ESG CONSULTING

负责任人工智能与 AI 道德使用

Responsible AI & Ethical Use

远东宏信 · 员工培训

Far East Horizon · ESG Training

合规•公平•安全

金融+产业场景下的 AI 责任边界

培训对象	远东宏信员工
配套服务	行动清单
培训目标	知道、识别、找谁

六个章节，一份责任

Responsible AI Training · Table of Contents

合规是底线 • 公平是核心 • 安全是基石 —— 每个人都是 AI 责任的守门人

01

第 一章

为什么要学

当一笔贷款申请被 AI 预审通过——你敢直接放款吗？

公司 AI 已深入业务 — 从信贷到产业的全面渗透

金融 + 产业 + 运营 + 员工赋能，AI 已成为日常工作的底层能力

AI 部署关键数据

10
秒生成预审报告
「远鉴」AI 预审速度

50%
审核时效提升
近 50% · 客户经理周审批量翻倍

11
套等保认证
AI 系统数据安全底座

AI 已渗透四大业务板块

金融业务
远鉴风控 · AI 审核助手

产业运营
医疗 AI · 教育 AI · 设备 AI

内部运营
智能财务 · 人力 AI

员工赋能
远东 AI 助理 · AI 陪练

AI 治理关键数据

22,228 人
员工总数（培训覆盖）

11 套
等保认证系统

100%
数据安全培训覆盖

0 起
重大信息安全事件

4,000+ 家
普惠金融中小微客户

SO WHAT • 为什么是现在

AI 已从工具变成决策伙伴。22,228 名员工的每一个日常动作都连着 AI 系统的责任边界——理解这个边界，是本次培训的核心命题。

培训的三个目标 — 知道、识别、找谁

这次培训解决三件事，让你下班前都能用得上

01

Know the rules

知道规定

公司在 AI 方面有哪些红线、哪些边界、哪些审批要求

- DeepSeek 三层权限架构
- 11 套系统等保认证
- 委员会监督

02

Spot red lines & gray zones

识别红线和灰区

日常工作里哪些动作合规、哪些违规、哪些模糊地带容易踩雷

- "四必须" 行为规范
- 5 工具 × 4 权限维度
- 4 个红色禁区场景

03

Know who to call

找对人、找对路

出问题或看到可疑时，按什么流程上报，找谁处理，多久闭环

- 逐级申诉流程
- AI 治理小组
- 信息技术团队兜底

本次培训总原则：合规是底线 • 公平是核心 • 安全是基石

02

第 二章

我们身边的 AI

金融+产业+运营+员工——AI 到底渗透到了哪里?

AI 应用全景 — 金融 / 产业 / 运营 / 员工赋能

「远鉴」风控 · AI 审核助手 · 医疗 AI · 教育 AI · 智能财务 · 远东 AI 助理

金融业务

「远鉴」智能风控

AI 预审 · 客户画像 · 尽调助手 · 监控预警

AI 审核助手

信用评估报告智能生成

AI 智能助手

行业动态推送 · 融资方案自动生成

普惠金融

累计近 4 万家中小微企业、超 960 亿元

产业运营

医疗 AI

智能病历质控 · 慢病管理 · 医院资源监控

教育 AI

AI 评语 · 精准教学 · AI 备课

设备运营 AI

AI 多语言信息沟通，支撑全球化

应用规模

覆盖医疗、教育、设备三大产业

内部运营

智能财务

合同比对 · 异常交易预警 · 报销稽核

人力 AI

AI 陪练系统 · 7×24 智能咨询

远东 AI 助理

全员智能问答、报告公文生成

覆盖范围

已面向全员推出

员工赋能

新员工培训

AI 陪练系统模拟真实业务场景

日常办公

远东 AI 助理辅助检索与公文

合规学习

本次培训及延续课程

数据安全

员工培训覆盖率 100%

SO WHAT • 理解 AI 的广度

AI 不是一个工具，而是覆盖金融、产业、运营、员工赋能的「能力底座」。在远东宏信，AI 是效率工具，但不替代专业判断——这是后续所有章节的认知前提。

DeepSeek 本地化部署 — 数据不出公司

AI 关键里程碑

- 2025-04 DeepSeek V3 与 R1 本地化部署
- 2025 AI 三层应用架构搭建
- 2025 与阿里、字节、腾讯、百度、月之暗面合作

本地化部署的三大价值

①	**数据不出公司**	客户信息安全从架构层得到保障
②	**权限分级管控**	谁看什么、谁用什么、用什么数据
③	**三层架构可治理**	模型层 / 中台层 / 应用层可审计可追溯

