人工智能驱动下银行业合规治理的范式跃迁

王晨烨[1] 支沛[1*] 余沂莲[2] 周西庆[2] 丁玄[3*]

（1.北京服装学院时尚管理学院，北京 朝阳 110105；2.江西服装学院商学院，江西 南昌 330201；3.阜阳师范大学法学院，安徽 阜阳 236000）

摘 要：随着生成式人工智能等技术的深度渗透，银行业的合规治理正面临由“算法黑箱”、模型幻觉与数据伦理引发的根本性挑战。本研究基于 2024 至 2025 年间全球银行业的代表性实践，系统剖析了人工智能合规治理从“被动响应”到“主动嵌入”、从“工具应用”到“操作系统级能力”的范式跃迁。分析表明，领先机构正通过构建覆盖全生命周期的自动化治理框架、倡导“通过设计构建信任”的伦理原则，以及开展“监管沙箱”内的协同创新，将合规转化为核心竞争力。研究指出，未来银行业的竞争不仅是技术场景的多寡，更是治理体系能否将伦理约束内化为技术基因，从而在创新与稳健之间建立动态平衡。

关键词：人工智能治理；银行合规；模型风险管理；全生命周期治理；伦理对齐；监管科技

1.引言

人工智能，特别是大模型技术，正以前所未有的速度重塑银行业的价值链。从智能投顾、自动化信贷审批到实时反洗钱监控，AI 的应用已从提升效率的辅助工具，演进为驱动核心业务决策的关键引擎。然而，这种深度整合也暴露并放大了传统合规体系的脆弱性。模型自身的“黑箱”特性、难以根除的“幻觉”问题，以及在训练数据中可能固化的社会偏见，使得金融活动面临着全新的、系统性的合规与伦理风险。中国人民银行科技司司长李伟明确指出，若缺乏有效治理，大模型可能生成歧视性内容，或通过“标签化”实施不公平定价，从根本上损害金融的公平性与普惠性。

在此背景下，银行业的合规治理范式已无法停留于对既定规则的事后检查与人工复核，而必须向 “主动式、嵌入式、智能化” 的新范式演进。这一演进的核心，是将合规与伦理要求前瞻性地嵌入人工智能模型从设计、开发、部署到监控退役的全生命周期，使之成为技术架构的内生属性。本文旨在通过聚焦 2024 至 2025 年间国内外领先银行的创新实践，揭示这一范式跃迁的具体路径、技术内核与战略价值，为银行业构建面向人工智能时代的韧性治理体系提供学术参考与实践镜鉴。

2.范式跃迁的三重维度：从治理实践到核心能力

全球领先银行的探索表明，人工智能合规治理的演进并非单一技术的升级，而是沿着技术嵌入、伦理内化与监管协同三个维度同步展开的体系化能力建设。

2.1 技术之维：从人工校验到全生命周期自动化治理

作者简介：王晨烨（2003-），男，研究生，研究方向为中国服装企业国际化、时尚产业分析、跨国纺织企业投资与经营。
支沛（2000-），男，研究生，研究方向为服装品牌运营、服装国际商务。
余沂莲（2001-），女，本科，研究方向为财会职业教育、财务管理、会计。
周西庆（2002-），男，本科，研究方向为财会职业教育。
丁玄（2003-），男，本科，研究方向为法学、法学教育、经济法。
通讯作者：支沛、丁玄

传统风控依赖规则引擎与人工抽样，而智能时代的合规要求治理贯穿数据、模型与决策全链条。巴西银行（Banco do Brasil）的实践为此提供了全球性标杆。面对旗下超过 8000 万客户和数百个生产级 AI 模型，该行于 2025 年与安永（EY）及 IBM 合作，实施了一项以治理为先的战略。其核心是借助 IBM watsonx.governance 平台，构建一个自动化的、统一的 AI 治理层。该体系实现了对模型公平性、性能漂移的实时仪表盘监控，并在指标突破阈值时自动触发警报。更重要的是，它通过自动捕获模型元数据、实现全链路可追溯性，将原本分散且依赖人工的文档、评估与审计流程标准化、自动化，从而将新模型的部署审批速度大幅提升，在严控风险的同时加速了创新落地。这种“治理即代码”的理念，标志着合规从一项成本中心业务，转型为可扩展的技术基础设施。

