扩大，并为经济和社会带来深远的变革。

（二）期货公司对于人工智能以及智能风控的新建设背景

1. 政策驱动

提高科技标准化水平是《证券期货业科技发展十四五规划》所确定的行业数字化转型的基本思路之一。当前，金融行业正处于推进数字化转型发展的关键阶段，期货行业应当加快数字化转型步伐，通过科技赋能，在业务上打造并形成核心竞争力。企业数字化转型的三大阶段分别为信息化、数字化和数智化，其中人工智能的应用是数字化转型中至关重要的数智化过程。通过研究人工智能大模型在期货行业中的应用，探索创新方法以提高服务效能和风险管理的可行性，能够提升期货公司的市场竞争力，优化客户体验，并为期货行业的金融科技创新提供理论支持和参考。

随着全球经济的发展与金融行业的扩张，期货公司的业务在过去几年的时间里得到了较大规模的提升，期货公司在进行业务的时候会面临五大风险，经纪委托风险、流动性风险、强行平仓风险、交割风险以及市场风险，期货公司及其他涉及经纪业务、期现业务、套保业务的金融机构在风险管理上面对更大的挑战，海量的数据尤其是非结构化的数据让人工处理变得及其耗时，而潜在的关联交易、合同造假等也给公司造成了极大的困扰。

为了更好的完善执行《期货公司风险监管指标管理办法》第四条，期货公司应当建立与风险监管指标相适应的内部控制制度及风险管理制度，建立动态的风险监控和资本补充机制，确保净资本等风险监管指标持续符合标准。在这个过程中运用大模型构建相应的风控监控体系既能保证风控系统及时监控及时预警，也能大幅减少人力消耗，提升效率。

2. 行业发展需求

随着期货市场的迅猛发展，交易量稳步提升，而伴随着客户数量的增长和市场波动性的加剧，期货公司对行情分析和风险预警的需求愈显迫切。全球范围内突发事件和各国政策变化对期货市场的影响往往是迅速而直接的，尤其是重大事件，会在短时间内导致剧烈的市场波动。在这样的市场环境中，期货公司以及各类分析机构对舆情分析的重视程度与日俱增，他们希望通过全面且快速的舆情监控系统来及时捕捉和分析潜在的市场风险。

然而，传统的舆情分析方法主要依赖关键词匹配和研究员的主观判断，这种方法不仅速度慢，而且在信息爆炸的时代，难以高效应对瞬息万变的市场变化。幸运的是，随着人工智能技术的飞速发展，新的文本分析和自然语言处理技术应运而生。这些技术可以自动处理海量文字信息，并进行情绪分析，从而能够实时捕捉潜在的风险信号，大大提高了舆情分析的速度和准确性。

基于 AI 的舆情分析系统实现了从数据收集、信息筛选到情绪判断的全流程自动化，显著减少了人工干预，提高了市场反应的速度。这种智能化的舆情分析

手段能帮助期货公司更有效地进行风险预警，并为客户提供更为及时和全面的风险管理支持。在市场瞬息万变的环境中，它为期货公司提供了强有力的保障，使其能够更从容地引导客户的投资决策。

最终，AI 驱动的舆情分析不仅提升了期货公司的竞争力，同时也推动了整个行业向更为智能化和高效化方向发展。在未来，随着 AI 技术的进一步成熟，舆情分析系统还将具备更强大的预测能力和更精准的市场洞察力，为金融市场的稳健发展提供更加坚实的基础。。

3. 技术革新带来的探索

人工智能技术的快速发展以及 GPU 计算能力的显著提升，为大语言模型的崛起和生成式 AI 技术的突破奠定了坚实的基础。据估算，自 2012 年至 2023 年，全球用于 AI 模型训练的算力增长已超过 30 万倍。这一算力的爆炸式增长直接推动了 AI 模型性能的提升。图 2 显示了部分大模型的评分排名，可以看出，GPT 等闭源大模型在综合评分上表现优异，并在今年实现了显著的性能突破。同时，多款开源大模型也相继涌现，展现出强劲的竞争力。从图 3 可以看出，今年模型的整体准确率较去年从 31% 提高至 63%，呈现出极高的提升速度和显著的效果。未来，随着算力的持续增长，AI 模型的性能和应用将会更加深入广泛。

从应用角度来看，麦肯锡和普华永道的数据显示，全球范围内 AI 的采用率正在迅速上升，尤其是在金融行业，AI 应用已十分广泛。然而，传统的智能风控系统大多依赖于规则驱动的决策式 AI，在既定的框架内运行，难以完全适应当前大模型带来的技术进步。对于期货行业而言，当前的风控体系仍然较为依赖人工审核。然而，大语言模型强大的自然语言理解能力为金融行业，特别是期货行业的智能风控普及和发展带来了新的契机。通过大模型技术，风控系统可以实现更高的自动化水平，显著提高在复杂市场环境中识别和应对风险的能力，从而帮助期货公司在信息爆炸和市场变化加速的时代中保持敏捷与精确的风险管理。

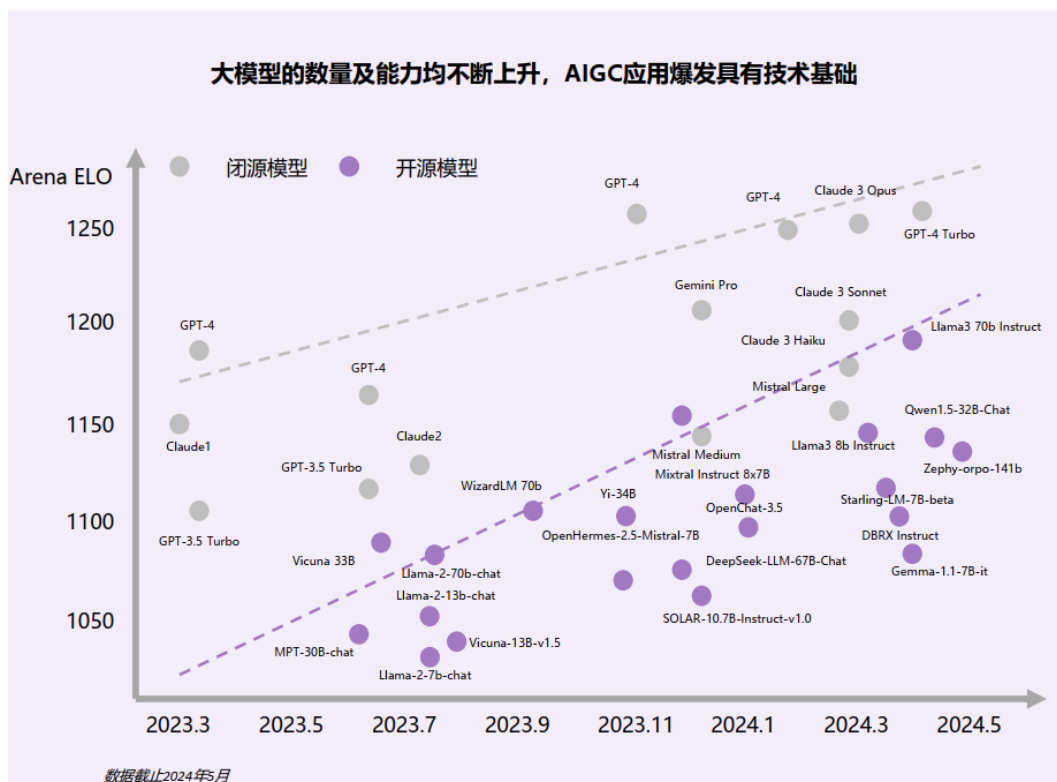


图 2 大模型在 Arena 综合评分与开发时间点阵图

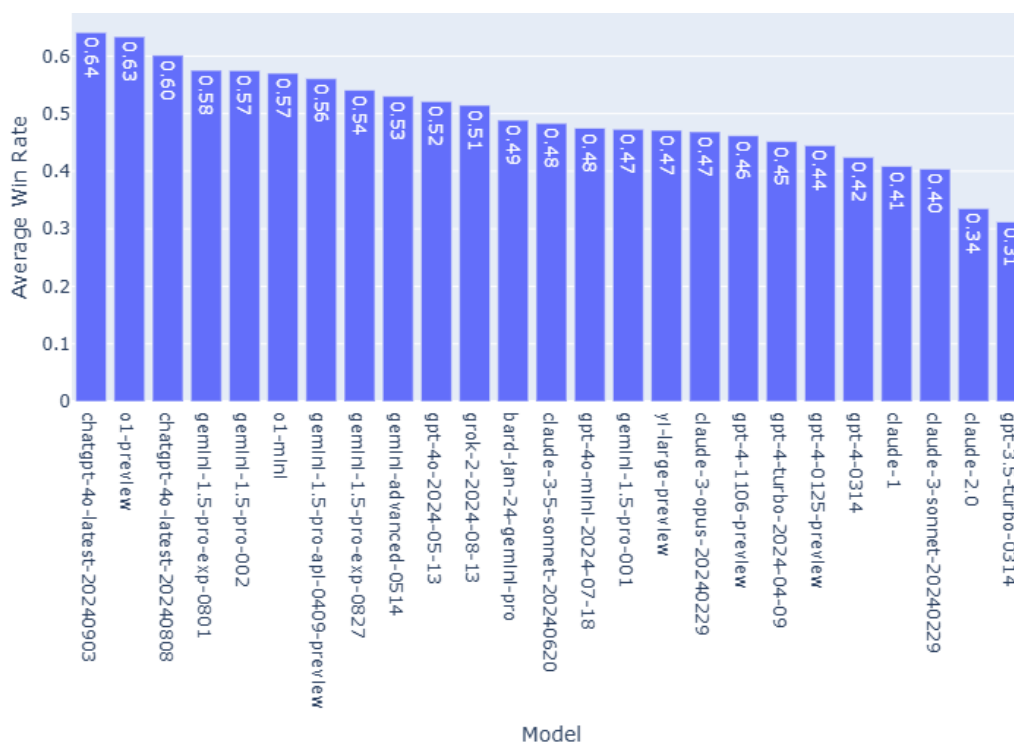


图 3 部分语言大模型平均综合正确率统计排序

三、国内外金融机构在智能风控上成功的案例

总体而言，风险管理框架可以分为四个阶段：

1. 识别：识别可能对组织战略或运营产生重大影响的风险，并监控外部和内部环境，以适应新的风险。
2. 评估：建立评估潜在风险暴露水平的流程，帮助理解整体风险状况。
3. 控制：应用控制框架，将风险降低至符合组织风险偏好的水平。
4. 监测与报告：制定方法评估控制措施的有效性，使用指标衡量成功，并定期报告风险状况和控制状态。

目前，国内外金融机构在智能风控上的应用大多以优化这四个阶段为主发展，传统决策式 AI 在建立了相对完善的识别和评估模型后能在特定领域达到比较好的效果，但随着大模型的问世以及其超预期表现，后续对于传统的智能风控仍有非常大的提升空间。目前多数成功运用 AI 技术并实际落地的风控系统主要来自于银行，且由于风控大多为公司中台系统，大部分为公司内部文件，尚未公开，但一些可以调查到的智能风控系统，其中也不乏一些重要的经验可供期货行业借鉴。下面具体展示国内外智能风控的成功案例。

1. 摩根大通运用 AI 在金融科技和智能风控上的实践

作为一家全球性全能服务银行，摩根大通涉及多方位繁杂的业务，其业务的繁杂也让摩根大通最早开始了数字化转型，2006 年，摩根大通的新上任 CEO 就极其重视金融科技应用，是金融行业最早引领技术革命的高管之一，并且在近些年大力推进“移动第一，数字渗透”的数字化战略，经历多年的发展，摩根大通已经成功开发了多款数字化产品大大提高了进行业务的效率，其中在智能风控领域，摩根大通同样非常激进。2017 年初，摩根大通开发了一款以机器学习技术推动的金融合同解析平台 COiN（Contract Intelligence），是一个利用人工智能和机器学习技术来处理法律文件的系统，为了处理在日常运营中需要的大量复杂的法律和财务文件，这包括贷款协议、交易合同等，这些文件通常由大量的法律和金融专业人员手动审阅和处理。这种做法不仅耗费大量时间，而且容易因为人为错误导致高昂的成本和风险。“COiN”平台也因此孕育而生，在接下来很多年不断进化，并且在业务中一直发挥重要作用。

该平台主要运用了 NLP 和机器学习技术，从非结构化的文本中提取相关信息，如合同条款、日期、金额、当事人等信息，再通过自动化方式来解读和分析合同条款，检查文件合规性，并识别潜在的风险和异常情况。

其主要功能为：

1. 合同信息提取：自动从合同中提取关键信息，包括合同的条款、日期、金额、当事人等。通过自然语言处理（NLP）技术，该平台可以高效解析合同文本，将非结构化数据转化为结构化信息，方便后续的分析和管理。
2. 合同分类归类：平台会对合同进行自动分类，将合同条款和内容归类到预定

义的合同类型中，比如贷款协议、租赁合同、服务协议等。这种分类和归类的功能有助于标准化合同的管理，简化了查找和检索的过程。

3. 合规性检查：通过内置的规则引擎和合规标准，对合同条款进行自动检查，以确保合同符合摩根大通内部的合规要求和行业标准。合规性检查功能帮助公司在合同签署前识别潜在的合规风险，降低了公司面临的法律风险。