AI 三层应用架构

应用层 • Application

远鉴风控 · 远东 AI 助理 · 业务 AI

中台层 • Middle Platform

能力沉淀 · 权限管控 · 数据中台

模型层 • Foundation Model

DeepSeek V3 / R1 本地化部署

▲ *上层调用下层能力，下层为上层提供基础*

SO WHAT • 数据安全底盘

本地化部署意味着客户数据始终在远东宏信内部流转——这既是合规优势，也是后续权限与边界（第三章）能够落地的基础设施前提。

03

第 三章

权限与边界

同一个「远东 AI 助理」，不同岗位看到的可用功能完全不同——凭什么？

三层架构、分级管控 — DeepSeek 权限全景

敏感层 · 业务层 · 基础层 —— 越敏感越需要专门审批

各层可用工具与典型功能

敏感层	「远鉴」智能风控（信用评估、客户画像） · 涉及客户财务数据的 AI 分析 · 涉及个人信用信息的 AI 处理
业务层	行业动态推送 · AI 审核助手 / 尽调助手 · 合同比对 / 报销稽核
基础层	远东 AI 助理（日常问答、文档辅助） · DeepSeek 基础功能（非敏感信息处理）

SO WHAT

权限与责任

权限不是束缚，是让专业的人做专业的事。同一个 AI 系统，在不同岗位、不同业务下，看到的功能和审批要求都不一样——这是对客户、对你自己的保护。

四个「严禁」— 红线即底线

任何一条触及，都会进入信息安全事件问责流程

① 严禁将客户财务数据、个人信用信息输入未经授权的 AI 工具

客户财务和信用信息属于最高敏感级别；外发即触发数据安全事件。

② 严禁使用公网 AI 工具处理公司业务数据

公共版大模型、第三方 AI App 均不属授权范围；本地化部署是唯一合规路径。

③ 严禁将 DeepSeek 本地化部署中的敏感业务数据导出至外网

本地化优势在于数据不出公司，导出即丧失合规基础。

④ 严禁代他人使用或被他人代用自己的 AI 权限账号

账号即责任主体，借账号等于把责任转嫁他人和自己。

违规后果： 信息安全事故问责机制启动 • 11 套等保系统全程留痕

SO WHAT • 红线即底线

四条严禁不是规则清单，而是远东方洪信对客户、监管和自己的三重承诺。触线即问责——这不是恐吓，是底线。

04

第 四章

AI 的信任与质疑

当 AI 给出「合理」结论——你怎么判断它是「正确」还是「自信」？

透明与可解释 — 标注、推理、客户知情

AI 输出的可追溯性，是信任的最低门槛

① AI 生成内容 | ② 标注「AI 辅助生成」 | ③ 附推理依据 | ④ 客户知情 + 人工复核

①

AI 生成的所有报告/建议必须标注

「由 AI 辅助生成」不是免责声明，而是职业诚信声明。

②

AI 信贷决策必须有可追溯的推理依据