2.2 伦理之维：从外部约束到“通过设计构建信任”

应对算法偏见与歧视风险，需超越事后审查，在系统设计源头植入伦理准则。巴西银行的治理框架不仅关注合规，更强调使 AI 输出与机构的价值观及负责任 AI 的最佳实践保持一致。这与国际监管趋势同频共振。例如，台湾地区金融监督管理委员会于 2025 年发布《金融业应用人工智慧指引》，中国信托商业银行旋即携手 SAS 启动了台湾首个完整的“AI 治理验证计划”。该计划并非泛泛而谈，而是以分行票据审核这一具体 AI 服务为试点，系统验证其生命周期中的风险识别、监测与管理流程，确保其决策透明度与问责制符合全球标准。此举将抽象的伦理原则转化为可审计、可验证的技术与管理流程，体现了“通过设计构建信任”的核心理念。

2.3 协同之维：从封闭自控到“监管沙箱”内的开放式创新

在高度不确定的技术前沿，领先银行选择与监管机构携手，在受控环境中共同探索治理边界。这种“监管沙箱”模式成为平衡创新与风险的关键机制。2025 年世界人工智能大会上，交通银行展示的“数字分身”远程视频服务系统，以及上海银行发布的“对话即服务”AI 手机银行，都代表了人机协同服务的前沿探索。这类深度介入客户交互与金融决策的应用，其合规性需在真实业务流中接受检验。通过主动进入监管机构设定的创新测试场域，银行能够在有限范围（如特定客群、额度）内，与监管方共同设定监控指标、评估伦理影响并迭代治理规则。这种开放式协同，将监管从纯粹的“外部裁判”转化为部分的“共同设计者”，为高创新性、高不确定性应用的合规落地开辟了安全通道。

3. 技术原理

3.1 核心技术原理：从“黑箱”走向“可解析”与“可干预”

传统机器学习模型，尤其是深度学习模型，常因复杂的非线性变换与海量参数而被视为“黑箱”，这直接与金融监管要求的透明度、可解释性、可审计性相冲突。为解决这一根本矛盾，领先银行采用了以下技术路径：

基于知识增强与推理链路的可解释性生成：以宁波银行与蚂蚁数科合作的实践为例，其“Agentar 知识工程平台”的核心原理是“知识增强生成”。它并非让大模型完全依赖参数记忆生成答案，而是引入了一个“规划-检索-推理”的确定性工作流。当收到查询时，系统首先进行问题解析与规划，然后从经过审核的结构化知识图谱和非结构化文档库中进行精准检索，最后将检索到的证据片段通过推理引擎合成最终答案。这一原理的关键在于，系统输出的每一个结论都附带清晰的推理路径和溯源至具体知识源的能力，从而将生成式 AI 的“幻觉”风险控制在有限范围内，满足了合规审计对决策依据的追溯要求。这种技术原理的本质，是用确定性的知识检索与推理框架，约束大模型概率化生成的不可控性。

基于多智能体协作的模块化风险决策：某股份制商业银行构建的智能反欺诈系统，其原理是“多智能体系统”架构。它将复杂的反欺诈任务分解为信用评估、交易核验、行为分析、关联网络排查等子任务，并训练多个专用的“子智能体”分别负责。一个中央的“主智能体”或“协调引擎”负责理解总体风险场景，并按照预设的工作流动态调度和组合这些子智能体的输出。例如，对于一笔跨境交易，系统可能依次调用“地理位置核验体”、“交易对手历史行为分析体”和“反洗钱规则匹配体”。这种设计的治理原理在于：a）模块化：每个子智能体功能单一，易于监控和评估其性能与公平性；b）可审计：整个决策链条被分解为清晰的步骤，便于定位错误或偏差的来源；c）鲁棒性：单一模块的更新或失效不会导致整个系统崩溃。这标志着风控从“端到端黑箱模型”向“白盒化协作系统”的演进。

