

摘要

在金融行业数字化转型的浪潮中，金融对账报表系统作为保障资金安全、业务合规的关键环节，面临着数据量大、业务规则复杂、人工处理效率低且易出错等挑战。传统的对账方式已难以满足日益增长的业务需求，为了提升对账效率、增强数据准确性和挖掘数据价值，我司决定将大模型技术应用于金融对账报表系统。我作为该项目的技术骨干，负责大模型的选型与适配，结合金融对账业务特点对模型进行微调训练。同时，主导大模型与现有系统的集成开发工作，优化数据处理流程，确保大模型能够高效准确地处理海量金融数据。通过引入大模型技术，系统实现了智能对账、异常预警和数据洞察等功能，显著提高了对账效率和准确性，为金融业务的稳健发展提供了有力支持。

正文

作为支付清算领域核心服务商，公司每年支撑超 10 亿笔跨机构交易（覆盖银行、第三方支付、商户等 12 类参与方），传统对账系统依赖“固定规则+人工核查”，暴露三大战略级痛点：月均对账周期 5 天，无法匹配业务“T+1”结算的规模化扩张需求；异常交易漏判率超 2%，曾因未识别重复入账面临监管质询；监管报表需人工整理 10 类数据，耗时 3 天且易出错——这些问题不仅侵蚀运营效率，更威胁公司在金融科技赛道的技术壁垒。

为破解痛点、巩固战略竞争力，公司将大模型技术与金融对账场景深度融合，启动“智能金融对账报表系统”项目（系公司“大模型赋能金融核心业务智能化”战略的关键落地场景）。作为项目负责人，我牵头联动业务、AI 研发及合规部门，聚焦“智能提效、精准判异、合规赋能”三大目标，设计核心模块：一是跨渠道交易归一化模块，整合异构交易数据（银行报文、支付订单、商户清算文件等），实现交易要素（时间、金额、对手方）的统一映射；二是智能对账规则引擎，基于大模型学习 3 年历史案例，自动生成动态规则（如识别“同一交易号金额不符”“跨机构重复提交”等异常），替代 80% 固定规则；三是合规风险预警模块，联动反洗钱、财务准则，自动标记对账差异中的疑似违规行为（如高频小额拆分交易）；四是监管报表可视化模块，按人行《支付清算统计报表》等要求自动生成多维度报表，缩短报送时间 50%。

项目是公司大模型技术在金融核心场景的首次规模化应用，旨在通过智能化对账能力，支撑业务年交易规模向 20 亿笔扩张，同时筑牢合规底线——既是业务发展的“刚需”，更是公司金融科技能力的“标杆”。

大模型技术是人工智能领域的重要进展，指具有大量参数和强大计算能力的深度学习模型。大模型通过在大规模数据集上进行训练，学习到丰富的语言知识和模式，从而具备强大的语言理解和生成能力。大模型的核心在于其深度神经网络结构，如 Transformer 架构，它采用自注意力机制，能有效捕捉长距离依赖关系，提升模型性能。在训练方面，通常运用无监督

学习和有监督学习相结合的方式，先通过无监督学习从海量文本数据中学习通用语言知识，再通过有监督学习针对特定任务进行微调。大模型的应用极为广泛，在自然语言处理领域，可用于智能客服、机器翻译、文本生成等，能显著提高工作效率和质量；在医疗领域，辅助疾病诊断、医学文献分析等；在金融领域，用于风险评估、投资分析等。大模型技术虽带来诸多机遇，但也面临数据隐私、计算资源消耗大、模型可解释性差等挑战，需要在发展过程中不断优化和解决。

在医疗数据处理与应用的实际场景中，有许多技术手段和方法可以助力我们达成目标。对于医学领域数据的处理，既要保证信息提取的精准度，又要兼顾多机构用户隐私数据的保护。接下来，我将围绕两个关键方面展开阐述，一是如何通过无监督预训练结合医疗领域标注数据微调来精准提取医学文献中疾病诊断相关信息，二是怎样利用联邦学习框架实现多机构用户隐私数据的大模型联合训练以避免原始数据直接共享。

我们通过无监督预训练结合医疗领域标注数据微调，实现医学文献中疾病诊断相关信息的精准提取。在金融对账报表系统开发中，我们发现医保费用报销对账需要验证医学文献中的疾病诊断是否符合报销规则，而传统关键词匹配法经常因为遗漏同义词或误判歧义导致人工复核率高达 60%，严重拖慢对账效率。为解决这一问题，我们依托国产 ERNIE - Medical 预训练模型，借助模型积累的医学语言理解能力把握文献中专业术语的上下文关联；整合合作医院覆盖糖尿病高血压等 12 类常见病种的 10 万条疾病诊断标注样本，围绕疾病实体识别任务优化模型输出层，强化诊断信息的定位精度；同时在微调中融入领域适配策略，提升模型对并发症合并症等复杂诊断的识别能力。最终疾病诊断信息提取准确率从 78% 提升至 92%，对账系统因提取错误产生的人工复核率下降 40%，有效支撑报销规则的自动化验证，大幅提升金融对账报表的处理效率。

我们用联邦学习框架实现多机构用户隐私数据的大模型联合训练，避免原始数据直接共享，解决金融对账报表系统中跨机构数据协同的隐私痛点。在金融对账场景中，银行、支付机构、清算平台需协同用户交易记录、账户余额、手续费明细等数据识别对账单差异，但直接共享原始数据违反《个人信息保护法》和《金融数据安全管理规范》，且单机构数据覆盖的用户场景有限，导致对账模型泛化能力弱、差异识别漏检率高。为解决该问题，我们基于国产联邦学习框架 FATE 构建隐私对齐 - 本地训练 - 加密聚合的联合训练体系：首先用隐私集合求交技术对齐多机构加密后的用户身份标识，仅筛选共同用户数据参与训练，避免泄露非交集用户信息；然后各机构用本地对账数据训练模型基础参数，通过同态加密将参数加密后上传至联邦服务器聚合，全程不接触原始数据；最后用重复记账金额不符等标注差异账单样本微调模型，优化识别精度。最终联合训练模型准确率较单机构提升 15%，差异漏检率降至 2% 以内，所有机构未共享原始数据且符合监管要求，跨机构对账处理效率从每周缩短至每日。

在金融对账报表系统中应用大模型技术后，取得了显著成效。大模型凭借其强大的自然语言处理和数据分析能力，极大提升了对账报表生成的效率和准确性。以往需要人工花费数天时间完成的复杂对账任务，现在借助大模型技术仅需数小时即可完成，并且错误率大幅降低，为企业节省了大量的人力和时间成本。同时，大模型还能对海量的金融数据进行深度挖掘和分析，为企业提供更具价值的决策支持。

然而，在实践过程中，我也对大模型技术有了更进一步的思考。尽管大模型技术优势明显，但它并非十全十美。一方面，大模型的训练需要大量的数据和强大的计算资源，这对于一些小型金融企业来说是难以承受的，可能会加剧行业内的资源不平衡。另一方面，大模型的可解释性较差，在金融领域这种对风险控制要求极高的场景下，一旦模型出现错误或异常，很难快速定位问题根源，这可能会给企业带来潜在的风险。此外，大模型技术也面临着数据隐私和安全方面的挑战，金融数据的敏感性使得数据泄露的后果更加严重。因此，在应用大模型技术时，我们需要全面权衡其利弊，谨慎应对可能出现的问题。