不能只给结论；模型用了哪些数据维度、什么权重，要可解释。

③

客户有权知道决策环节是否使用 AI

客户对 AI 决策有知情权；这是金融监管的硬要求。

④

标注 + 复核 = 完整可追溯链

任何 AI 结论进入正式决策前必须经过人工复核。

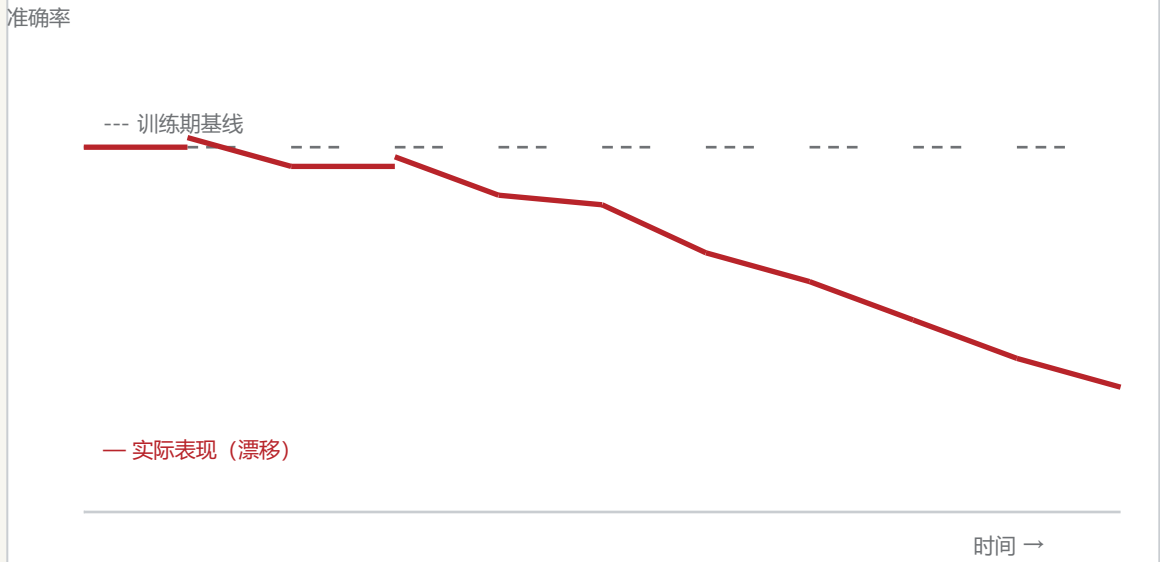
SO WHAT • 透明度 = 信任

AI 不是黑箱，也不是替罪羊。标注 + 推理 + 复核，让 AI 决策可解释、可追溯、可问责——这是金融行业对 AI 的最低要求。

持续监控 — 模型不是一劳永逸

什么是「模型漂移 Model Drift」？远鉴如何发现并上报

模型漂移示意



提示：实际业务场景随时间变化（监管、客群、宏观），模型准确率会逐渐偏离基线——这就是「模型漂移」。

员工日常关注 3 个信号

① 同一客户的 AI 评估结果前后矛盾

② AI 推荐的融资方案明显不合理

③ AI 预警频繁触发但经核实多为误报

发现异常时的上报路径

你 → 部门主管 → 风控/技术团队 → AI 治理小组

任何环节都可以启动；越早上报，越早止损。

SO WHAT • AI 也需要校准

远鉴上线不是终点，是起点。模型漂移是 AI 系统的常态——关键是我们有没有持续监控、有没有让员工在日常中成为「早期预警」的眼睛。

公平与偏见 — 信贷决策的反偏见（本章重点）

AI 没有立场，但 AI 学的是历史——历史里的偏见，会被模型继承

AI 偏见从何而来

- **A** 历史数据中隐含的歧视性模式
- **B** 地域/行业/规模训练数据不均衡
- **C** 「幸存者偏差」— 只学了获贷客户
- **D** 特征工程的隐性变量选择

金融偏见的严重后果

- ⚠ 特定地区/行业/规模企业被系统性低估
- ⚠ 违反金融监管的公平信贷要求
- ⚠ 客户信任丧失、舆情与诉讼风险
- ⚠ 监管处罚、罚款与业务限制

远东宏信的做法

- ✓ 多元化客户画像维度设计
- ✓ 风险模型定期接受公平性审查
- ✓ 人工复核作为偏见纠正的「最后防线」
- ✓ 员工主动上报机制（任何环节）

员工能做什么

- → 发现系统性高风险，主动质疑
- → 尽调中补充「软信息」
- → 对 AI 评估有异议，按申诉流程提报

SO WHAT • 公平是金融 AI 的核心命题

效率可以由 AI 提效，但公平必须由人来守——特别是当模型从历史中学习时，员工的质疑和软信息是 AI 系统学不会的「最后一道防线」。

质疑与申诉 — 当 AI 决策被挑战

谁可以提、什么场景、什么流程、多久闭环

逐级申诉流程

① 提出	客户、业务、风控或任何发现问题的员工
② 部门主管	第一时间接收并初步判断
③ 风控/技术团队	技术与业务事实核查
④ AI 治理小组	治理层面裁决与改进