基于动态工具调用的复杂任务可靠执行：江苏银行在授信材料智能鉴伪中应用的大模型技术，体现了“大模型即大脑，专业工具即四肢”的原理。其技术框架允许大模型根据对材料内容的实时理解，动态调用一系列外部工具，例如：调用 OCR 引擎提取文字、调用图像真伪检测模型分析水印与篡改痕迹、调用外部工商数据库 API 验证企业信息。大模型的核心角色是“任务规划者”和“信息整合者”，而非全能执行者。这一原理的治理优势在于，它将大模型的通用认知能力与经过验证的、高精度的专业工具相结合，既发挥了前者理解复杂场景的优势，又通过后者保证了关键环节（如真伪判断）的确定性和高准确率，有效控制了因大模型自身知识局限或幻想带来的合规风险。

3.2 治理机制原理：构建覆盖模型生命周期的管控闭环

仅有技术方案不足以构成治理，必须建立与之匹配的管理机制。以巴西银行和中国信托商业银行的实践为代表，其治理机制的原理是构建一个 “全生命周期、自动化监控、持续验证” 的管控闭环。

治理即元数据管理：实现全景可见性。巴西银行部署的 watsonx.governance 平台，其底层原理是将治理活动转化为对“元数据”的系统性管理。该平台自动捕获并关联 AI 模型从数据来源、特征工程、模型版本、训练参数、公平性评估结果到生产环境性能指标的全链路元数据。这建立了一个统一的“数字孪生”视图。其治理作用在于，它将原本分散在数据科学家、工程师和业务部门手中的碎片化信息，整合为可追溯、可审计的权威记录，使模型的任何一次决策都能回溯到其训练数据和版本，满足了《巴塞尔协议》操作风险高级计量法（AMA）等监管框架对模型文档化的严格要求。

验证即压力测试：将原则转化为可执行标准。中国信托商业银行携手 SAS 开展的“AI 治理验证计划”，其原理借鉴了金融领域的“压力测试”和软件工程的“验证与确认”思想。它并非抽象地讨论伦理，而是将“公平性”、“可解释性”、“稳健性”等原则，转化为一套针对“分行票据审核”这一具体模型的可执行测试用例。例如，通过注入包含不同行业、规模企业票据的测试集，验证模型是否存在系统性偏见；通过模拟图像模糊、格式异常等对抗性样本，测试模型的鲁棒性。这一机制的原理是，通过在一个可控的试点场景中进行极限测试，为全行范围的 AI 治理政策、技术标准和审计流程提供实证基础和可复用的方法论，实现了从治理原则到落地实践的“最后一公里”贯通。

监控即持续学习：从静态合规到动态适应。现代 AI 治理机制的核心原理之一，是认识到模型上线并非终点，其性能会随数据分布变化（概念漂移）而衰减。因此，治理平台（如上述 watsonx 平台）集成了自动化持续监控模块。其原理是设定关键绩效指标（如准确率、公平性指标）和风险指标（如输入数据分布偏移度）的阈值，进行实时跟踪。一旦监测到模型性能退化或出现预期外的偏差，系统能自动告警甚至触发模型重训练流程。这种机制将合

