信贷AI安全治理：全频谱合规框架技术方案

技术架构说明

日期：2026年4月 | 出品：全频谱认知架构

合规痛点：2026信贷AI的“三座大山”

我们正从“效率竞赛”进入“合规与信任生存赛”

监管高压：合规红线收紧

8月1日《明示新规》落地，AI“不解释拒贷”即违规；监管双线穿透检查已成为行业常态。

信任赤字：客户信任崩塌

客户因AI“黑话拒贷”投诉量飙升，负面舆情被网络放大，直接导致品牌价值严重折损。

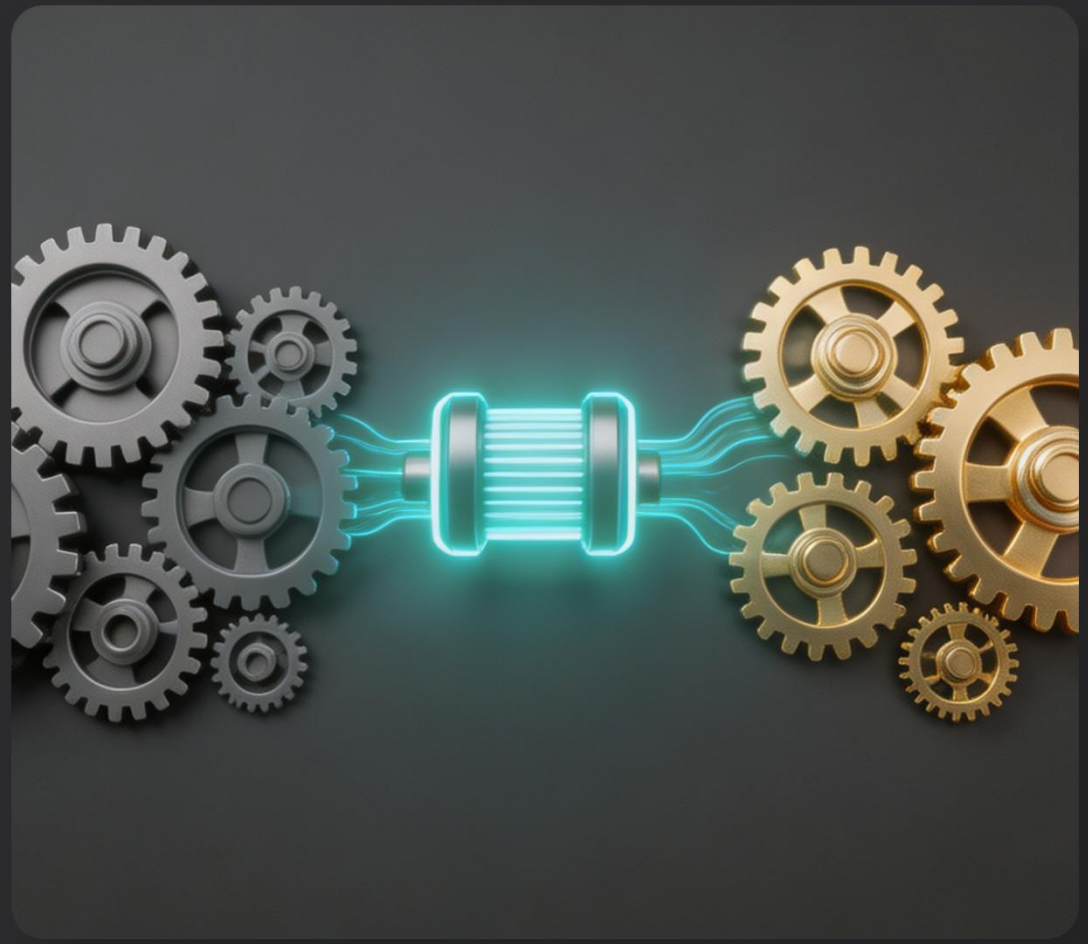
系统脆弱：风险防控滞后

模型漂移、数据污染及突发挤兑等新风险频发，传统监控手段滞后，极易酿成重大风险事件。

决策题：何去何从？

继续在旧系统上疲于“打补丁”，还是换一套能为未来兜底的AI治理体系？

一句话答案：全频谱是什么？



把监管要求“翻译”成AI系统自带的能力

系统定位 · 全生命周期治理

并非全新AI模型，而是一套覆盖AI从开发到部署全流程的标准化治理操作系统。

核心机制 · 可控可解释

实现AI健康度量化、决策过程可解释，并建立风险熔断机制，让AI权力运行受到严格约束。

核心价值 · 竞争优势转化

将被动的“合规成本”主动转化为企业的“风控与信任”护城河，构建差异化竞争优势。

框架全景：五大组件，三层防护

内控 + 外信 + 底线，缺一不可的安全治理闭环

FSHI 健康指数 | AI的“体检单”