处理原则：有记录 · 有回应 · 有结论 · AI 相关投诉时效不高于人工决策

案例：AI 误判与人工纠偏

一家制造企业的贷款申请被「远鉴」AI 预审为「高风险」并自动拒绝。

客户经理在实地尽调中发现：AI 未考虑该企业的核心技术专利和近期订单增长。

按申诉流程提交 → 人工复核 → 调整为「中风险」→ 放款 → 事后回访正常。

反例速写：

另一家企业 AI 评估为「低风险」快速放款，因未充分人工复核关键风险点导致不良。

→ 申诉不是找麻烦，是 AI 持续改进的重要输入。

SO WHAT • 申诉是 AI 持续改进的输入

每个申诉都让 AI 模型更接近真实业务。越多人愿意提、敢于提，AI 系统的偏见和盲区就越早被发现和修正。

05

第 五章

AI 与可持续发展

AI 是高能耗技术——但它也能为可持续经营贡献巨大价值。

AI 为可持续经营贡献了什么

降不良 · 提普惠 · 优医疗 · 强教育 — AI 价值的社会维度

降不良

Reduce NPL

近 50%

审核时效提升

远鉴风控拦截潜在不良贷款，降低不良率 → 资产质量提升 → 可持续经营能力增强。

提普惠

Inclusive Finance

4 万家

中小微企业

累计服务近 4 万家中小微企业，超 960 亿元资金支持，让金融活水精准灌溉。

优医疗

Better Healthcare

3+

医疗 AI 场景

智能病历质控、慢病管理、医院资源监控 — 提升医疗服务效率，患者直接获益。

强教育

Smarter Education

4 项

教育 AI 应用

AI 评语、精准教学、AI 备课、智能升学规划 — 优化教育资源配置，创造社会价值。

SO WHAT • AI 的可持续价值

AI 不是单纯的「技术秀」，它正在转化为远东宏信可持续经营能力的一部分——降风险、惠小微、优医疗、强教育。

绿色用 AI — 节能意识与能效责任

每多算一次，都是能源消耗。理解并践行「按需用 AI」

日常 AI 使用 3 条节能意识

① 按需使用，不做无意义重复调用

能一次问完的不要分十次；能用模板的不要每次重新生成。

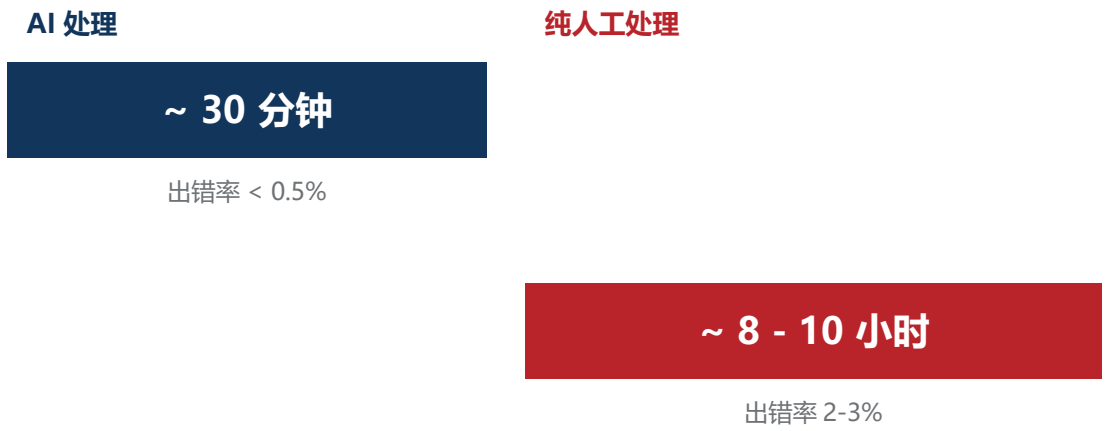
② 理解并尊重集中部署的能效优化

公司 AI 集中调度是为了降低整体能耗；统一时段使用更环保。

③ 培训完成后关闭不必要的 AI 会话窗口

会话窗口占用资源；用完即关，是日常最基本的能效动作。

合同比对：100 份的人机对比



示意数据：100 份标准合同条款比对的相对参考。

时间成本 ↓ 95% | 出错率 ↓ 80% | 能源消耗 → 集中优化

SO WHAT • 绿色 = 高效 + 集中

绿色用 AI 不是要求大家不用 AI，而是用得聪明、用得集中。集中部署 + 按需调用，是远东宏信能效优化的核心动作。

06

第 六章

日常 AI 行为规范

会做更要会防——今天哪些事绝对不能做?

AI 使用「四必须」 + 工具边界速查表

凡事先问四个问题，再打开 AI 工具

① 必须

工具在公司授权清单内

远东 AI 助理、DeepSeek 内部、远鉴、AI 审核助手

② 必须

数据不含客户敏感信息