规从一次性的上线审批，转变为覆盖模型整个服役期的动态、持续的保障过程，确保 AI 系统在快速变化的市场环境中始终保持合规与可靠。

4.核心实践场景：智能合规的纵深突破

上述治理范式的演进，在银行业务的具体场景中正结出实质性成果，尤其在智能风控、数据洞察及运营自动化领域实现了纵深突破。

4.1 智能风控：从规则匹配到多智能体协同决策

在反洗钱、反欺诈等高复杂性风控领域，AI 正从执行单一规则向自主协同决策演进。IDC 2025 年的报告显示，某股份制商业银行构建了一个“主智能体-子智能体”分层架构的多智能体协作引擎。主智能体能够协调“贸易数据核验”、“信用评估”、“风险预警”等 8 个专业子智能体，通过工作流引擎实现毫秒级任务解析与嵌套调用，最终将欺诈识别准确率从 85%显著提升至 99.2%，年挽回损失超 2 亿元人民币。这一案例表明，合规风控已进入“系统性对抗系统性风险”的阶段，其治理重点也转向确保多智能体协作流程的透明度、稳定性与整体合规性。

4.2 数据洞察与报告：从经验驱动到算法驱动

在监管报告、内部审计和行业研究领域，AI 正在改变知识生产的模式。中国邮政储蓄银行于 2025 年获奖的“邮赢洞见”全景数据分析引擎，首创了“找数—问数—算数—用数”的智能体系。该引擎基于自然语言处理技术，允许业务人员以问答方式提取数据，并自动化撰写涵盖资产、负债、盈利及风险维度的深度分析报告。这不仅仅是效率提升，更意味着合规与战略决策的基础从高管经验直觉，转向由算法驱动的、标准化的全景数据洞察，使决策过程本身变得更可追溯、可分析。

4.3 运营自动化：从流程优化到风险主动拦截

在运营合规场景，AI 的应用从优化单点效率升级为对全流程风险的主动扫描与封堵。例如，深圳中国人寿研发的“收付费 AI 智能监控平台”，通过 AI 智能检索引擎，将合规数据提取时间从 5 天缩短至 5 分钟，并通过实时比对与智能预警，将每笔异常交易的定位时间从 30 分钟压缩至 1 分钟，实现了反洗钱等合规工作从“人防”到“技防”的质变。这与前文提及的某些银行在消保审查、票据审核等场景的实践异曲同工，其共同特征是将合规检查点深度嵌入业务操作流，从事后抽查变为事中即时阻断。

机构名称	实践领域	核心举措与技术特征	治理理念与成效
巴西银行	全行级AI治理体系	与安永（EY）、IBM合作，部署watsonx.governance平台，实现对全行数百个生产级AI模型的自动化监控、元数据捕获与全生命周期可追溯管理。	“治理即战略”：将治理从成本中心转变为加速创新、建立内外部信任的核心战略能力，为规模化AI应用奠定合规基础。
中国信托商业银行	AI治理验证计划	响应监管指引，携手SAS启动中国台湾地区首个完整的“AI治理验证计划”，以“分行票据审核”AI服务为具体试点，系统验证其全生命周期的风险管理流程。	“通过设计构建信任”：将抽象的监管与伦理要求，转化为可审计、可验证的具体技术与管理流程，树立行业合规实践标杆。
某股份制商业银行	智能反欺诈风控	构建“主智能体-子智能体”分层协作引擎，协调8个专业子智能体（如贸易核验、信用评估等）完成毫秒级任务解析与决策，实现复杂欺诈模式的系统性识别。	“系统性对抗系统性风险”：通过多智能体协同，将欺诈识别准确率显著提升至99.2%，实现风控效能从量变到质变的跨越。
中国邮政储蓄银行	智能数据分析与报告	开发“邮赢洞见”全景数据分析引擎，首创“找数—问数—算数—用数”智能体系，支持自然语言交互的“智能问数”与多维度分析报告的自动化生成。	“决策算法化”：将内部报告、合规分析与战略决策的基础，从高管经验直觉转向算法驱动的全景数据洞察，提升决策过程的可追溯性与标准化。
交通银行/上海银行	人机协同客户服务	在2025年世界人工智能大会（WAIC）分别展示“数字分身”远程视频服务系统与“对话即服务”AI手机银行，探索高拟真、高互动性的人机协同前沿模式。	“协同式创新”：在“监管沙箱”思维指导下，于可控环境中探索高互动性AI应用的合规边界与伦理准则，平衡客户体验与未知风险。