根据报告，该平台能够自动分析成百上千页的法律文件，在几秒钟内完成过去需要 360,000 小时才能手动完成的工作。真正意义上为摩根大通达成了提升效率、节约成本、提高准确性的作用。

在未来的发展方向中，摩根大通还将继续扩展以下领域：

- 扩展应用领域：摩根大通可能会继续扩展 COIN 的应用领域，不仅限于法律文件，还可能应用于其他需要高精度数据处理的金融服务业务中。
- 增强智能化：通过进一步的发展 AI 技术，包括更先进的 NLP 技术，提升 COIN 对更复杂文本的理解能力。
- 集成创新：将 COIN 与企业的其他科技创新如区块链、云计算结合，提升整体数据管理和合规运营的效率。

COIN 平台的成功标志着 AI 在金融行业的一个重要应用方向，有效地展示了如何应用技术来优化传统业务流程，提高运营效率和准确性，并在智能风控领域拿下头筹。而对于传统 NLP 模型来说，由于其依赖于手工设计特征，先验特征，需要专家知识去从零构建，特征工程复杂且容易出错，需要投入大量的时间和专门知识。而随着大语言模型的急速发展，公司不再需要投入大量的人力和资源从零开发，只需通过海量数据预训练，学习通用语言结构和知识，然后在特定任务上进行微调，就能够有非常好的表现。开发语言理解模型的成本及效率大大提升，对于期货公司及其他金融机构，目前需要开发合同审阅及合规性检查的成本降低，且伴随着目前期货行业各类违规合同、欺诈问题的飙升，这类应用毫无疑问将在未来期货行业有非常多的应用场景和广泛需求。

2. 平安银行在智能风控领域的应用

近年来，平安银行在智慧风控平台上的持续升级，不仅为其在供应链金融领域提供了稳固的风险控制基础，也为整个金融行业树立了创新风控的标杆。通过“1 个核心企业+N 个关联企业”的供应链场景风控模式，平安银行有效整合了核心企业与关联企业之间的风险关系，打造了基于客户分层管理和风险自动化监测的完整风控生态。

平安银行智慧风控平台的关键模块主要如下：

- a) 智能预警模块：智能预警模块加大了对潜在风险客户的实时监测，确保在风险尚未扩大前对客户进行精准查验和风险评估。AI 技术的持续优化，使预警模型能够在不同场景中更高效地识别潜在风险，通过减少无效预警，平安银行的客户经理能够专注于高风险客户的应对与管理，提升整体风控效率。

- b) 智慧审批模块：通过数据分析和智能筛选，智慧审批模块帮助平安银行识别出信用优良、具有成长潜力的客户群体，从而优化资产配置和业务结构。通过 AI 技术的赋能，该模块不仅大幅提升了审批速度，还显著降低了审批误差，使新增业务的资产质量不断提高。
- c) 智能控制：平安银行将政策制度的要求转化为具体参数，并在系统中实现严格的准入和授权管控。智能控制模块的自动化执行，杜绝了“越权准入”和“越权审批”等风险，实现了技术防控的标准化与流程化，大大减少了人为操作风险。这种“技防机控”的方式使得风险控制更具一致性和透明度。
- d) 智能放款：智能放款模块依托内外部数据源，构建了风险信息、特殊管控、缓释工具和额度管控等多层次的风险模型体系。放款过程通过“三查”机制（即贷前调查、贷中审查、贷后检查）的机器自动化审批，使放款过程更加高效和严谨，从而确保资产的安全性。
- e) 智能监控：平安银行的智能监控模块实现了对人员、机构、效率、产品和行业的多维度监控，确保风控管理全面覆盖业务的每个细节。通过数据实时追踪和分析，管理层能够快速发现和响应风险，提高了风控和业务的协同效能。

平安银行的 AI 风控系统将风险管理框架中的所有阶段整合为一个高效的智能体系。通过智能预警、智慧审批、智能控制、智能放款和智能监控等模块，平安银行不仅提升了信贷业务流程的自动化和智能化水平，还显著提高了风险管理的效率和准确性。在运营成本方面，AI 技术的应用帮助银行减少了大量人工成本，将更多资源集中到高价值的客户服务中。

平安银行的创新实践为金融行业，特别是为期货公司提供了可借鉴的风险控制框架。当前，期货市场中反欺诈、反洗钱等风险监控和管理需求与日俱增，期货公司可以借鉴银行的风控模式，通过智能化风控系统对潜在风险客户进行监测和预警，通过大量的信息数据和交易特征提高整体的风险管理能力。此外，随着全球市场的不断波动和政策的快速变化，智能风控系统的实时监控和预警功能也为期货公司在动态市场环境中提供了重要的风险防控手段。

未来，AI 技术的不断进步将进一步推动智能风控的发展，金融机构在实现风险管控自动化和智能化的同时，也将在风险识别、合规管理和客户体验等方面获得更多创新的应用场景。平安银行的成功实践无疑为金融行业的智能风控建设提供了宝贵的经验和参考方向。

四、基于大语言模型的舆情分析与智能风控层面框架与应用

（一）基于大模型的期货行业智能风控多场景系统整体框架介绍

随着金融市场的复杂性和不确定性日益增加，快速且精确的市场信息捕捉和舆情分析已成为提升客户和公司风控体系里不可或缺的因素。传统的风险预警方法多数依赖于价格数据、成交量、技术指标等等结构化信息，但由于信息来源的多样性和新闻、社交媒体等非结构化数据的涌入，如何及时解读这些非结构化数据成为了有效市场监控和风险管理的重大挑战。

为应对这一需求，我们尝试构建了基于人工智能的行情实时监控及风险预警系统，通过引入自动化内容生成（AIGC）技术及情感分析模型，采用“大模型+微调”的模式对来自财经网站的海量文本信息进行自动处理和分析，构建起了能够实时识别市场趋势、捕捉潜在风险的智能化系统。本项目的宗旨是通过大模型、自然语言、大数据分析等技术，搭建一个期货行业全方位行情与舆情风险监控、风险管理系统，设计一个公司层面能实际落地的系统应用架构，并且设计两个实际应用于客户端的场景，分别为行情舆情实时监控和套保客户风险预警模块。系统的整体架构如下图所示。

整体架构主要分 4 个模块：数据层、场景服务层、安全层及输出层。其中数据层主要包含数据采集、清洗及储存模块；场景服务层除了该节重点关注的舆情监控及套保预警模块外，还包含了智能投研、智能客服等特色化模块；最终的输出层主要面向不同的用户端，包括 PC 端、移动端等等。

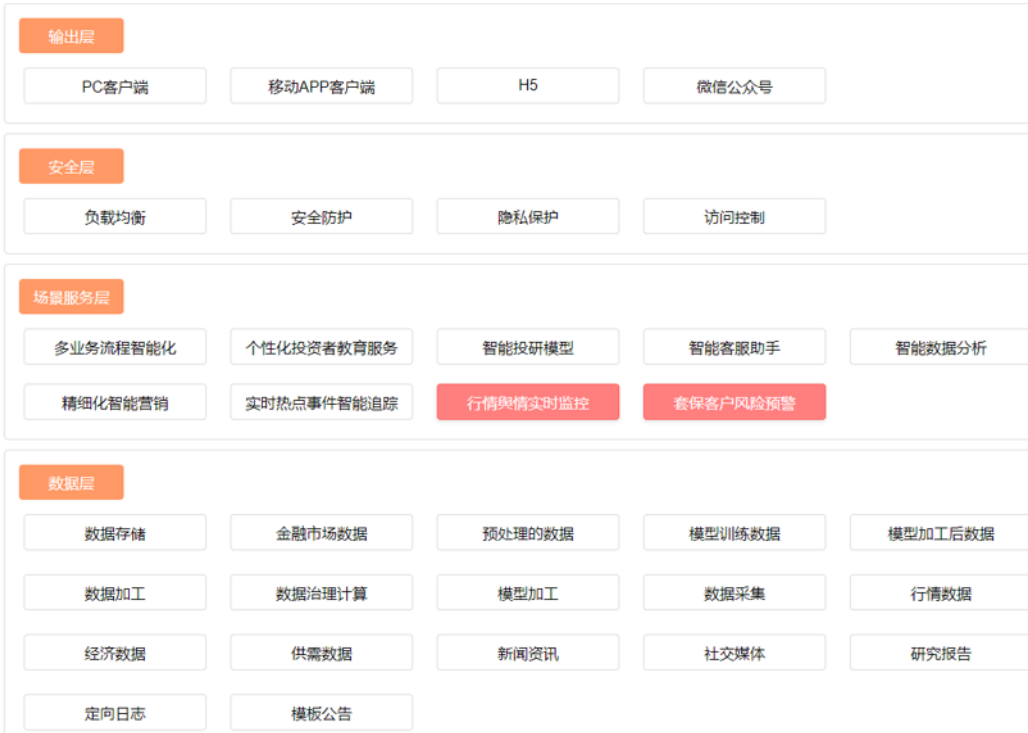


图 4 基于大模型的期货行业智能风控与市场监控多场景系统架构

为了进一步阐述该系统的核心架构，下面我们对应用架构中的具体模块展开

介绍。其中各模块的具体功能及技术方案如下：

(1) 数据层：

- a) 数据采集：从金融市场、新闻和社交媒体、专业网站、公司各部门等多个数据源收集数据。利用 DataX、智能采集器等方式获取期货相关数据。
- b) 数据加工：对采集的数据进行数据清洗、数据计算及模型加工。通过 Flink、MLIB、Spark、Flume、Hive、HBase、ES 等框架工具，实现原始数据接入并进行实时处理计算。
- c) 数据存储：对金融市场数据、预处理数据、模型训练数据等加工后的数据进行存储。

(2) 场景服务层：

包括多个场景服务层，这里主要介绍行情舆情实时监控和套保客户风险预警服务层。

项目通过 Pytorch、TecnorFlow 等深度学习算法框架基础上，采用 LLMs 大模型技术对数据进行挖掘抽取，并对数据进行品种分类。

(3) 安全层：

- a) 数据安全与隐私保护：由于该项目涉及数据较多，且会涉及到大语言模型的训练问题，实际上线需要非常注重数据安全问题，所以需要采用安全加密和权限控制等机制确保数据的安全和隐私。
- b) 安全与访问控制：实施平台级的安全措施，包括身份验证、访问控制和审计日志等，保护系统免受恶意攻击。
- c) 负载均衡：由于项目涉及到实时数据采集、计算和输出，需要确保项目能够承载一定的高负载，并且保证稳定性和性能。

(4) 输出层：

通过 PC 客户端、移动 APP、微信公众号等终端在不同的客户界面上提供服务自动化服务。

(二) 期货行业智能风控系统搭建难点及解决方案

由于期货行业数据参差不齐、冗杂凡多，这给智能风控系统的搭建带来了很多难点，如何有效整合多源异构数据、如何选择合适的算法及评估模型等等都是需要着重解决的问题。

针对其中可能面临到的难点及挑战，我们一一给出了具体的解决方案。具体结果如下。

首先，期货行业数据参差不齐、冗杂繁多。为了解决这个问题，在该项目中，我们运用 ChatGLM、BERT、UIE 等先进的生成式大语言模型（LLM）技术，从期货金融百科、期货行业教材书籍、行业权威网站等自动化式提取并生成与期货行业密切相关的语句内容并审核入库，提取的数据包含期货交易交割规则、保证

金制度、商品供需、产业链上下游、突发事件、宏观政策、品种讯息等。显著减少对人工的依赖，并且能够做到实时更新，确保信息数据库的时效性和准确性，后续将这些结构化和非结构化的部分数据作为语言大模型预训练数据，寻找监控作为提示风险预警的关键信息，提高期货行业获取信息的效率。在本项目中涉及到的具体模型总结对比详见下表 2。

表 2 各模型对比

模型	ChatGLM	BERT	UIE
全称	Chat Generative Language Model	Bidirectional Encoder Representations from Transformers	Unified Information Extraction
开发者	清华大学和智谱 AI 合作开发	Google AI Language	微软亚洲研究院开发
发行年份	2023	2018	2022
主要应用	对话生成及文本生成	语言理解任务（如问答、分类、填空）	信息抽取任务
架构	Transformer	Transformer	Transformer
优势	高效的中文处理能力 及对话能力优化	强大的语义理解能力	灵活适用于多种信息抽取场景
特点	高效的中文处理：对中文进行了深度优化，能在中文任务上取得更好的效果。 目前国内期货行业的信息需要较好的中文处理能力，目前海外的模型在中文处理上表现一般	BERT 模型的核心特点是其双向性，它能够同时考虑单词的左右上下文，而不是像传统模型那样按顺序处理文本。这种双向方法使得 BERT 在理解语言时能够提供更丰富的语境信息。	模型实现实体抽取、关系抽取、事件抽取、情感分析等任务的统一建模，并使得不同任务间具备良好的迁移和泛化能力。UIE 模型支持不限定行业领域和抽取目标的关键信息抽取，实现零样本快速冷启动，并具备优秀的小样本微调能力，快速适配特定的抽取目标。