财务、信用、个人隐私

③ 必须

对 AI 输出进行人工验证

复核、追问、必要时推翻

④ 必须

AI 生成内容明确标注

「由 AI 辅助生成」

5 工具边界速查表

AI 工具	权限层级	可以做	不可以做
远东 AI 助理	**基础层**	日常问答、文档辅助	输入客户财务数据
「远鉴」风控	**敏感层**	按权限查看风险报告	将风控结论外传客户
AI 审核助手	**业务层**	辅助生成报告初稿	不经复核直接提交
DeepSeek	**分层**	权限内数据处理	上传客户信用信息
公网 AI 工具	**— 严禁**	个人学习、非工作场景	任何工作数据处理

SO WHAT • 操作指南

「四必须」是行动清单，速查表是日常使用手册。下次打开 AI 工具前，先问自己：工具对吗？数据对吗？复核了吗？标注了吗？

高警示场景 — 4 个红色禁区

这 4 种操作一旦发生，进入信息安全事件问责流程

把企业客户的财务报表截图发给公网 AI 分析

外发即触发数据安全事件；客户财务信息属于最高敏感级别。

用个人手机上的 AI App 处理公司内部文件

未经授权的第三方 AI 工具不在授权清单内；本地化部署是唯一合规路径。

在出差时通过不安全的网络使用远东 AI 助理处理客户数据

公共 WiFi / 不可信网络存在中间人攻击风险；敏感数据应在可信网络环境处理。

把同事的 AI 系统账号借来用

账号即责任主体；借账号等于把责任转嫁他人和自己，违规即双责。

两位客户经理面对同样的 AI 工具——一位合规使用走完 5 年晋升；一位违规外发被问责——差距就在每一个日常动作。

SO WHAT • 红色禁区

这些不是建议，是禁止。培训覆盖率 100% 的延续，是你我都熟悉这些场景、共同守住底线。

六大行动准则 — 一图总览

一页带走，回去贴在工位上

① **定位清晰**

AI 是效率工具，人不推责

② **分层管控**

DeepSeek 三层架构，权限分明

③ **公平底线**

信贷决策反偏见，客户平等

④ **透明可溯**

AI 输出标注，决策可解释

⑤ **持续监督**

模型监控，异常即报

⑥ **绿色用 AI**

按需调用，不浪费算力

合规是底线 · 公平是核心 · 安全是基石 —— 每个人都是 AI 责任的守门人

Far East Horizon · Responsible AI · Six Principles

从今天起 — 6 条每日承诺

合规使用 AI，从每一个日常动作开始

- ✓ **每次使用 AI 工具前**
  确认在公司授权清单内
- ✓ **绝不将客户财务数据**
  或个人信息输入公网 AI
- ✓ **「远鉴」AI 预审结果**
  必须经过人工复核
- ✓ **AI 生成内容**
  必须标注「AI 辅助生成」再提交
- ✓ **发现 AI 输出异常**
  或可疑偏见时主动上报
- ✓ **不将自己的 AI 账号**
  借给他人使用

今日所承诺的，是明日所收获的

THANK YOU

合规是底线
公平是核心
安全是基石

—— 每个人都是 AI 责任的守门人

问题上报路径

部门主管 → 信息技术团队

感谢您完成本次培训 · 2026