表1：银行业人工智能合规治理代表性实践对比（2024-2025）

5.结论与展望

综上所述，2024至2025年间银行业的实践清晰地表明，人工智能驱动的合规治理正在经历一场深刻的范式革命。其内涵已从对外部监管条例的被动遵循，演变为一项需要前瞻性布局、深度技术融合与跨领域协同的战略性核心能力。成功的关键在于构建一个如“操作系统”般稳固的底层治理架构，该架构能实现技术调度的敏捷性、对业务流程的穿透力以及与组织目标的协同性。

展望未来，银行机构面临的挑战将愈发复杂。模型幻觉的完全消除、动态演进中算法公平性的持续保障、以及跨境数据流动与隐私计算中的合规问题，均是待解的难题。下一阶段的竞争，将不仅是人工智能应用场景的数量之争，更是治理体系韧性与适应性之争。那些能够将伦理考量深度编码进技术基因，并使其治理框架与AI技术及监管环境同步迭代的银行，方能将合规从成本负担转化为真正的品牌信任与竞争优势，从而在智能金融的新纪元中行稳致远。未来的研究可进一步聚焦于治理效能的可量化评估体系，以及跨文化、跨法域下人工智能治理原则的互认与协调机制。

参考文献：

[1]王柯瑾.竞逐3万亿元增量商业价值上市银行谋划AI战略图谱[N].中国经营报,2025-03-31(B05). DOI:10.38300/n.cnki.nzgjy.2025.000605.

[2]孙涛. 结合实践探索看DeepSeek等AI技术在商业银行的应用趋势[J].中国银行业,2025,(03):91-94+104.

[3]张祎.年薪最高80万!AI人才成为银行春招“香饽饽”[N].每日经济新闻,2025-03-26(002). DOI:10.28571/n.cnki.nmrjj.2025.000813.

[4]齐金钊.平安银行行长冀光恒:坚持以零售为主、带有科技基因战略方向[N].中国证券报,2025-03-25(A03). DOI:10.28162/n.cnki.nczjb.2025.001144.

[5]郑金纲,李洁,施妍萍. 以“RPA+AI”技术打造邮政手机银行营销新质生产力[J].邮政研究,2025,41(02):10-15.DOI:10.13955/j.yzyj.2025.02.02.06.

[6]许予朋.平安银行:坚持做强零售业务[N].中国银行保险报,2025-03-18(003). DOI:10.28049/n.cnki.ncbxb.2025.000860.

[7]黄钰霖.人工智能推动人才战略重构银行校园春招缩量提质[N].证券时报,2025-03-17(A03). DOI:10.38329/n.cnki.nzjsb.2025.001113.

[8]张小洁.借力DeepSeek银行业智能化变革提速[N].经济参考报,2025-03-11(006). DOI:10.28419/n.cnki.njjck.2025.000955.

[9]姜其林,陈焕雷,刘文. AI在智慧银行运营中的数据分析与客户行为预测研究[J].金融科技时代,2025,33(03):11-15.

[10]杜肖锦.激发民间创新创业热情[N].中国银行保险报,2025-03-06(006). DOI:10.28049/n.cnki.ncbxb.2025.000774.

[11]向红.从指尖到心间AI如何让手机银行更懂你[N].中国城乡金融报,2025-02-28(A04). DOI:10.28148/n.cnki.ncxjr.2025.000168.

[12]周萃.DeepSeek催化农商银行变革:低成本AI重构银行生态版图[N].金融时报,2025-02-27(010). DOI:10.28460/n.cnki.njrsb.2025.000740.