除此之外，期货行业中的咨询信息标准化程度较低，整体内容相对混杂，难以统一采集，且难以在短时间内进行归类分析。本项目主要使用前文提到的大模型技术对已有的财经新闻、分析报告、社交媒体帖子、网站短讯等咨询文本进行实体识别、文本分类、时间抽取等分词处理，对新的数据内容只是点和新的数据源进行内容的定向采集，从而极大地提升了数据采集的效率、质量及完整性，构建了全面完备的高质量行业数据库，为之后的风控体系提供了强有力的支持。

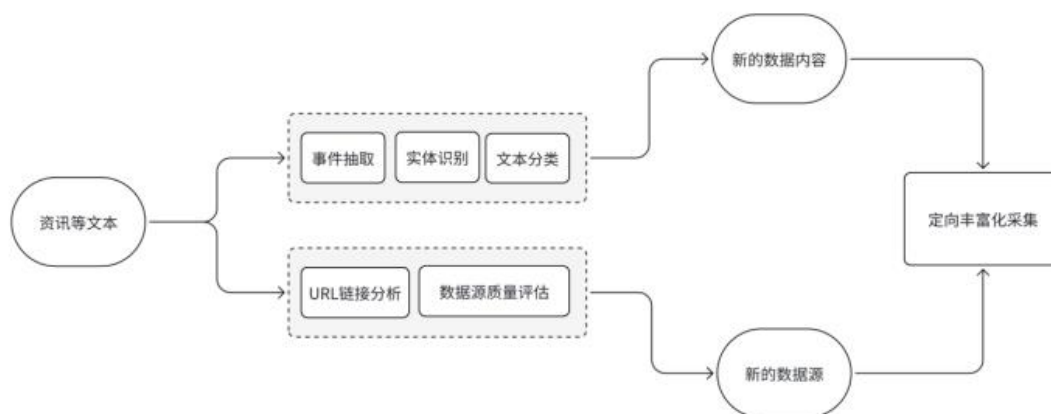


图 5 期货行业智能风控系统信息采集模块流程图

另外，由于选取的数据量较大且渠道众多，数据容易出现重复数据或者其他问题数据，尤其在一些敏感内容上，需要不断的对大模型进行微调而达到规避违规信息的目的。我们运用 UDA 无监督数据增强框架对少样本训练数据进行增强扩展，并在增强扩展后的数据基础上微调训练 BERT 低质量文本分类模型，该模型通过学习大量的高质量和低质量文本数据，过滤涉及政治、色情、暴力、赌博、攻击等敏感内容。最后，为了尽可能避免重复数据，主要通过 MinHash、SimHash 等算法进行去除。清洗后的数据导入数据库。

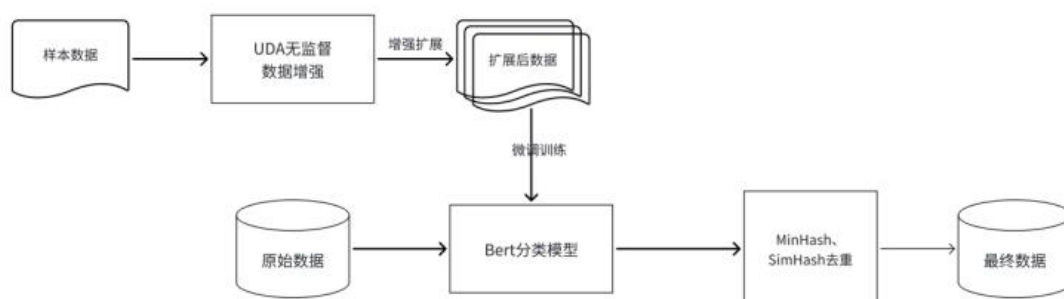


图 6 UDA 无监督数据增强框架流程图

最后，为了减少数据查询的错误，提高查询效率，我们在数据查询模块使用 NL2SQL 技术。NL2SQL 是一项将文字语句转为可执行 SQL 语句的技术，其本质是将自然语言语句转化为计算机可以理解的规范语义表示，是语义分析领域的一个子任务。基于大语言模型的 NL2SQL 能将信息咨询及社交网络上的文字信息高精度转化为数据库 SQL 查询语句，充当数据库的智能接口，实现期货风控业务场景中的数据高效输出和查询。其与传统数据查询的对比图如下。

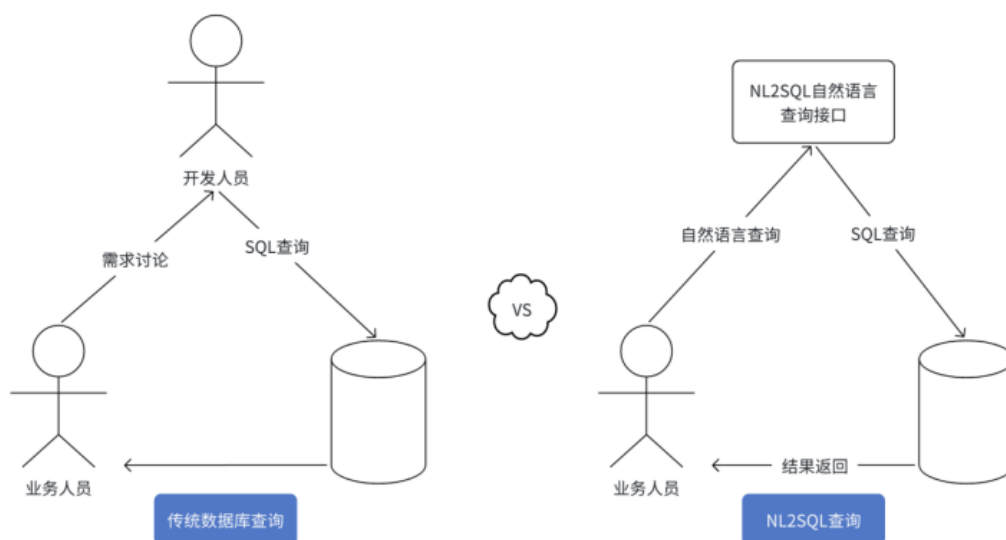


图 7 NL2SQL 查询 vs 传统数据库查询

（三）大语言模型在舆情分析与智能风控层面应用案例介绍

在本小节中，我们将展示利用大语言模型生成的舆情及风控的具体案例，具体流程如图 8 所示。

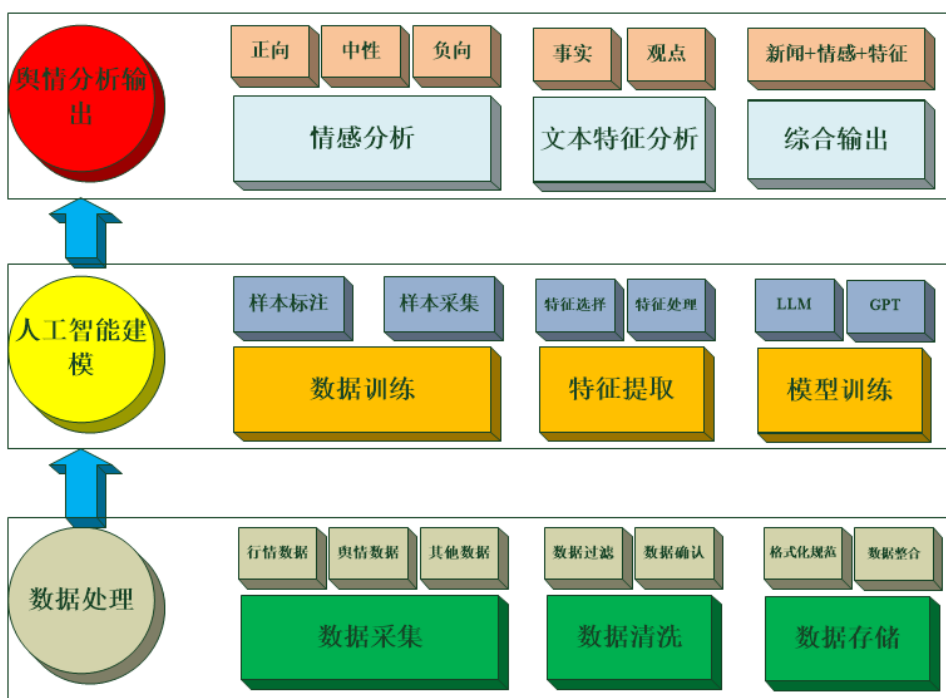


图 8 舆情实时监控场景流程图

首先，我们从新闻网站、API、社交媒体或 RSS 等数据源中提取新闻内容。通过编写爬虫工具或使用第三方新闻数据 API 进行数据收集。并且基于大模型的数据自动化丰富采集技术，对原始数据中的新闻字段（如发布时间、标题、内容等）提取并标准化存储在数据表中。通过 NL2SQL 技术自动生成 SQL 语言存入 MySQL 数据库。通过无监督模式的数据清洗技术去除无效数据和重复条目，

去除敏感词汇新闻，去除错误数据与空白数据。

清洗数据后，为了提升最终的模型的结果准确度。我们先对各个模型的准确度进行了初步测试。在原始数据中我们先抽取了 1000 条作为测试集，由于测试集中没有准确的标签，所以通过人工手动标签的方式制作对比测试结果。通过在不同的模型下拟合这些数据，并进行结果对照，结果如下表所示。

表 3 不同模型准确率对比展示

NLP 任务	模型	准确率	测试集	测试集规模
期货品种实	Bert+CRF	80%	句子级的实体识别测试集	1000 条
体识别	ChatGLM3	83%	句子级的实体识别测试集	1000 条
资讯文本分	Bert+UDA	79%	篇章级的文本分类测试集	1000 条
类	ChatGLM3	80%	篇章级的文本分类测试集	1000 条
产业链事件	UIE	91%	句段级的事件抽取测试集	1000 条
抽取				
事件因果关	UIE	78%	句段级的事件抽取测试集	1000 条
系识别				

从表中可以看出，在进行微调以后的 ChatGLM 模型整体的正确率会略高于 BERT 模型，这可能得益于 ChatGLM 在中文环境下的良好表现。UIE 模型在产业链名词抽取中有不错的表现，但因果关系识别上表现不佳。在后续的测试中，我们将使用 ChatGLM3 模型制作的标签进行进一步的风险管理测试。

针对预处理完成后的数据，我们使用 ChatGLM3 模型识别出新闻内容中的实体，例如人名、地点、组织等，并分别存入相应的列中，在表格中一共存储了以下这些标签：publish_time: 发布时间，event_topic: 事件核心，content: 文本内容，place: 地点，people: 人物，organization: 组织，country: 国家。使用情感分析工具，基于新闻内容计算出文本的情感倾向（正面、负面、中性），并填入“sentiment”列。根据新闻内容的主题，使用文本分类模型为每条新闻打上标签，填入“event_topic”和“content_tag”等列。并且运用语言模型判断其内容为观点还是事件描述。根据新闻中提到的地点，将对应的国家和区域信息标注在“country”和“place”列。整体标注效果可见图。

在本案例中，我们使用了金融新闻咨询 300+ 万篇、期货行业教材类书籍 700+ 本、结构化数据规模 4TB 以上。模型端的参数量为 ChatGLM3-6B 60 亿参数，文本摘要模型 PEGASUS-Large5.7 亿左右。平均日均生成期货咨询 3000+ 篇、产业链时间 300+ 条。

[illegible]

图 9 基于大模型的舆情信息标注结果

在上述舆情数据的基础上，对于有特定商品（在期货交易所上市的品种）相关讯息的，我们进行了进一步的自动化分类和整合，通过大模型将同一品种同一天信息按照市场情绪、宏观、供需风险等标签进行汇总，如图。整合信息后存储，再通过输出端传输给客户，整个过程实现完全自动化。

