

人工智能审计工具对上市公司财务舞弊识别的有效性研究

李倩

中国石油天然气股份有限公司重庆销售江北分公司，重庆，400000；

摘要：数字经济时代，上市公司财务舞弊日益隐蔽、复杂、海量化，传统审计方法在效率与精准度上遭遇瓶颈。人工智能审计系统利用机器学习、自然语言处理等技术，构建一体化智能识别体系。本文分析 AI 在财务舞弊识别中的有效性，包括技术原理、应用场景等，阐明其通过数据扫描、文本解读、风险建模实现精准识别。结果显示，AI 打破经验依赖，但受限于数据质量、算法可解释性及人机协作效率。未来需优化数据治理、创新算法模型、再造审计流程，深化技术与审计融合，贡献新范式。

关键词：人工智能审计；财务舞弊；人机协作；审计融合

DOI：10.64216/3080-1486.26.02.060

引言

随着数字技术在资本市场的应用^[1]，上市公司财务舞弊手段持续更新：从传统虚增收入、虚减成本演变为利用区块链跨境资金循环、SPV 表外融资、AI 伪造交易记录等新方式^[2]。中国证监会统计显示，2023 年 47% 的财务舞弊案件涉及复杂金融工具和跨境交易，较 2015 年上升 29 个百分点。传统审计依赖抽样和经验判断，处理 TB 级数据时存在漏判率约 35%、识别滞后平均 45 天等不足。

人工智能技术通过算法挖掘数据，为舞弊识别提供新工具：机器学习识别异常分布模式，自然语言处理解析 MD&A 语义矛盾，深度学习捕捉时序关联特征。本文聚焦 AI 审计工具的有效性，从技术赋能、应用场景创新、效能瓶颈突破等维度论述，旨在为构建“技术驱动+专业判断”的新舞弊识别体系提供理论支撑。

1 财务舞弊识别的传统困境与人工智能赋能逻辑

1.1 财务舞弊的新型特征与审计挑战

1.1.1 舞弊手段的隐蔽化升级

交易结构更趋复杂：通过多层嵌套的关联交易（例如某上市公司设立 7 级子公司转移利润）、金融衍生工具（如利用期权合约隐匿表外负债）来模糊交易实质，使得传统审计的穿透式核查成本大幅增加^[3]；

数据操纵更为智能：利用财务软件漏洞批量伪造凭证（如某企业通过宏命令每天生成 2000 多条虚假采购订单）、借助 AI 生成技术制作逼真的银行对账单影像，导致肉眼难以分辨真伪；

文本粉饰更加精细：在 MD&A 中使用“业绩稳健”等正面表述掩盖现金流断裂风险，同时通过行业术语进行包装（如将“存货减值”表述为“资产结构优化”），

加大了语义识别的难度。

1.1.2 传统审计的适配性困境

数据处理效率不高：人工审计单家上市公司年报的文本审阅需要 40 多小时，对 10 万多条交易流水的抽样检查覆盖率不到 5%，异常数据捕捉能力较弱；

舞弊模式识别滞后：依赖历史案例总结的舞弊特征库（如“毛利率明显高于行业平均水平”），对新型舞弊模式（如“通过加密货币交易转移资金”）的识别存在 2-3 年的滞后；

跨维度关联分析不足：难以将财务数据（如应收账款账龄）、市场数据（如股价波动）、舆情数据（如媒体质疑报道）关联起来进行综合判断，从而导致重大风险点漏判。

1.2 人工智能审计的技术赋能原理

人工智能审计工具通过“数据输入-特征工程-模型训练-舞弊识别”的技术链条，实现对舞弊特征的智能化捕捉，核心技术支撑包括：

1.2.1 机器学习算法的模式识别

监督学习模型：利用历史舞弊样本训练随机森林、支持向量机等分类模型，提取关键特征（如资产负债率突变率、经营活动现金流与净利润比值小于 0.5），对已知舞弊模式的识别准确率能达到 85%以上^[4]；

无监督学习模型：通过 K-means 聚类检测数据分布异常点（如某季度管理费用增长率与收入增长率差值超过 40%），利用 DBSCAN 算法识别交易流水的密度异常区域（如同一账户单日有 10 笔以上整数金额转账）；

集成学习模型：构建梯度提升树（GBDT）集成多个弱分类器，综合判断复杂的舞弊场景（如同时存在收入虚增、存货异常、关联交易频繁等多维度异常）。

1.2.2 自然语言处理的语义解析

情感分析技术：借助 BERT 预训练模型解析 MD&A 文本情感值，当文本情感值（+0.7）与净利润增长率（-15%）出现背离时触发预警；

实体关系提取：识别文本中的关联方名称、交易金额、会计政策术语，构建“公司-关联方-交易金额-时间”关系图谱，以发现异常关联交易网络（如某公司与 3 家注册地址相同的关联方存在亿元级资金往来）；