[13]中国农业银行审计局武汉分局课题组,桂艳芳,刘莹. AI技术在商业银行反洗钱、反欺诈领域的应用场景研究[J].农银学刊,2025,(01):39-42.DOI:10.16678/j.cnki.42-1864/f.2025.01.008.

[14]夏晖.北京银行创新推动AI算力产业链加速发展[N].首都建设报,2025-02-19(001). DOI:10.28681/n.cnki.nsdjs.2025.000256.

[15]周艾琳.AI浪潮将重塑金融五大领域“成长的烦恼”与机会并存[N].第一财经日报,2025-02-19(A01). DOI:10.28207/n.cnki.ndycj.2025.000576.

[16]蔡越坤,洪小棠.全面重估中国科技股[N].经济观察报,2025-02-17(009). DOI:10.28421/n.cnki.njjgc.2025.000128.

[17]李静.银行的“DeepSeek时刻”:信贷审核等场景已搭上AI快车[N].中国证券报,2025-02-11(A03). DOI:10.28162/n.cnki.nczjb.2025.000504.

[18]李若菡.“未来银行”加速到来[N].国际金融报,2025-02-10(012). DOI:10.28403/n.cnki.nifnb.2025.000088.

[19]李国辉.金融大模型:“奇点”时刻正在迫近?[N].金融时报,2025-02-06(004). DOI:10.28460/n.cnki.njrsb.2025.000419.

[20]陈璐,王小彩. 商业银行生成式AI发展趋势[J].中国金融,2025,(03):68-69.

[21]秦志刚.生成式AI技术赋能银行转型增长[N].国际商报,2023-08-08(004). DOI:10.28270/n.cnki.ngjsb.2023.002955.

[22]温婷.图计算、大模型“各显神通”国产软件创新百花齐放[N].上海证券报,2023-05-09(006). DOI:10.28719/n.cnki.nshzj.2023.002032.

[23]艾丽达娜. AI赛道崛起农村中小银行向何处突破[J].中国农村金融,2023,(06):73-74.

[24]秦玉芳.从“在线”转向“在场”银行深掘AI新赛道[N].中国经营报,2023-02-13(B05). DOI:10.38300/n.cnki.nzgjy.2023.000198.

[25]刘海玲.科技助力经济高质量发展[N].中国会计报,2022-11-25(010). DOI:10.38301/n.cnki.nzgkj.2022.001506.
[26]姚琥. “AI+知识图谱”在信用风险管理领域的探索[J].金融电子化,2019,(05):55-57.

Paradigm Shift in Banking Compliance Governance Driven by Artificial Intelligence

WANG Chenye[1], ZHI Pei[1*], YU Yilian[2], ZHOU Xiqing[2], DING Xuan[3*]

(1. School of Fashion Management, Beijing Institute of Fashion Technology, Chaoyang, Beijing 110105, China; 2. School of Business, Jiangxi Institute of Fashion Technology, Nanchang, Jiangxi 330201, China; 3. School of Law, Fuyang Normal University, Fuyang, Anhui 236000, China)

Abstract: With the deep integration of technologies such as generative artificial intelligence, compliance governance in the banking sector is facing fundamental challenges arising from "algorithmic black boxes," model hallucinations, and data ethics. Based on representative practices in the global banking industry from 2024 to 2025, this study systematically analyzes the paradigm shift in AI-driven compliance governance—from "passive response" to "active embedding," and from "tool application" to "operating system-level capabilities." The analysis indicates that leading institutions are transforming compliance into a core competitive advantage by constructing automated governance frameworks covering the entire lifecycle, advocating for the ethical principle of "building trust through design," and promoting collaborative innovation within "regulatory sandboxes." The research suggests that the future competitiveness of the banking industry will depend not only on the diversity of technological applications but also on whether governance systems can internalize ethical constraints as technical foundations, thereby establishing a dynamic balance between innovation and stability.

Keywords: AI governance; Banking compliance; Model risk management; Full lifecycle governance; Ethical alignment; Regulatory technology (RegTech)