2024-08-26	上证50股指期货	HO	2024-08-26上证50股指期货期权新闻汇总	<p>市场预期与情绪</p> <p>1. 上证50股指期货品种的波动率上升。</p> <p>2. 市场对上证50股指期货的波动性预期增加。</p> <p>3. 投资者对上证50股指期货品种的风险厌恶情绪上升。</p> <p>宏观经济与外部因素</p> <p>1. 宏观经济政策变化可能影响上证50股指期货品种的价格。</p> <p>2. 上证50指数期权品种的隐含波动率上升。</p> <p>市场预期与情绪</p> <p>1. 尽管供需基本面压力存在，但成本支撑限制行情调整空间，部分业者对“金九银十”持乐观心态。</p> <p>2. 不锈钢期货今日偏强运行，现货市场情绪好转，304冷轧成交活跃。</p> <p>3. 不锈钢现货市场成交好转，低价货源减少，平盘出货压力存在。</p> <p>4. 腾达科技2024年上半年净利润395.68万元，同比下降11.94%，主要受不锈钢价格波动影响。</p> <p>5. 国内市场不锈钢价格波动较大，导致腾达科技产品售价价格下降，营业收入减少。</p>
2024-08-26	不锈钢	SS	2024-08-26不锈钢新闻汇总	<p>宏观经济与外部因素</p> <p>1. 镍铁供需双价价差拉大，博弈加剧；镍矿供需偏紧，矿价高企；不锈钢价格走低，钢厂加大压力力度。</p> <p>2. 8月份无锡地区大约304冷卷价格下跌50元/吨，跌幅0.35%，同比下跌9.72%，市场交投清淡。</p> <p>3. 镍生铁价格回落导致冶炼厂利润承压，生产驱动力受阻；不锈钢现货价格低位，库存高企，钢厂采购持观望态度。</p> <p>4. 商务部决定对原产于欧盟、英国、韩国和印度尼西亚的不锈钢锅还和不锈钢热轧板/卷进行期终复审调查。</p> <p>5. 本周不锈钢库存增量明显，主要由于到货增多和市场需求谨慎，成交表现平淡。</p> <p>6. 本周市场到货正常，现货行情震荡，交投氛围平静，下游刚需采购，导致不锈钢社会库存窄幅增量。</p> <p>7. 商务部决定对原产于欧盟、英国、韩国和印度尼西亚的不锈钢锅还和不锈钢热轧板/卷的反倾销措施进行期终复审调查。</p> <p>8. 商务部发布2024年第30号公告，对不锈钢锅还和不锈钢热轧板/卷的反倾销措施进行期终复审调查，并发送调查问卷。</p> <p>9. 商务部对欧盟、英国、韩国和印度尼西亚原产的不锈钢锅还和不锈钢热轧板/卷实施反倾销措施的期终复审调查。</p> <p>10. 商务部已发放不锈钢锅还和不锈钢热轧板/卷反倾销措施期终复审调查问卷，调查自7月22日开始。</p> <p>11. 宏观回调但基本面疲软，镍不锈钢价格震荡运行。</p> <p>12. 受宏观回调影响，镍不锈钢价格出现震荡，但基本面疲软。</p> <p>13. 镍、铬和铁矿石等大宗原料价格跳水，导致不锈钢盘条价格大幅下降，影响公司营业收入。</p> <p>14. 近两月国内市场不锈钢价格波动较大，特别是镍、铬和铁矿石等大宗原料价格持续跳水，导致不锈钢盘条价格大幅下降。</p> <p>15. 不锈钢行情持续低迷，钢厂开工减弱，废铁回收指导价下跌。</p>
2024-08-26	乙二醇	EG	2024-08-26乙二醇新闻汇总	<p>市场预期与情绪</p> <p>1. 检修装置增多，国内供应缩量，需求端支撑稳定，市场预期弱势反弹，关注国际原油及港口库存变化。</p> <p>2. 检修装置增多导致供应缩量，需求端刚需支撑稳定，港口去库预期，市场缺乏持续性驱动。</p> <p>3. 外围不稳定因素多，市场缺乏持续性驱动，关注国际原油及港口库存变化。</p> <p>4. 商品整体情绪回暖，短期国内乙二醇市场偏强震荡为主。</p> <p>5. 国际油价反弹，成本端支撑仍在，乙二醇市场短期偏强震荡。</p> <p>6. 国际油价反弹，成本端支撑仍在，乙二醇市场短期偏强震荡。</p> <p>供应风险</p> <p>1. 检修装置增多导致供应缩量。</p> <p>2. 连云港—乙二醇企业因上游原料短缺，乙二醇装置负荷下降至5-6成。</p> <p>成本风险</p> <p>1. 国际油价持续下降，成本端支撑薄弱，主港有阶段性累库预期，但供需维持平衡，需求端有走强预期。</p> <p>2. 华东主港库存大幅累积，下游需求端驱动不足，采购积极性不佳，但国际油价反弹，成本端支撑仍在，供需维持平衡。</p> <p>价格风险</p> <p>1. 张家港乙二醇现货价格上涨至4646元/吨，预计华东现货价格商谈区间在4550-4700元/吨。</p> <p>2. 主力合约多数上涨，碳酸锂涨幅最大，超过3%。</p> <p>市场预期与情绪</p> <p>1. 预计9月份内贸用180C或有望试探回暖表现，受国内经济持续修复和传统消费旺季提振。</p> <p>2. 月末未采需求提振燃料油价格，低硫燃料油表现强于高硫燃料油。</p> <p>3. 燃料油月末未采需求增加，低硫燃料油价格相对较高。</p>
2024-08-26	低硫燃料油	LU	2024-08-26低硫燃料油新闻汇总	<p>供需状况</p> <p>1. 终端船运需求清淡，船东购进谨慎，小单成交；供应方面，市场资源供应量或处于中低水平。</p> <p>2. 发电需求逐步走出旺季，导致高硫燃料油市场边际宽松。</p> <p>3. 新加坡燃料油库存上涨至两周高点，轻馏分油库存达到11周高点，中馏分油库存下降。</p> <p>4. 原油系期货跌势持续，低硫燃料油、SC原油、燃料油主力合约均跌超2%，市场忽视OIA库存报告利好，供需面未看到有效支撑。</p> <p>价格变动</p> <p>1. 国内期货主力合约多数跌多，碳酸锂、集运欧线、铁矿石等品种涨幅较大，低硫燃料油、氯化铝等品种跌幅较大。</p> <p>2. 多数商品品种收盘上涨，集运欧线涨幅最大，碳酸锂、铁矿石等品种涨幅超过4%。</p>

图 10 基于大模型的新闻信息整合

（四）基于大语言模型生成信息的收益与波动率分析

上述案例给出的风控信息整合结果可能过于繁多，且用户可能无法直接落地到交易层面。为了解决这个问题，在本节中，我们尝试利用大模型生成的信息来编写一些落地到交易层面的策略，以期在未来收益分析及波动率分析两个角度给用户带来新的风控启发。

首先，为了测试本系统构建的情感色彩的准确性及预测性。我们摘取了 2024 年 7 月至 9 月基于该系统生成的数据进行分析。由于盘中的信息错综复杂，且及时性较强，在此，我们选定未来的 5 分钟收益率作为预测目标。

具体做法为根据系统评定的情感色彩来预测该商品未来 5 分钟的收益率，希望得到的结果是情感色彩与未来收益呈正相关关系。由于各家财经机构给出的新闻舆情信息可能会存在重合，因此在进行本部分实证前，我们使用结巴分词进行去重，保存重合度在 60%以下的新闻资讯进行相关性分析及收益分析。

图 11、图 12 分别展示了各品种实时的正面及负面新闻出现的次数及胜率的统计情况。其中系统判定实时信息为正面情感后，未来 5 分钟收益上涨的概率是 60.86%；当系统判定实时信息为负面情感后，未来 5 分钟收益下降的概率是 52.17%。

从结果来看，商品未来的收益率与舆情的色彩不一定呈现明显的相关关系，其中部分品种的胜率较低，这一点可能跟商品本身的价格影响因素较多及舆情信息存在滞后性等原因有关，但如果做多个商品舆情策略的投资者可以参考该系统给出的判断，如果是从某一个品种出发，该系统提供的信息只能作为其中一部分因素来判断，投资者还需要结合基本面的信息来整体把控未来的行情。

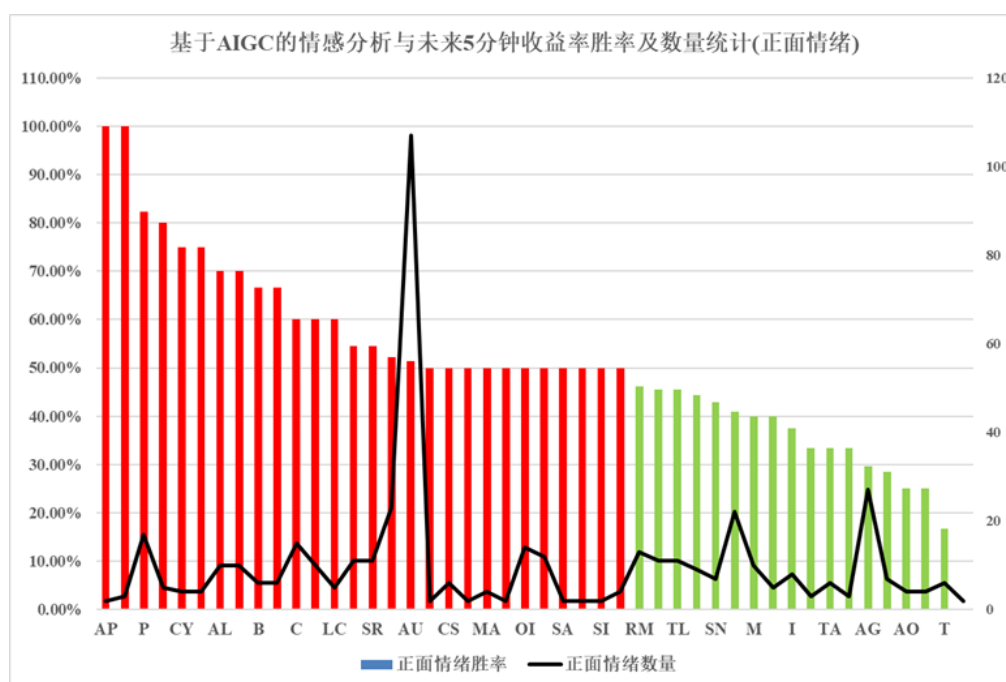


图 11 基于 AIGC 的行情情感分析与收益率关系统计(正面)

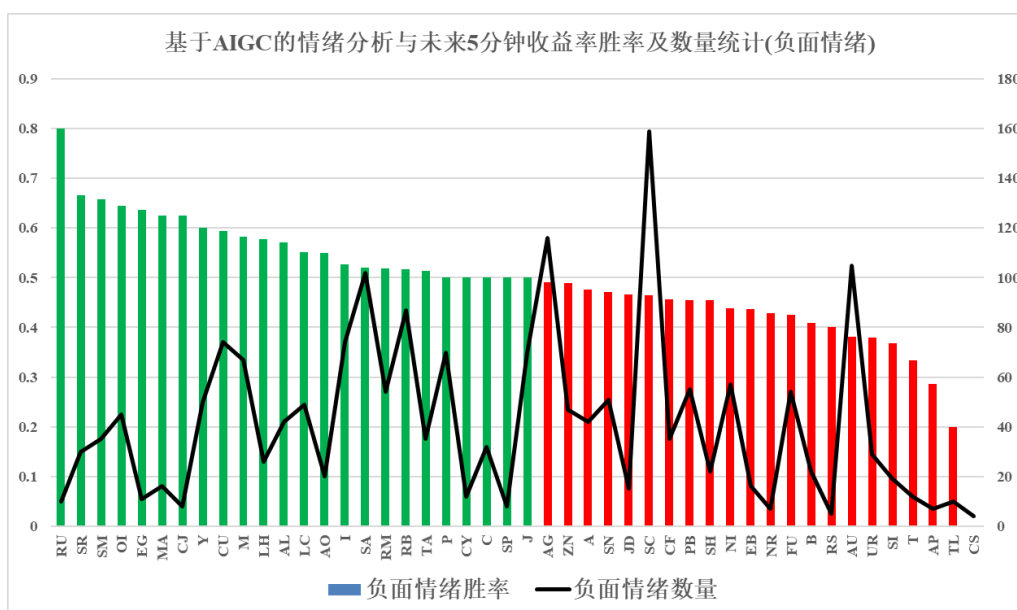


图 12 基于 AIGC 的行情情感分析与收益率关系统计(负面)

另外，对于一些有较大套保头寸的机构或者客户，相较于盘中舆情信息带来的日内波动外，可能他们更关心的是期货未来几天的波动率水平的变化，如果能够根据盘后的汇总信息来预测未来商品可能会出现波动率变化情况，这一点可能会便于使用期权做套期保值的用户及时调整自己的套保头寸，以期获得波动率方向上的收益。

在该部分，我们测试了不同商品日度的舆情信息数目变化率与下一个交易日波动率变化率的相关性，希望得到的结果是如果该商品出现的信息增加，未来该品种的波动率会放大。但具体测试的结果显示并不是所有的商品都是正相关关系。但从所有商品的胜率来看，正相关的品种占比接近 60%，这一点可能跟筛选的信息本身的质量以及波动率本身较难预测有关。从交易层面而言，如果是一个做多商品波动率套利的交易者，该胜率还是有一定的参考意义的。图 13 展示了其中按照相关系数过筛后的品种的相关系数及品种出现天数的统计结果，其中正相关的品种为 19 个，负相关的品种为 13 个，相关性胜率为 59.375%。

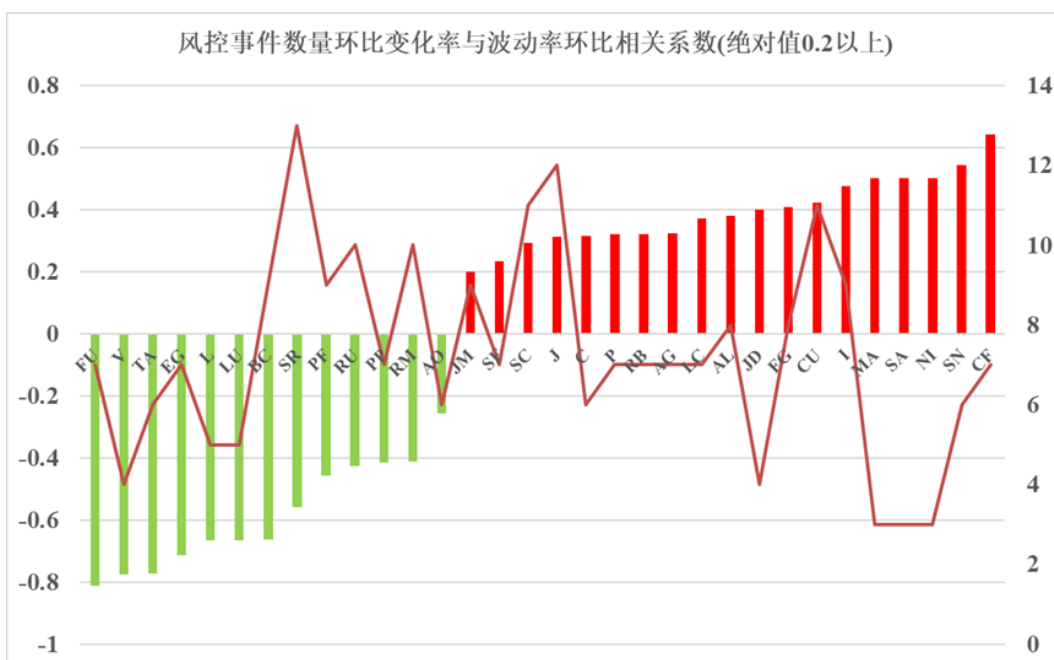


图 13 风控事件数量环比变化率与波动率环比相关系数及品种出现天数结果

总体来看，正向舆情信息出现的次数越多，未来波动率上涨的概率越高。如果在客户在盘面有较多的套期保值头寸，可以参考该信息系统及当前的行情来合理的设置自己的套保头寸，比如将部分期货头寸用组合期权或者场外期权的形式来代替，获取波动率方向上的盈利，避免较大的波动。