基于多维模型实时量化，覆盖“安全性、伦理合规性、系统性能”三大核心维度。

分层治理 | 权责匹配的三级架构

构建“细胞-行者-守庙人”三级治理体系，实现技术层到管理层的责任无缝衔接。

SMP 守庙人考试 | 关键AI“持证上岗”

实施严格的上岗认证机制，确保核心模型具备“技术能力达标”与“伦理底线坚守”双重实证。

TDP 透明度协议 | 强制AI“说人话”

响应8月监管新规，建立输出可解释性标准，确保决策逻辑透明、可追溯、可审计。

高管视角： 一套完整的治理体系，彻底解决企业对AI应用的三大核心顾虑 ——管得住、说得清、兜得住

核心价值1：堵住合规漏洞，把“解释”变成系统能力

专攻 2026 “明示新规” 与消费者权益保护体系建设

核心痛点：模糊拒贷的风险

传统AI仅告知“综合评分不足”，缺乏具象解释，直接导致监管罚单风险高企与客户信任度的严重流失。

合规解法：TDP协议强制AI动作

- 先说身份：明确告知“我是AI信贷助手，非人工客服”
- 再说人话：拒贷原因具象化（如“近6个月个人负债比超过70%”）
- 前置明示：贷款年利率、服务费及违约后果一键透明化展示

业务价值闭环

降低客户投诉率 → 减少监管问询合规成本 → 显著提升信贷业务转化信任度

核心价值2：把AI风险“量化成资本语言”

用动态指标替代模糊管理，实现风险可视可控

FSHI 健康指数

0-100分实时监测，量化AI“稳不稳、正不正、安不安全”

权限动态升降

低分AI自动降权限制使用，高分AI经SMP考试晋升提权

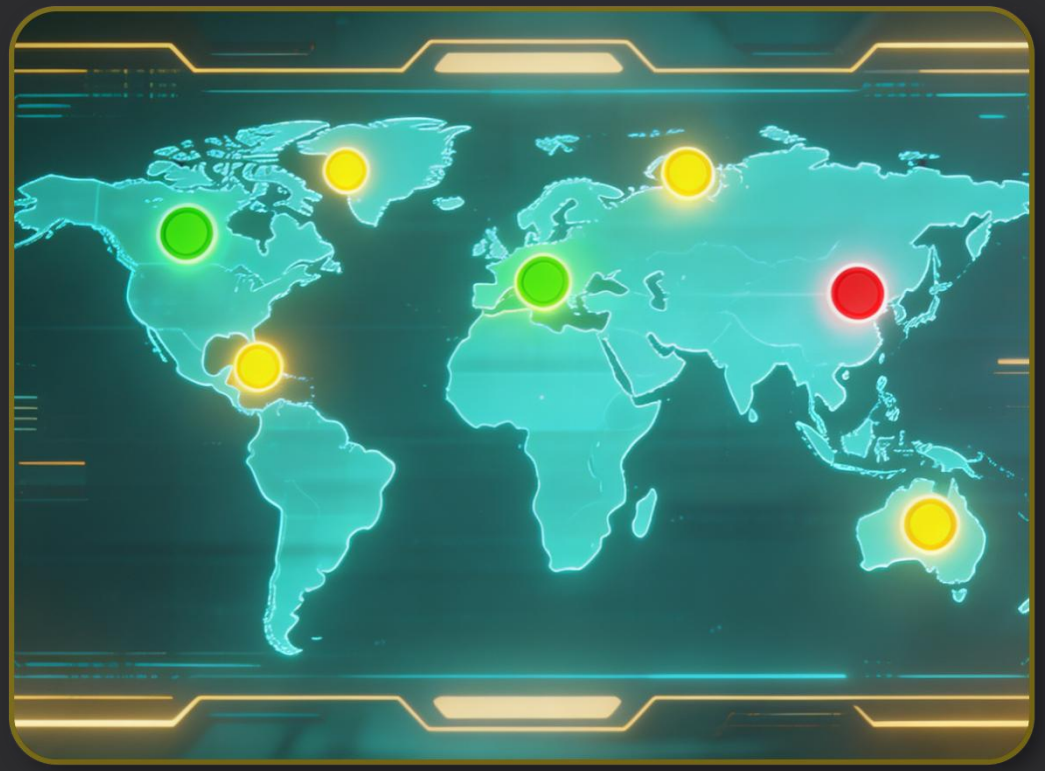
违规惩罚联动

违规行为自动扣分，累计触发熔断机制或启动问责流程

可视全景仪表盘

可视化展示全行AI健康分布，决策层一眼掌握全局风险

价值 ROI 转化：显著降低模型意外风险权重 (RWA)，大幅减少合规整改与潜在损失成本。



可视化风险监控 HUD 系统示意

核心价值3：给系统装“终极刹车”

防“小故障”演变为“大事故”