文本生成对抗网络（GAN）：通过生成模型模拟正常财务报告的语言模式，对抗模型检测文本偏离度，从而识别异常表述（如某公司年报中“风险提示”章节字数不到 300 字，且未提及具体应对措施）。

1.2.3 深度学习的时序动态建模

LSTM 神经网络：捕捉财务数据的时序异常，例如连续三个季度“营业成本增长率大于收入增长率且差值扩大”，将其识别为利润操纵风险；

图神经网络（GNN）：构建企业股权关系、资金流动、业务合作的图结构，检测跨公司的舞弊网络（如 5 家关联企业形成闭环交易链，年交易规模为 20 亿元）；

数字孪生技术：对上市公司财务状况进行虚拟建模，通过压力测试模拟不同舞弊场景下的指标波动，提前识别潜在的风险点。

2 人工智能审计工具的核心应用场景

2.1 数据审计场景：从抽样检查到全域扫描

2.1.1 异常交易精准定位

人工智能工具通过交易流水的多维度分析，精准定位三类异常交易：

时间序列异常：运用动态时间规整（DTW）算法，识别交易时间分布与历史模式的偏离（如某公司在财报发布前 15 天密集出现大额费用冲回，频次达日均 3 倍）；

金额特征异常：依据 Benford 定律验证数字频率分布，当某科目数据中“1”开头数值占比低于 30%（理论值 30.1%）且“9”开头占比超过 15%时，提示数据操纵风险；

网络关联异常：采用社区发现算法（如 Louvain 算法）识别交易对手方集群，若某公司与 10 家成立不足 1 年、股权高度分散的企业形成交易网络，且交易价格偏离市场均值 40%以上，触发红色预警^[5]。

2.1.2 科目联动智能分析

传统审计依赖人工经验判断科目勾稽关系，人工智能则通过关联规则算法（如 Apriori 算法）挖掘异常联动模式：

当“存货周转率下降 25%以上+应付账款账期延长 45 天+毛利率上升 8%”时，判定为“存货虚增利润嫌疑”；

当“研发费用资本化率超 75%+专利授权量同比下降 50%”时，识别“研发支出虚增风险”；

当“商誉/净资产比率超 30%+被收购方业绩承诺完成率低于 60%”时，预警“商誉减值准备计提不足”。

2.2 文本审计场景：从人工审阅到智能解析

2.2.1 管理层讨论与分析（MD&A）舞弊信号捕捉

借助自然语言处理技术，捕捉文本中的三类舞弊信号：

情感背离信号：净利润同比下滑 20%，但 MD&A 中“业绩稳健”“增长可期”等积极词汇出现频次为行业均值 2 倍以上；

逻辑矛盾信号：同时声称“强化成本控制”与“扩大研发投入”，但销售费用率与管理费用率合计下降 15%（资源配置逻辑冲突）；

术语异常信号：频繁使用“创新业务”“战略调整”等模糊表述，回避商誉减值、存货跌价等具体会计处理细节（信息披露透明度不足）。

2.2.2 审计底稿智能校验

针对审计底稿中的文本与图像数据，实现三方面校验：

一致性校验：发票日期与合同签署日期间隔超 90 天（疑似后补票据），或发票金额与入库单数量不匹配（如采购 100 件商品，发票金额对应 120 件）；

完整性校验：银行对账单关键流水记录缺失（如亿元级存款变动未附凭证），或存货盘点表连续 3 页出现相同盘盈盘亏描述（疑似复制填报）；

合规性校验：关联交易协议未明确具体定价机制（如仅标注“按市场价格结算”但无参考依据），或担保合同中担保方资质文件不全（如缺少营业执照复印件）。

2.3 风险预警场景：从滞后响应到实时监控

2.3.1 动态风险评估模型

构建包含 30 余项核心指标的动态评估体系，实时计算风险等级：

财务指标层：资产负债率超 75%且 EBITDA 利息保障倍数低于 1.2（偿债风险预警），应收账款周转率低于行业中位数 60%（资金回收风险预警）；

市场指标层：股价波动幅度超大盘 60%且分析师评级集中下调（市场信心异常预警），融券余额占流通股比例超 5%且持续增持（做空压力预警）；

舆情指标层：股吧负面情绪指数超 0.6 且媒体质疑报道月均超 10 篇（舆情风险预警），企业信用评级展望为“负面”且高管减持比例超 3%（内部信心危机预

警)。

2.3.2 舞弊模式进化追踪

通过强化学习算法,工具可自动识别舞弊手段的迭代趋势:

当传统“虚增收入”案件减少时,模型捕捉到“表外负债隐匿”案件中“售后回租协议未披露”“衍生品合约嵌套”等新特征,并更新识别维度;

针对跨境舞弊,模型整合国别风险数据(如避税地关联交易占比超40%)、资金流向异常(如海外账户资金滞留超180天)、汇率波动影响(如结汇损失占净利润15%以上)构建新识别模型,将跨境舞弊识别准确率提升至82%。

3 人工智能审计工具的有效性瓶颈与突破路径

AI审计有效性提升面临数据质量、算法局限及人机协同三重核心障碍,其突破需通过数据治理优化、算法迭代创新及人机协同重构实现。数据质量层面,内外部数据(如企业ERP/CRM/OA系统与工商、税务、舆情平台数据)未有效整合形成数据孤岛,导致某AI工具因缺银行流水漏判30%资金挪用案件;财务数据合理波动(会计政策变更、重组并购等)与舞弊特征重叠超40%,正常商誉减值误判率达25%;审计数据滞后业务发生30-60天,难以及时捕捉舞弊萌芽。算法层面,深度学习模型存在黑箱效应(如LSTM模型特征权重不透明),审计师信任度仅45%;过度依赖历史样本致过拟合,对“区块链跨境资金循环”等新舞弊识别准确率32%;制造业样本占比超70%引发偏差传递,服务业识别效能下降40%。人机协同层面,仅23%审计师能解读随机森林特征重要性排序,工具应用停留在“结果接收”;AI漏判时责任边界模糊(审计准则未明确条款);审计流程未嵌AI接口,异常交易未入实质性测试。为此,数据治理需构建“内部数据仓+外部数据湖”混合架构,API实时接入多源数据提升样本量5倍,异常识别覆盖率从60%升至92%;开发智能清洗引擎过滤噪声,清洗效率85%、误判率降至10%以下;通过Flink框架实现交易数据分钟级响应,滞后压缩至24小时内。算法迭代可采用SHAP技术解析模型决策(如可视化“存货周转率下降”贡献度,透明度提升60%),利用GAN模拟新舞弊数据将未知模式识别准确率提至78%,并按行业动态调整指标权重(房地产“存货/资产”权重0.3、科技行业“研发费用增长率”权重0.25,跨行业效能差异缩至15%以内)。人机协同需重构流程(AI初筛80%财务数据提升现场效率40%,报告阶段AI生成分析报

告)、建设“审计知识+AI应用”二维能力矩阵(认证审计师工具应用深度提3倍),并制定指引明确复核责任(对前20大异常交易100%穿透核查),形成“机器筛异常-人类判性质-系统持续学”闭环。

4 结论与未来展望

人工智能审计工具凭借技术赋能,在财务舞弊识别领域展现出传统审计难以企及的优势,不仅将数据处理效率提升5倍以上,异常交易识别覆盖率超90%,还能将舞弊信号识别滞后时间压缩至24小时以内,其核心价值在于突破经验依赖,构建起“数据驱动-模型决策-人机协同”的新型识别体系;然而,其有效性提升仍需克服数据质量、算法解释性及责任界定等核心挑战。未来研究可聚焦于:一是发展跨模态融合技术,整合财务数据、文本信息与图像证据构建多模态识别模型,提升对复杂舞弊场景的穿透能力;二是深化区块链审计应用,利用其不可篡改特性构建审计数据共享平台,从源头保障数据真实性并降低舞弊识别验证成本;三是构建智能监管生态,探索与监管科技(RegTech)的协同机制以实现舞弊风险“识别-预警-处置”的全链条智能化,为资本市场财务安全提供技术保障。人工智能审计工具的发展并非颠覆传统审计,而是通过技术赋能实现审计能力的范式升级,为此,需持续推进技术创新与审计实务的深度融合,以构建适应数字经济时代的财务舞弊治理体系,为资本市场的健康发展筑牢防线。

参考文献

- [1]刘毅楠.新质生产力视域下数字金融科技治理论纲[J].中国科技论坛,2025,(08):32-42.
- [2]纪玉俊,公洁.数字技术与制造业现代化产业体系建设:渐进与空间双重效应下的新证据[J/OL].产业经济评论,1-19[2025-09-15].
- [3]付小川,周杰.不公允关联交易的法律认定困境与规制路径研究——基于《公司法》与司法实践的实证分析[J].中国证券期货,2025,(04):72-77.
- [4]邱捷.上市公司财务舞弊智能化预警模型构建及应用研究[D].首都经济贸易大学,2021.
- [5]付安玲,张书铭.生成式人工智能背景下网络意识形态风险治理的逻辑理路与治理实践[J].学习论坛,2025,(05):5-13.

作者简介:李倩,1991.10.20,女,汉族,重庆人,本科学历,职称:高级职称,研究方向:财会。