（五）大语言模型在舆情分析与智能风控层面总结与展望

在本节中，我们重点介绍了大语言模型在舆情分析与智能风控层面的应用案例。我们利用财经网站的数据进行大语言模型标注分类，同时进行情感分析，及时输出到用户端。在此基础上，我们还对日内舆情信息及日间新闻汇总分别进行了收益率及波动率分析，以期获得可以落实到交易层面的策略。该模块目前已经落地，套保模块风控如图 14 所示。

在此基础上，还有部分问题尚未解决。

首先需要解决的是数据内容安全问题，由于在后续模型进行预训练过程中，由于数据量庞大，且缺乏人工审核，在实时数据录入过程中很容易出现敏感词或者违规信息，如果这部分输出给终端客户将给期货公司带来极大风险，所以在未来需要继续微调模型过滤问题数据，同时建立一个维护数据内容的框架。另外在实时分析数据的过程中，同样要关注大量数据高速传输的网络传输安全性，实时采集和实时输出将会对系统和网络有较高的要求。图 15 为一个简单的基于大模型的期货行业平台安全架构。

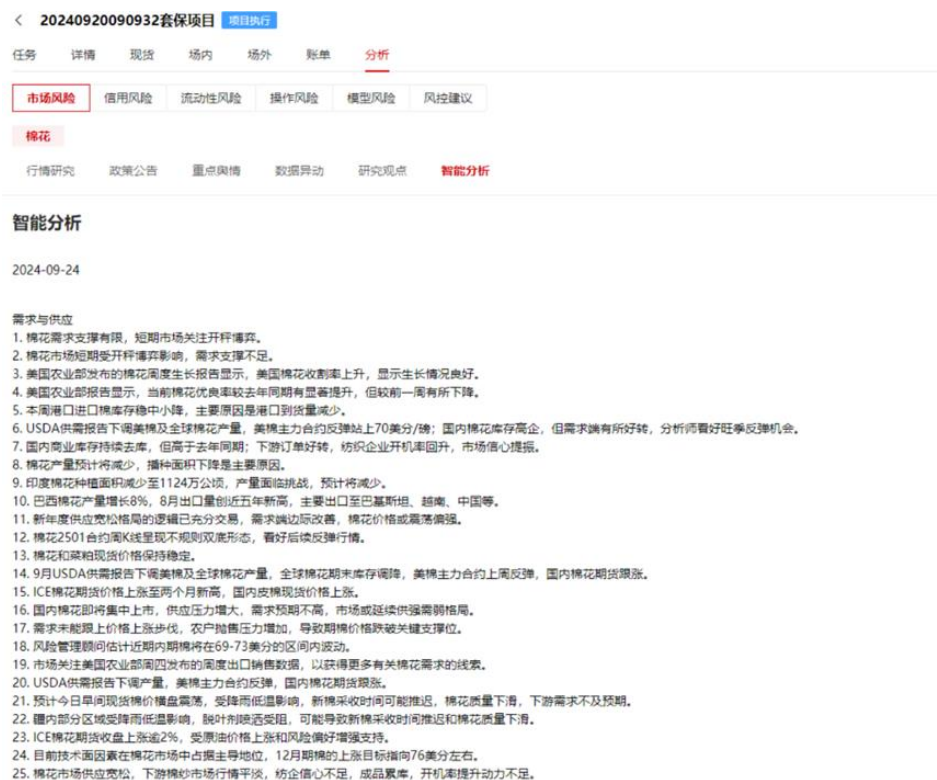


图 14 棉花期货套保模块智能分析图示



图 15 基于大模型的期货行业平台安全架构

另外，我们可以展望一下未来舆情层面的智能风控板块。假设你是一个客户，首先我们会区分你是投机户还是套保户，根据不同的需求，我们会匹配不同的智能风控模型。对于投机者，我们希望根据盘中的舆情信息给交易者提供未来收益率的分析，让交易者更敏锐的洞察该信息背后存在的投资价值，同时还可以增加相似行情历史分析模块，比如盘中出现了某个价格或者库存变化的及时性消息，我们可以通过快速回溯匹配历史上的相似事件，给出未来行情可能出现的各种概率分析以及风险分析；对于套保用户，我们希望根据盘后信息的汇总给出未来行

情的分析及波动率的分析，实时调整推荐的套保头寸，比如预期未来波动率会降低，可以使用卖权的方式代替期货单边头寸。

综上所述，本研究通过大语言模型和自动化数据处理技术，显著提升了市场信息捕捉和舆情分析的效率，构建了一个自动化、精确的风险预警系统。该系统满足了期货行业在复杂市场环境下对高效风险管理的需求，展示了 AI 技术在期货风控中的广泛应用潜力。

五、基于 AGI 的关联交易识别框架

（一）基于 AGI 的关联交易识别框架的应用背景

期货交易涉及大额资金流动，容易成为非法资金的洗钱渠道。识别关联交易是反洗钱检测中的核心环节：关联账户间的频繁交易和复杂资金路径往往是洗钱行为的典型特征，能够有效揭示团伙协作的洗钱模式，帮助锁定高风险账户群体。

当前期货公司相关部门主要通过 KYC、异常交易监控、跨境交易审查等手段履行反洗钱职责。然而，这些手段在识别复杂的关联交易时存在局限性，难以有效应对多层次、跨账户的洗钱模式。我们设计的框架通过多算法协同分析（如节点中心性、社区检测、孤立森林等），能够精准识别关联交易网络中的高风险账户及其交互模式，并通过实时数据整合与可视化工具，为稽核人员提供直观、动态的风险监控视角。该设计能高效识别团体协作洗钱风险。

（二）基于 AGI 的关联交易识别框架的具体结构

基于 AGI 的关联交易识别框架旨在通过智能化的图分析技术，提升对期货市场中潜在洗钱行为的监控与识别能力。该框架以多层次模块化设计为基础，涵盖数据收集、储存、分析、可视化四大模块，实现对复杂资金流动和多账户操作的全面监控与分析。

首先，数据收集与整合模块负责从不同来源获取多维数据，包括账户信息、交易记录、资金流动及外部风险数据，并将其标准化和整合，以确保后续分析的准确性和完整性。接着，图数据库构建模块利用图数据库技术，将收集到的多维数据转化为图结构，以直观展示账户及其交易行为的关系网络。

随后，图分析算法模块通过应用多种图算法，如社区检测、节点中心性分析等，识别多账户间的潜在关联和异常交易模式。

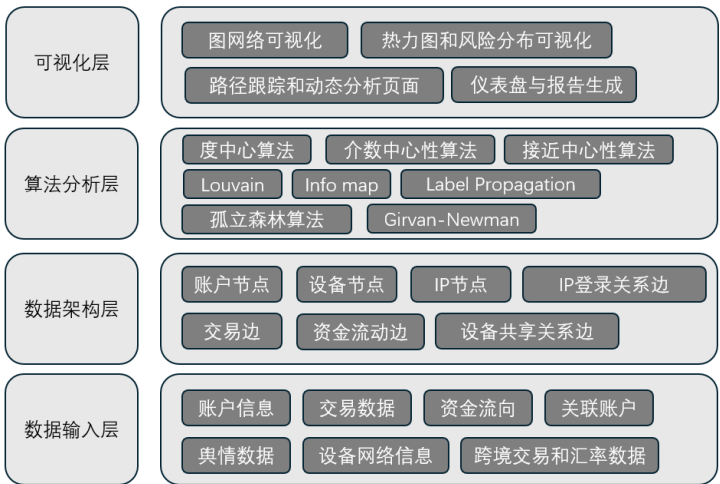


图 16 基于 AGI 的关联交易识别框架

最后，为了提高用户对分析结果的理解和应用，框架引入了可视化与用户界

面模块，以交互式图谱呈现关联账户和资金流动路径，帮助用户直观识别高风险账户。整体项目结构如图 16 所示。下面我们将针对其中的具体模块进行展开。

1. 基于 AGI 的关联交易识别框架——数据输入层

第一模块也是最基础的模块是数据输入层，其主要目的是从多个数据源中提取、整合和标准化各种信息，以支持后续的图数据库构建和分析。该模块的核心在于确保数据的全面性和准确性，从而为识别潜在的关联账户和异常交易模式奠定坚实的基础。首先，在账户信息层面，该模块会收集客户的基本信息、注册信息和联系方式等。这些信息有助于建立账户的唯一标识，并分析账户间的共性和潜在的关联关系。其次，在交易数据方面，模块会提取所有交易的时间、金额、交易类型和对手信息。这些数据有助于分析资金的流动路径和交易行为的模式。例如，高频交易、分拆交易和跨市场交易行为，往往会通过交易频率和金额的分布表现出来。该模块还涵盖资金流动、跨境交易和关联账户的追踪数据。这些数据能够揭示资金来源与去向、跨境资金的转换模式，以及多账户之间的频繁交易情况。此外，外部数据源和舆情数据的整合也是该模块的关键任务。负面名单、法律记录和公开报道等数据源能够补充内部数据的不足，提供更多的背景信息。下图展示了具体的数据收集目录。

待搜集数据和搜集目的		
信息大类	信息项	搜集目的
账户信息数据	账户ID	用于唯一标识客户账户并区分不同客户。
	账户类型	识别账户的属性和类型，如个人、企业、离岸等。
	开户时间	分析账户的活跃周期及关联账户的时间关系。
交易数据	交易ID	用于追踪资金流动路径，识别交易模式。
	交易时间	分析交易的时间分布和频率，识别高频交易。
	交易金额	识别大额交易和分拆交易，分析可疑资金流动。
资金流动数据	资金来源账户ID	识别资金的最初来源和渠道，进行合法性评估。
	资金去向账户ID	识别资金的最终去向和汇聚模式，进行追踪分析。
	资金流动路径	分析资金在不同账户间的流动顺序，识别异常路径。
跨境交易与汇率数据	跨境交易标识	监控跨境资金流动，判断是否符合合法要求。
	交易汇率	分析跨境资金转换的合理性，识别异常汇率使用。
关联账户数据	同一IP地址/设备登录账户	识别多个账户的潜在关联，尤其是在同一设备上登录的账户。
	相似交易模式的账户	识别交易模式相似的账户群体，标记潜在的关联账户。
	账户间资金往来频繁	标记账户之间的频繁资金往来，识别洗钱行为。
外部数据源与舆情数据	负面名单	识别账户或交易对手的高风险背景，提高风险预警。
	媒体报道与舆情信息	识别客户或公司在公开报道中的信息，了解潜在风险。
	公共数据库信息	核查账户的法律背景，如法院记录、法律纠纷等。
设备与网络信息	设备指纹	识别账户的设备特征，分析设备共享情况。
	网络信息	分析账户间的网络连接关系，识别潜在关联。

图 17 基于 AGI 的关联交易识别框架数据收集目录

2. 基于 AGI 的关联交易识别框架——数据储存层

收集完数据后我们需要单独的储存模块来保存数据。该模块的设计旨在通过图数据库的高效性和灵活性，实现多账户和多路径的快速查询与分析。借助 Neo4j

的图模型能力，可以有效支持复杂的路径搜索、节点关系分析和社区检测等高级图计算功能，从而提高对潜在洗钱行为的识别精度。

在数据储存模块，我们首先介绍图数据库的基本概念和优势；其次，我们将通过实例演示如何将收集的数据导入 Neo4j 数据库；最后，我们将借助简单的 query 对 Neo4j 数据库的查询功能做演示。图数据库核心概念是将数据看作一个图，图中的节点代表实体（如账户、交易、设备），边则表示这些实体之间的关系（如交易、资金流动、设备共享等）。Neo4j 是目前最流行的图数据库之一。它储存、计算效率高，查询能力灵活，在处理大量数据时有独特的优势。合理的数据模型能使查询事半功倍。