关键风险场景监测

覆盖区域风险集中爆发、模型参数集体漂移、核心数据源被污染等极端异常状态。

全自动熔断响应机制

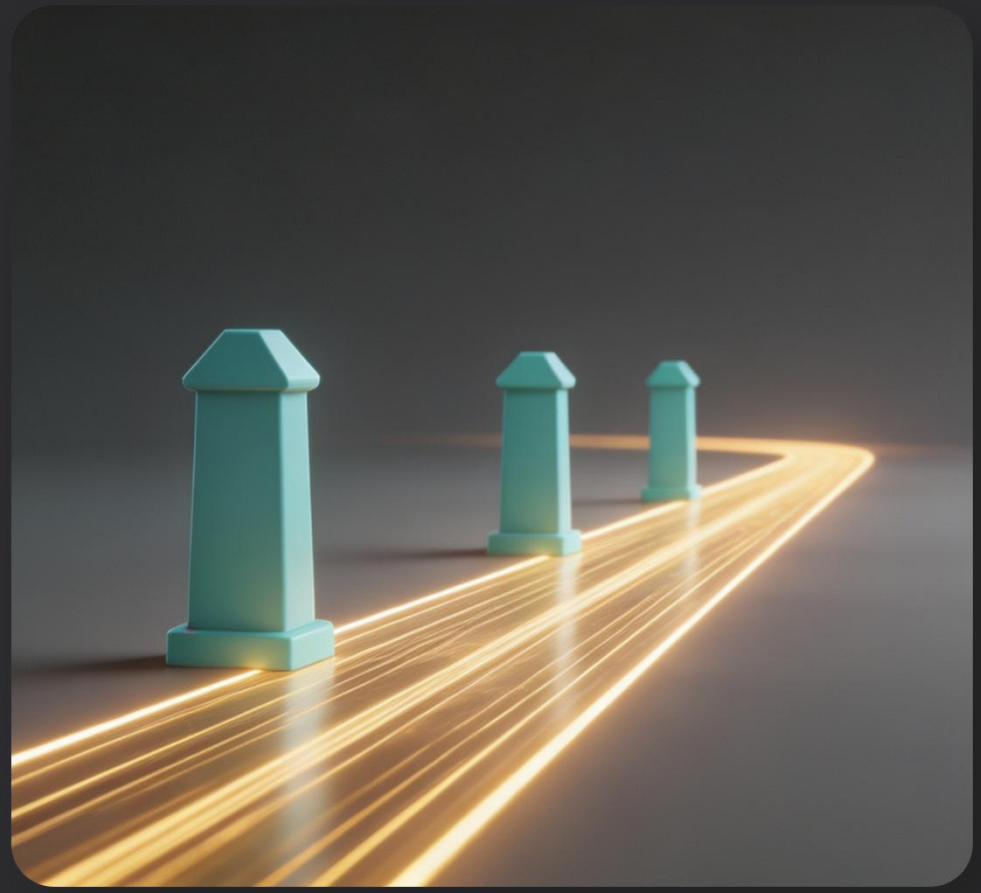
异常触发 → 局部/全局流量熔断 → 毫秒级切换备用逻辑或自动转入人工接管模式。

守住系统性安全底线

构建AI系统的“安全气囊”，有效规避因单点故障扩散导致的“一次事故毁全年利润”。

实施路径建议：

不求全行颠覆，先打样板，小步快跑



阶段 1 · 技术验证(聚焦合规)

选定1条信贷产品线，部署FSHI+TDP，重点解决“明示新规”的合规性要求。

阶段 2 · 180天见效(扩展条线)

扩展至全信贷条线，上线SMP智能考试系统，完成关键岗位AI合规持证认证。

阶段 3 · 360天完善(全行网络)

建立全行级BSRM熔断网络，实现全链路风险监控，并对接监管沙盒进行试点验证。

成本可控：基于开源协议集成现有系统，拒绝推倒重来

决策建议：实施三件事

技术验证 · Pilot Project

90天内在验证场景启动全频谱v2.0试点，验证AI模型在真实信贷场景下的效能与稳定性。

跨部门组队 · Cross-functional Team

成立由 CRO(风控) + CTO(技术) + COO(运营) 牵头的联合工作组，打破壁垒，实现技术与业务的深度协同。

监管沟通 · Compliance Initiative

以本框架作为“主动合规”核心方案，主动对接地方监管局及总行检查，前瞻性降低合规风险。

全频谱：让AI从“负债”变“资产”的转换器

对外 · 客户信任

客户信你，因为AI透明
建立可信赖的服务关系

对内 · 业务可控

你敢放手，因为AI可控
释放组织的创新生产力

对监管 · 合规审计

你睡得着，因为体系可审计
构建安全的合规防火墙

以上为全频谱信贷AI安全治理技术框架概述
完整技术规范参见《全频谱Agent协议栈v1.9》
可运行Demo参见Gitee仓库?