结合数据输入层的数据，我们计划采用以下结构存储数据。为了方便数据库进行捕捉，我们还需要做一些定义，包括节点定义及边定义。

其中节点定义包括以下几个部分：

- A. 账户节点 (Account)：代表客户账户，并进一步细化账户的多维属性，包括：
 - a) `account_id`：账户唯一标识符。
 - b) `account_type`：账户类型（如个人账户、企业账户、离岸账户）。
 - c) `registration_date`：账户注册日期。
 - d) `risk_level`：账户风险等级（如“低风险”、“中风险”、“高风险”）。
 - e) `owner_info`：账户持有人信息（如姓名、年龄、国籍等）。
 - f) `status`：账户状态（如“活跃”、“冻结”、“关闭”）。
 - g) `negative_flag`：是否在负面名单中（如“是”或“否”）。
 - h) `media_score`：舆情评分（根据外部舆情分析结果给出的风险分数）。
- B. 设备节点 (Device)：记录账户使用的设备信息，包含更多动态属性，如：
 - a) `device_id`：设备唯一标识符。
 - b) `type`：设备类型（如手机、电脑、平板）。
 - c) `location`：设备的地理位置。
 - d) `usage_frequency`：设备的使用频率（如登录次数、活跃时间）。
- C. IP 地址节点 (IP)：记录账户的 IP 登录信息，并进一步加入动态属性
 - a) `ip_address`：IP 地址。
 - b) `location`：IP 地址的地理位置。
 - c) `change_frequency`：IP 变更频率（记录 IP 的动态变化频次）。

边定义包括以下几个部分：

- D. 交易边 (TRANSACTION)：在账户间创建交易关系，细化属性以支持多级交易识别：
 - a) `transaction_id`：交易唯一标识符。
 - b) `amount`：交易金额。
 - c) `date`：交易日期。

- d) **type**: 交易类型（如买入、卖出、对冲）。
- e) **strategy**: 交易策略（如套利、高频交易、对冲等）。
- f) **market_level**: 交易的市场级别（如一级市场、二级市场）。
- E. **资金流动边 (FUNDS_FLOW)**: 描述资金在账户间的具体流向，进一步增加资金的来源和去向属性：
 - a) **funds_id**: 资金流动唯一标识符。
 - b) **amount**: 资金流动金额。
 - c) **source_type**: 资金来源类型（如工资收入、投资收益、贷款等）。
 - d) **destination_type**: 资金去向类型（如个人账户转账、企业账户入账等）。
 - e) **date**: 资金流动日期。
- F. **设备共享关系边 (USES)**: 连接账户和设备节点，描述设备的共享情况：
- G. **IP 登录关系边 (LOGGED_IN_FROM)**: 描述账户的 IP 登录情况，包含 IP 的变动属性：
 - a) **login_time**: 登录时间。
 - b) **change_detected**: 是否在短时间内发生 IP 变动（如“是”或“否”）。

节点	边	关系类型
账户节点	交易边 (TRANSACTION)	表示两个账户之间的交易行为，包括金额、类型、策略等
账户节点	资金流动边 (FUNDS_FLOW)	表示资金在账户间的流向，包括金额、来源和去向
账户节点	设备共享关系边 (USES)	描述账户使用的设备情况，识别多账户共用设备的风险
账户节点	IP登录关系边 (LOGGED_IN_FROM)	记录账户的IP登录情况，标记潜在的IP共享和变动情况
账户节点	交易边 + 资金流动边	通过结合交易和资金流动边，分析账户间的资金和交易链条
设备节点	设备共享关系边	分析同一设备的多账户登录情况，揭示潜在的关联账户
设备节点	设备共享关系边 + IP登录关系边	通过设备和IP的关系，标记多账户可能的设备共享和IP交叉使用
IP节点	IP登录关系边	分析单一或多账户的IP登录情况，评估IP的风险性
IP节点	IP登录关系边 + 资金流动边	通过IP和资金流动的结合，揭示跨账户的潜在洗钱路径

图 18 账户节点关系类型

其中图 18 展示了各个节点的关系类型。为了更详细的阐述该模块的功能，最后我们用代码做个简单展示。我们选取交易关系作为示例。演示上述设计如何在 Neo4j 中实现。假设我们有两个 CSV 文件：

- a) 账户数据文件 (accounts.csv)，示例如下：

account_id	account_type	registration_date	risk_level	owner_info	status	negative_flag	media_score
12345	个人	2021-05-10	中风险	张三	活跃	否	60
67890	企业	2020-07-15	高风险	李四	活跃	是	80
82855	企业	2019-05-18	高风险	李四	冻结	否	61
51827	企业	2020-12-29	低风险	周七	禁用	否	75
84468	企业	2022-07-13	中风险	李四	冻结	是	90
45170	个人	2019-07-03	中风险	周七	活跃	否	85
44508	个人	2019-07-31	高风险	李四	活跃	是	96
60126	个人	2020-04-14	低风险	赵六	活跃	是	52
31997	企业	2019-07-03	高风险	钱八	禁用	否	41
86432	个人	2022-02-04	高风险	王五	冻结	否	60
58277	个人	2019-03-16	中风险	周七	活跃	是	20
66788	企业	2019-04-15	低风险	赵六	禁用	是	47
11111	企业	2021-03-12	高风险	周八	活跃	否	85
22222	个人	2021-11-25	高风险	钱九	活跃	是	90

图 19 Neo4j 匹配算法账户文件示例图

可以使用 Cypher 导入账户数据，将其存为 Account 节点：

```
LOAD CSV WITH HEADERS FROM 'file:///accounts.csv' AS row
MERGE (a:Account {account_id: row.account_id})
SET a.account_type = row.account_type,
    a.registration_date = row.registration_date,
    a.risk_level = row.risk_level,
    a.owner_info = row.owner_info,
    a.status = row.status,
    a.negative_flag = row.negative_flag,
    a.media_score = toInteger(row.media_score);
```

图 20 Cypher 导入账户数据代码

b) 交易数据文件 (transactions.csv)，示例如下：

sender_id	receiver_id	transaction_id	amount	date	type	strategy	market_level
12345	67890	tx001	10000	2023-08-15	买入	高频交易	二级市场
67890	12345	tx002	5000	2023-08-16	卖出	对冲	一级市场
21922	74843	tx003	1409	2021-12-14	卖出	长期投资	一级市场
86858	92202	tx004	5058	2021-05-28	买入	套利	场外交易
18256	65680	tx005	9441	2020-03-05	买入	高频交易	二级市场
34534	14282	tx006	5031	2021-10-20	买入	套利	一级市场
72925	73500	tx007	19384	2021-05-25	买入	套利	一级市场
87270	64126	tx008	6761	2021-03-02	卖出	套利	场外交易
98620	10161	tx009	2156	2022-10-10	卖出	高频交易	一级市场
88545	58623	tx010	16584	2020-11-29	卖出	对冲	场外交易
94232	10656	tx011	10205	2021-04-30	买入	长期投资	二级市场
27282	44598	tx012	2700	2022-02-17	卖出	长期投资	场外交易
11111	22222	tx_high_01	12000	2023-09-10	买入	高频交易	一级市场
22222	11111	tx_high_02	8000	2023-09-11	买入	对冲	二级市场

图 21 Neo4j 匹配算法交易文件示例图

若交易数据存为交易关系如下：

```
LOAD CSV WITH HEADERS FROM 'file:///transactions.csv' AS row
MATCH (sender:Account {account_id: row.sender_id})
MATCH (receiver:Account {account_id: row.receiver_id})
MERGE (sender)-[t:TRANSACTION {transaction_id: row.transaction_id}]->(receiver)
SET t.amount = toFloat(row.amount),
    t.date = row.date,
    t.type = row.type,
    t.strategy = row.strategy,
    t.market_level = row.market_level;
```

图 22 交易数据关系案例

我们可以通过以下语句查询所有高风险账户的买入交易：

```

MATCH (a:Account {risk_level: '高风险'})-[t:TRANSACTION {type: '买入'}]->(b:Account)
RETURN
    a.account_id AS 发送者,
    t.transaction_id AS 交易ID,
    t.amount AS 交易金额,
    t.date AS 日期,
    t.type AS 交易类型,
    t.strategy AS 策略,
    t.market_level AS 市场层级,
    b.account_id AS 接收者

```

图 23 高风险账户查询语句

最终，通过示例数据，我们共发现两笔高风险账户的买入交易，细节如下：

发送者	交易ID	交易金额	日期	交易类型	策略	市场层级	接收者
11111	tx_high 01	12000	2023-09-10	买入	高频交易	一级市场	22222
22222	tx_high 02	8000	2023-09-11	买入	对冲	二级市场	11111

图 24 Neo4j 查询高风险账户买入交易结果

3. 基于 AGI 的关联交易识别框架——图分析算法模块

首先，该项目计划使用度中心性算法来识别与多个账户有频繁交易联系的“枢纽账户”。这些账户通常是资金的集散地或频繁交易的节点，是反洗钱监控的重点对象。度中心性通过计算一个节点的直接连接数来衡量其重要性。在交易网络中，这类高频交易的账户可能是分拆交易或快速资金调度的一部分。

此外，项目计划采用介数中心性算法来发现充当“资金中介”的账户。这类账户在多个账户之间进行资金中转，是洗钱链条中的重要环节。介数中心性通过计算一个节点在最短路径上的出现频率，衡量其作为“桥梁”的能力。这类账户通常连接着多个资金来源和去向，承担着资金分散或汇聚的功能。介数中心性分析能够揭示洗钱路径中的中转账户，特别适用于多账户间复杂的资金路径分析。

另外，项目也会使用接近中心性算法，来识别资金网络中的“中心账户”。这些账户能够快速接触到其他账户，通常是资金汇聚或分散的枢纽。接近中心性通过计算一个节点到其他所有节点的平均最短路径长度，来衡量其在网络中的紧密程度。这类账户在洗钱网络中可能是控制账户，具备快速资金调度的能力。接近中心性分析有助于揭示具备快速资金分散能力的账户，并帮助锁定洗钱网络中的核心控制点。

在本项目中还计划使用社区检测算法被用于揭示交易网络中的账户集群，以识别多账户之间的紧密联系和协同行为。具体包括 Louvain 算法、Girvan-Newman 算法、Label Propagation 算法和 Infomap 算法。

Louvain 算法是本步骤的核心方法。Louvain 算法的核心在于**模块度 (modularity)**，其定义为在给定的划分下，实际的边数与随机情况下的期望边数之间的差值。模块度计算公式为

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

其中， A_{ij} 为 ij 之间的实际连接数， k_i 和 k_j 为节点的度数， m 为总边数， δ 为 indicator，表示两个节点是否属于同一个社区。

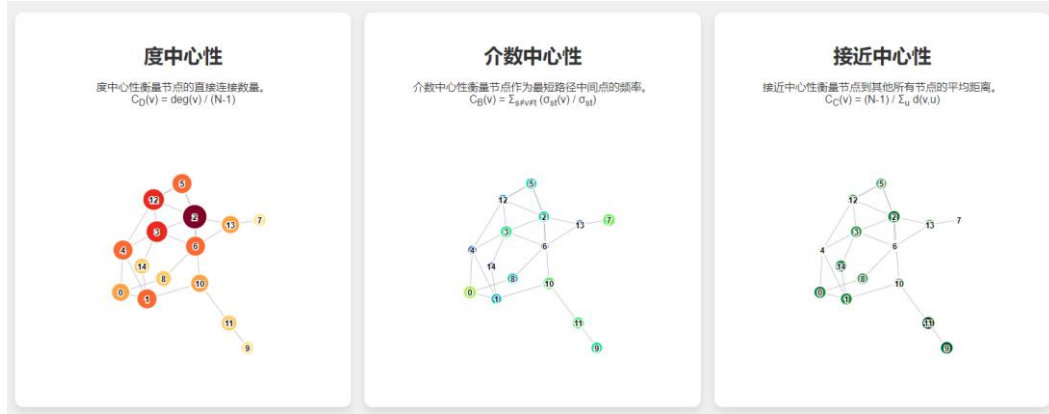


图 25 各中心性对比

Louvain 算法通过迭代的方式逐步优化模块度，分为两个主要阶段：

A. 局部优化阶段

- 初始化：**将每个节点视为一个独立的社区。
- 逐节点优化：**对于每个节点 iii ，尝试将其加入到邻居节点所在的社区中，然后计算模块度的增量。如果某个邻居社区的模块度增量最大且为正，就将节点 iii 加入该社区。
- 循环迭代：**重复上述步骤，直到没有节点可以被进一步移动，以提升模块度。

B. 社区聚合阶段

- 社区合并：**将第一阶段生成的社区视为超级节点，将超级节点之间的边合并为新的边，形成一个缩小后的新网络。
- 再次优化：**在这个新的网络上重复第一阶段的操作，继续优化模块度。
- 终止条件：**当模块度不再显著提升时，算法终止。

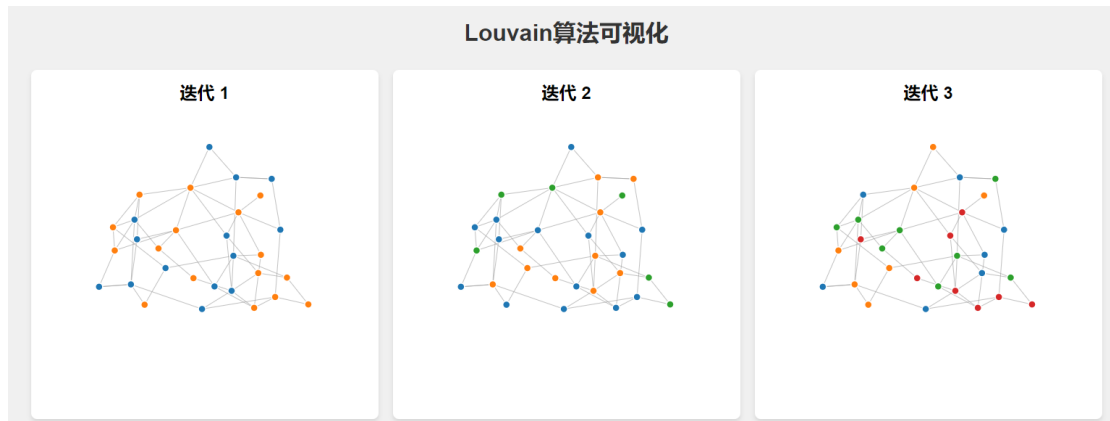


图 26 Louvain 算法可视化结果

在关联交易领域，群体可能涉及团伙协作进行洗钱。算法通过模块度最大化将这些潜在的团伙交易划分为多个社区，从而帮助反洗钱分析人员锁定风险更高的账户组合。

另外，项目还打算辅助采用 Girvan-Newma、Label Propagation 算法和 infomap 算法辅助划分社群。Girvan-Newman 通过逐步移除网络中的“桥梁”连接，将图分割成多个独立的社区。Label Propagation 该算法通过标签传播的方式，将图中的节点逐步划分到不同的社区中，适合实时交易网络的快速分析。Infomap 算法则通过模拟信息在网络中的流动，揭示具有相似信息传播路径的账户集群。该算法能揭示网络中的复杂资金路径，能够准确识别账户间的深层次关联。

最后，本项目还计划使用孤立森林算法检测交易网络中的异常账户和异常交易行为。孤立森林算法的核心原理是基于“孤立”的概念：与大部分节点不同，异常节点在图中通常具有不寻常的属性或行为，如交易频率过高、交易金额异常大或资金来源复杂。算法通过构建多个随机决策树来分割数据，并根据数据点被孤立的“深度”来评估其异常性。深度越小，意味着该节点越容易被孤立，也就是越可能是异常点。本项目的涉及到的算法总结如下图所示。

算法类型	主要算法	核心原理	应用目的
节点中心性算法	度中心性	通过计算节点的直接连接数，衡量其在网络中的重要性，度数越高的节点为频繁交易枢纽	识别频繁交易的枢纽账户
节点中心性算法	介数中心性	计算节点在最短路径上的出现频率，识别多账户间的中介节点，介数越高代表其中介角色越强	揭示洗钱链条中的中转账户
节点中心性算法	接近中心性	根据节点到其他所有节点的平均最短路径长度，接近中心性高的节点为快速接触其他账户的控制账户	定位快速资金汇聚或分散的控制账户
社区检测算法	Louvain	通过模块度优化，将图划分为多个紧密连接的社区，揭示账户间的群体性联系	发现多账户间的紧密集群
社区检测算法	Girvan-Newman	逐步移除边介数最高的边，分离图中的社区结构，揭示多账户间的交互路径和潜在边界	识别账户间的潜在交互路径和社区边界
社区检测算法	Label Propagation	通过标签传播的方式，节点根据邻居标签调整自身标签，逐步形成稳定的社区结构	检测高频协同交易的账户群体
社区检测算法	Infomap	模拟信息在网络中的流动，将信息传播路径相似的节点划分到同一社区，揭示多级洗钱路径	揭示资金的复杂分散和汇聚模式
孤立森林算法	孤立森林	通过随机分割数据，计算节点的孤立程度，离群点更可能是异常账户或交易行为	标记不符合常规模式的异常账户和交易

图 27 基于 AGI 的关联交易识别框架相关算法总结

4. 基于 AGI 的关联交易识别框架——可视化与用户界面模块

图网络可视化是整个用户界面的核心部分，用于动态呈现多账户间的复杂关联关系。可视化界面将账户节点、交易边、资金流动路径以及设备和 IP 的共享情况以直观的方式展示。不同类型的节点（如高风险账户、设备、IP 地址）可以用不同的颜色、形状或大小进行区分，而交易或资金流动的边则通过颜色深浅或粗细来体现交易强度和风险等级。用户可以通过点击节点查看账户的详细信息，并选定某条交易路径或资金流动进行进一步的分析。技术上可采用 D3.js、Sigma.js 等库进行实现。

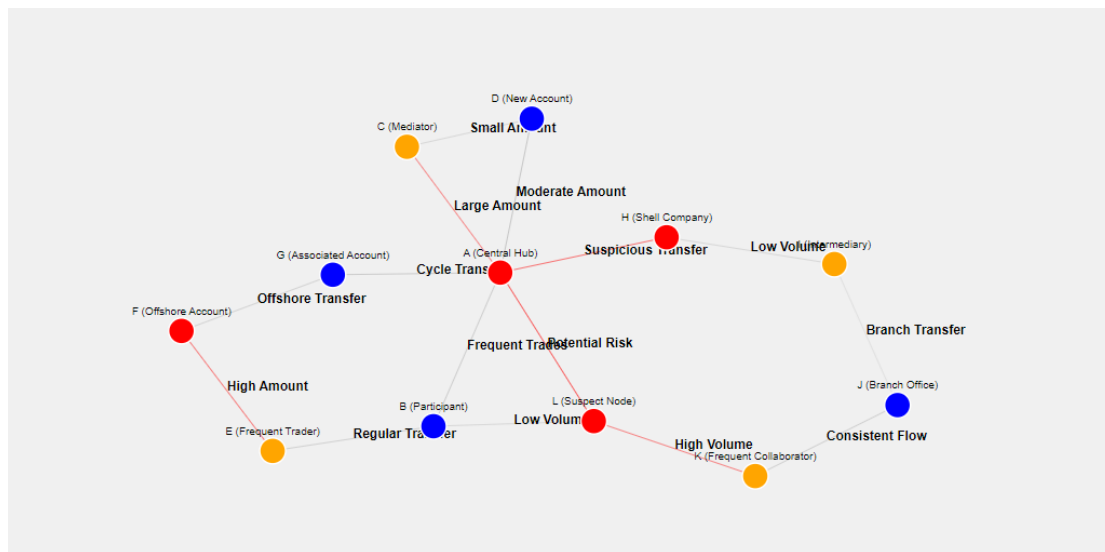


图 28 账户图谱示例图

在上图网络结构中，高风险节点通常表现为与多个节点有高金额交易的账户，并且扮演着资金的中枢角色。例如，节点 A 是一个典型的高风险节点，它与多个节点（如 H 和 L）有大额交易连接，且交易金额超过 30000。这类中心节点在洗钱网络中往往扮演着资金集散地的角色。此外，节点 F 作为离岸账户，与节点 E 和 G 有高金额的资金转移，这类离岸账户在洗钱中常被用于隐藏资金的实际来源或去向。高风险节点的特点是交易路径复杂、金额较大，且多为多级转账。相比之下，低风险节点的交易模式则较为简单、金额较小。例如，节点 B 与节点 A、E 和 L 的交易金额大多在 10000 以下，且交易路径不复杂，这类账户在网络中主要作为参与者角色。节点 D 作为新账户，交易金额较低且频率不高，符合常规的账户行为。低风险节点的典型特征是交易金额小、连接节点少，且大多没有复杂的多级资金流动路径。

另外，热图与风险可视化图也是重要的可视化方法。热图与风险分布可视化用于展示高风险账户或交易的集中区域，帮助用户直观了解风险的分布情况。通过热图形式，系统可以展示账户的地理分布情况，标记出高风险账户所在的区域。这有助于用户识别风险集中的地理区域，便于进一步的深入调查。

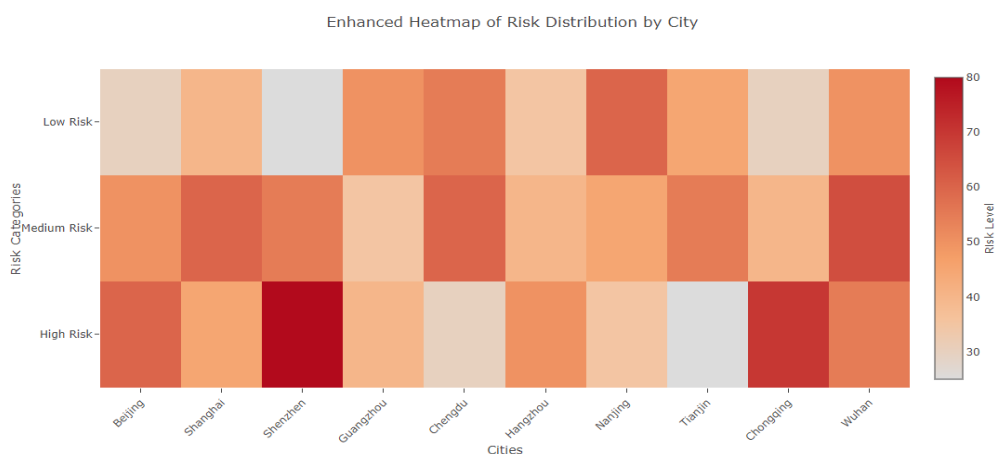


图 29 不同地区账户风险热力图

上图可以清晰地展示不同城市的风险分布情况。深圳和重庆的高风险账户数量分别为 80 和 70，是所有城市中最高的，这表明这两个城市的资金交易行为可能存在更高的洗钱风险，建议加强监控和预防措施。与此同时，成都和武汉的中等风险账户数量较多，分别为 60 和 65，这些城市的商业活动频繁，需要进行定期的交易模式分析，以防风险进一步升级。而在南京和广州，低风险账户数量相对较高，显示出较为稳定的交易环境。这些城市虽然整体风险较低，但仍需维持基本的合规检查，确保低风险账户不会突然转变为高风险账户。

除上述两种方式外，路径跟踪与动态分析界面也是用户实时探索和分析交易网络的关键工具。用户可以通过界面跟踪特定账户的资金流动路径，查看资金从源头到目标账户的整个流动过程，包括交易金额、时间和中转账户等详细信息。系统支持动态分析，允许用户设置自定义筛选条件（如资金流动方向、金额范围、时间段等），并实时生成符合条件的资金路径图。用户可以通过点击路径上的节点或边，获取该路径中每个账户或交易的详细信息，并标记可疑的中转账户。这种设计不仅便于用户进行深入的关联分析，还能帮助快速识别和锁定高风险的资金路径和关键账户。可采用 Dash（基于 Plotly）、Grafana 或 Kibana 等工具来实现该功能。

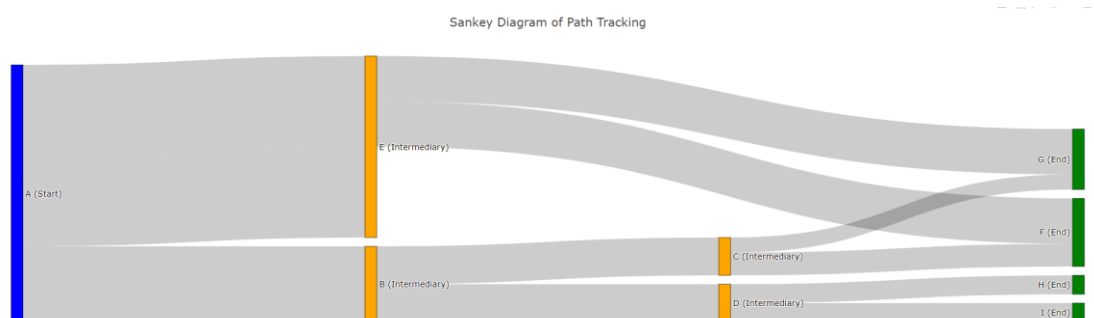


图 30 路径跟踪可视化图示

在这个路径跟踪可视化中，可以看到起始账户 A 将资金通过不同的路径进行分配，这些路径涉及多个中转账户（如 B、C、D、E），再分流到不同的最终账户（如 F、G、H、I）。桑基图清晰地展示了资金的流动方向和每一步的金额变化。

例如，A 通过路径将 12,000 直接转移至 E，而 E 又将资金分成 6,000 分别流向 F 和 G。这种多级分散的资金流动方式是典型的洗钱模式，尤其是在中转账户中进行的多次分拆和转移，进一步隐藏了资金的真实来源和去向。

使用上述可视化方案后，我们最终可以通过仪表盘提供系统的全局监控视图，实时展示反洗钱监控的关键指标，包括总交易量、高风险账户数量、异常交易数量等。用户可以通过图表形式（如柱状图、饼图、折线图）直观了解系统的整体运行状态和风险趋势。仪表盘还支持多维度分析，用户可以选择不同的时间段、风险类型或交易维度进行筛选。除此之外，系统还提供报告生成功能，用户可以根据分析结果生成可下载的详细风险报告。这些报告包含风险分析结果和交易网络的图示，有助于记录分析过程、提供合规证明或支持进一步的决策。实现上可以使用 Tableau、Power BI 或 Metabase 等工具，支持多种图表类型和可导出报告的功能。

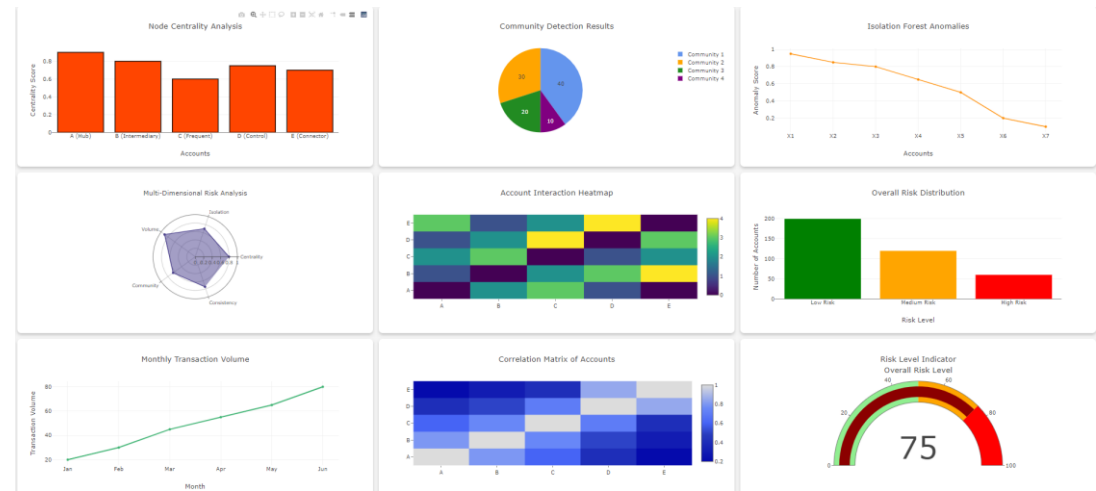


图 31 仪表盘示意图

（三）基于 AGI 的关联交易识别框架落地挑战和解决方案

我们认为，该项目落地的最重大挑战在于所需数据的共享。该项目所需的数据包括客户身份、账户信息、交易记录（尤其是账户间的频繁转账）、跨境资金流动等。不同期货公司的客户账户信息属于机密，无法直接共享，但反洗钱风险监测又需要跨公司数据的协同分析。为了解决这一矛盾，需要在保护客户隐私的前提下实现安全的数据共享。

具体的解决方案是采用联邦学习和数据脱敏技术。联邦学习允许多家公司在不交换原始数据的情况下，通过分布式模型训练实现协同分析。每家公司可以在本地训练模型，将模型参数而非数据进行共享，从而在保护客户隐私的同时，实现对高风险关联交易的监控。数据脱敏则确保在数据共享前对敏感信息（如账户 ID、交易详情）进行加密或匿名化处理，降低泄露风险。此外，通过引入区块链技术来保证数据的溯源性和不可篡改性，进一步增强跨公司数据协作的安全性和透明性。

六、总结与未来展望

（一）总结

本文首先简要回顾了人工智能在过去一段时间的发展历程，并详细讨论了人工智能在期货领域风控发展的可行路径，并通过实例的方式重点分析可以借鉴提高的案例。摩根大通的案例体现出现在企业的普遍化需求，通过 AI 高效的检索信息、阅读能力，在大量合同文件上寻找漏洞避免风险，这样的应用可以在合同初步筛查过程中为企业节约大量的时间，同时在应用层面上利用大语言模型，通过预训练合同文字数据，后续可能可以进行简单的合同解读，法律回复等等，这在期货行业尤其是涉及很多大宗贸易的业务来说非常有意义。在平安银行案例中，如何运用 AI 对于客户账户存在违规行为的监管同样非常重要。

其次，本文设计了运用大语言模型的舆情及风控的具体案例，在构建期货行业的智能风控系统中实现了多项创新，展示了显著的成果价值。首先，基于大模型（如 ChatGLM、BERT）和自动化内容生成（AIGC）技术，研究将自然语言处理和数据分析相结合，构建了智能风控体系，实现了非结构化财经新闻、社交媒体内容和市场数据的自动提取与分析，从而支持了实时舆情分析和风险预警，提升了市场响应速度和决策支持能力。其次，项目通过多源整合技术采集金融市场、社交媒体及外部风险数据，并利用无监督数据清洗和大模型技术，显著提升了非结构化数据的利用效率，保障了风险管理的精准性。

最后，针对关联交易的复杂性，本研究设计了基于图分析的交易识别框架，应用社区检测、度中心性、介数中心性等算法，构建了账户关系网络的结构化展示和实时监控，为高风险关联交易的识别和反洗钱工作提供了技术支持。为实现跨公司数据的安全共享，研究提出了结合联邦学习和数据脱敏技术的隐私保护方案，并引入区块链技术，保障了数据协作的透明性和溯源性，提升了行业间反洗钱和风险识别的协作效能。此外，系统还引入了图网络可视化和多维度风险展示，通过风险热图和路径跟踪等直观方式，使稽核人员能够更加清晰地识别和分析高风险账户、资金流动和关联关系，显著提升了风控系统的可操作性。总体而言，这些创新成果构建了一个实时、高效、自动化的智能风控系统，为期货行业提供了从数据采集、处理、分析到风险识别的全流程技术方案，不仅提升了应对市场复杂性和交易风险的能力，也为金融行业在智能风控中的创新实践提供了有力支持，展示了大模型在风控系统中的广泛应用潜力，为未来的金融科技发展奠定了基础。

（二）未来展望

尽管 AI 在期货行业的风控中表现出显著的优势，但其在以下这些方面仍有待提高：

数据质量和覆盖面：AI 系统的准确性依赖于数据的质量和覆盖面。在期货市

场中，数据源的多样性和数据质量参差不齐会直接影响 AI 模型的表现。例如，舆情监控系统依赖于社交媒体、新闻等外部数据源，但这些数据往往存在噪声，信息的可靠性和完整性也难以保障。因此，提高数据质量和覆盖面、优化数据清洗和过滤流程，是提升 AI 风控系统准确性的关键。

模型解释性和透明性：目前的 AI 模型，尤其是深度学习模型，往往存在“黑箱”问题，难以解释其预测结果。这在风险控制领域可能导致信任度不足。期货行业对合规性和透明度有较高的要求，因此，提升模型的可解释性，增加对预测结果的可视化和解读功能，使风险管理人员能够理解和追踪模型的决策过程，是 AI 风控系统的一大改进方向。

应对极端市场条件的鲁棒性：AI 模型在极端市场条件（如金融危机、市场剧烈波动）下的表现可能不稳定。由于模型在常规市场条件下训练，极端事件的数据较少，导致模型在异常情况下容易产生误判。因此，加强模型对极端市场条件的适应能力，设计更具鲁棒性的风控算法，能够进一步提高 AI 系统在风控中的可靠性。

监管合规与隐私保护：随着数据隐私和监管要求的加强，AI 风控系统在数据收集、处理和分析过程中需要严格遵守相关法律法规。在跨国交易的背景下，如何在合规的前提下实现高效的数据利用，也是 AI 在期货行业中应用的一个重要挑战。未来的 AI 风控系统需要增强对数据隐私的保护能力，并确保合规性，才能适应不断变化的监管环境。

高昂的技术成本和资源需求：构建和维护一个高效的 AI 风控系统需要投入大量的资源和资金，特别是在高频率的数据更新、模型迭代和系统优化方面。因此，如何降低成本、提高系统运行效率，同时确保系统的实时性和准确性，是 AI 风控系统在期货行业中的一个亟待解决的问题。

综上所述，AI 智能风控在期货行业中具备显著优势，能够有效提升风控效率和准确性。但在数据质量、模型解释性、极端市场应对、合规性及成本方面仍有提升空间。

展望未来，随着技术的不断发展和优化，推动金融行业在风险管理效率、准确性和响应速度等方面取得显著提升。基于大语言模型和图分析技术的突破，未来的智能风控系统将具备更强的数据整合和实时分析能力，实现多源数据的自动化整合，不仅限于新闻和社交媒体，还包括实时的全球经济数据和政策动态；生成式 AI 将支持从数据采集、清洗、分析到风险识别的全流程自动化，实现个性化和自适应的风险管理，从而在复杂市场中提供多维度、灵活的风险提示。随着图神经网络和多层次图分析算法的应用，系统将在关联交易识别上更加精准，联邦学习等隐私保护技术也将推动跨公司数据协同，形成全行业的反洗钱与风险防控网络，显著提升对复杂交易网络的识别能力。此外，未来的智能风控系统将更加注重模型的解释性，通过可解释性 AI 和可视化技术，帮助风控人员更清晰地理解风险成因和资金流动路径，支持透明、合规的智能决策。智能风控在金融行

业的深入应用将提高安全性与合规性，并逐步拓展到能源、物流、医疗等高风险行业，推动跨行业的智能化风险管理。总体而言，AI 在智能风控中的未来发展将为期货公司乃至整个金融体系的风险管理提供强大技术支持，不仅助力提升行业安全性和管理效率，还为未来的数字化金融生态建设奠定坚实基础。

参考文献

- [1] 亿欧智库, 2024-07, "2024 中国 AI 商业落地投资价值研究报告".
- [2] 廖高可, 李庭辉, "人工智能在金融领域的应用研究进展", 经济学动态 2023 年第三期。
- [3] Hu W, Chen Y, 2022, Application of Artificial Intelligence in Financial Risk Management//International Conference on Adaptive and Intelligent Systems. Cham: Springer International Publishing.
- [4] Barde K, Kulkarni P A, 2023, Applications of Generative AI in Fintech//Proceedings of the Third International Conference on AI-ML Systems.
- [5] Metawa N, Hassan M K, Metawa S, 2022, Artificial Intelligence and Big Data for Financial Risk Management, Routledge.
- [6] Kim A, Muhn M, Nikolaev V, 2023, From transcripts to insights: Uncovering corporate risks using generative ai, arXiv preprint arXiv:2310.17721.
- [7] Mbanye W, 2024, Generative AI and ChatGPT in Financial Markets and Corporate Policy: A Comprehensive Review, Available at SSRN 4745990.
- [8] Wang Y, 2023, Generative AI in Operational Risk Management: Harnessing the Future of Finance, Operational Risk Management: Harnessing the Future of Finance.
- [9] Flaig S, Junike G, 2022, Scenario generation for market risk models using generative neural networks, Risks, 10(11): 199.
- [10] Vaswani A, Shazeer N, Parmar N, 2017, Attention is all you need, Advances in neural information processing systems.
- [11] Du, Zhengxiao, 2021, Glm: General language model pretraining with autoregressive blank infilling, arXiv preprint arXiv:2103.10360.
- [12] GLM, Team, et al., 2024, ChatGLM: A Family of Large Language Models from GLM-130B to GLM-4 All Tools, arXiv preprint arXiv:2406.12793.
- [13] 刘瑶, 2024, "中国 AIGC 行业应用价值研究报告", 甲子光年智库。
- [14] Hu, Wanting, Yixian Chen, 2022, Application of Artificial Intelligence in Financial Risk Management, International Conference on Adaptive and Intelligent Systems, Cham: Springer International Publishing.
- [15] Barde, Kalpesh, Parth Atul Kulkarni, 2023, Applications of Generative AI in Fintech, Proceedings of the Third International Conference on AI-ML Systems.
- [16] Tom Bigham, Valeria Gallo, 2024, AI and risk managementInnovating with confidence, CENTRE for REGULATORY STRATEGY EMEA.
- [17] Tim Rowan, Beena Ammanath, 2024, Now decides next:Moving from potentialto performance, Deloitte's State of Generative AI in the Enterprise Quarter three report.
- [18] Student number: 1028447, 2019, Risk Management with GenerativeAdversarial Networks, A thesis submitted in partial fulfillment of the MSc inMathematical and Computational Finance.
- [19] Iwanicz-Drozdowska, Małgorzata, 2023, New technologies in the financial industry: Case of Poland, Economics and Business Review.
- [20] Banizi, Amir Hossein Talebi, Azam Alikhademi, 2024, Using Intelligent Systems to Manage Risks and Reduce Financial Risks Using Artificial Intelligence in Large Companies, Journal of Financial Risk Management.

- [21] Lee, David Kuo Chuen, 2024, A Comprehensive Review of Generative AI in Finance, FinTech.
- [22] Dong, Mengming Michael, Theophanis C. Stratopoulos, Victor Xiaoqi Wang, 2024, A scoping review of ChatGPT research in accounting and finance, International Journal of Accounting Information Systems.
- [23] Bahrammirzaee, Arash, 2010, A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems, Neural Computing and Applications.
- [24] Chen T, Li M, Li Y, Liu X, 2021, Generative adversarial networks for financial data augmentation: A systematic survey, Journal of Financial Risk Management.
- [25] Huang Y, Li S, 2022, Application of generative AI in credit risk assessment: A review, Financial Technology and Economics.
- [26] Li J, Zhao P, 2021, A review of generative AI in anomaly detection for financial fraud prevention, Journal of Financial Crime.
- [27] Gao H, Zhang Y, Qian T, 2020, Application of GANs in detecting and preventing financial fraud, IEEE Access.
- [28] Yao L, Lin Q, 2021, Improving financial risk forecasting with GANs: A quantitative analysis, Quantitative Finance.
- [30] Lopez J, Gupta N, 2020, Enhancing risk management strategies with generative AI: A review, Risk Management Journal.
- [31] Lee S, Kim D, Lee J, 2021, Risk management in banking using GAN-based synthetic financial data, Computational Economics.