附件二：第四届中国研究生金融科技创新大赛“揭榜挂帅”赛题指南

大赛筹备阶段，承办方通过多轮深度沟通对接区域金融创新需求，从 30 余个企业提案中精选出 15 个兼具技术创新性、落地可行性与行业应用价值的企业“揭榜挂帅”赛题。赛题共包括大数据、人工智能、图像识别三大类别。

一、大数据

赛题 1：基于 Transformer 架构的信贷风控模型研究

赛题背景：赛题来源于苏银凯基消费金融有限公司。消费信贷金融机构信贷风险控制高度依赖风险数据与风控模型。人工智能算法作为风控模型的核心组成部分，影响着信贷风控审批的效率与精度。本赛题旨在解决信贷风险评估过程中，需要分析不等长的时序数据和非结构化的离散数据的技术堵点问题。

赛题要求：

总体目标：基于 Transformer 算法架构开发信贷风险评估模型。

技术任务：

① 传统的机器学习算法很难处理不等长的时序数据，难以挖掘某些特征在不同时间维度下的内生关系。需要基于 Embedding 技术和 Transformer 架构解码器部分的蒙版特性，解决传统时序模型无法处理不等长时序数据问题。

② 传统的风控模型通常对于训练数据本身的要求较高，无法处理非结构化数据、时序特征数据、稀疏数据和弱关联特征数据。需要将非结构化数据、时序特征数据、特征稀疏数据和弱关联特征数据通过 Embedding 一并纳入训练，学习数据特征之间关联关系。

③ 客群分类问题下，无法直接使用基于 LLM（large language model），在此场景下的 Transformer 算法需要改造 Transformer 架构，使其适配于信贷风控场景。

配套数据情况：

数据规模：训练集和测试集样本量 120 万条，违约样本量 12 万条；跨时间验证集样本量 60 万条，违约样本量 3 万条。

赛题 2：基于多场景预测的人伤自核引擎系统优化研究

赛题背景：赛题来源于紫金财产保险股份有限公司。保险理赔中，车险、意外险的人身伤害案件估损审核复杂且争议多，业务流程涵盖报案查勘、医疗费用审核、伤残等级评估、赔偿计算。当前从医疗费审核到赔偿计算人工操作多，存在效率低、主观性强、全流程自动化不足的问题。本赛题旨在解决人身伤害案件的估损审核复杂且争议多，需要形成标准的自动化定损工具的现实问题。

赛题要求：

总体目标：开发能够根据人身伤情状况生成合理赔偿区间的自动化工具。建立完整的人伤智能估损体系，实现从动态阈值生成到实时预警的全流程自动化。

技术任务：

① 根据保险理赔人伤案件历史数据，包括伤情诊断、治疗费用明细、伤者职业信息、地区经济数据（如平均工资、医疗成本）等，构建多维度伤情-费用关联特征库。

② 基于伤情、职业、地区等核心因素，将不同伤情按严重程度、治疗方式等维度细分群体，分别计算 95%分位数作为阈值上限，精准刻画各群体费用分布特征。

③ 基于生成的动态阈值，搭建实时监控体系，设计数据实时处理流程，对每一笔新产生的案件费用数据进行即时分析，判断是否超出阈值。当案件费用超出阈值时，按照超出比例和风险程度，划分为黄、橙、红三色预警等级。例如，超出阈值 10%-30%为黄色预警，30%-50%为橙色预警，50%以上为红色预警。

④ 最终建立完整的人伤智能估损体系，实现从动态阈值生成到实时预警的全流程自动化。

配套数据情况：

样本量：训练集 8000 条，测试集 2000 条，均为真实脱敏业务数据。结构化数据 29 项（数值型 12 项、分类型 17 项），短文本数据 1 项

时间区间：2020 年 1 月-2025 年 5 月。

赛题 3：对公贷款逾期预测模型构建与应用

赛题背景：赛题来源于苏州农商银行。对公贷款是银行向企业客户提供的资金支持，但是企业无法按时偿还贷款本息的情况时有发生，导致贷款逾期，给银行带来巨大的信用风险和财务损失。目前银行主要依赖人工贷前审查、财务报表分析以及历史信用评级等方式进行风险评估，但这些方法难以准确捕捉企业在贷后阶段的动态变化，缺乏前瞻性预警能力。本赛题旨在解决贷款客户的逾期概率难以提前判别，催收工作针对性不足的现实问题。

赛题要求：

总体目标：在逾期样本占比较少的前提下，构建能够识别客户逾期行为的预警模型。

技术任务：

① 攻克小样本、高不平衡比条件下的模型泛化能力弱和对少数类识别能力差的技术难题。构建具备强泛化能力的风险预测模型，在有限正样本下有效识别潜在高风险客户。探索并应用先进的样本重平衡策略、代价敏感学习方法以及集成学习框架以缓解类别不平衡带来的偏差；同时通过交叉验证、早停机制和模型稳定性评估手段，确保模型在实际业务场景中的鲁棒性和可解释性。

② 解决结构化数据特征表达力有限和高维特征空间噪声干扰的问题，实现高效、鲁棒的风险特征提取与压缩。研发面向金融时序数据的自适应特征构造算法：自动或半自动生成刻画趋势（如滑动平均、斜率)、波动性（如标准差、变异系数)、异常（如 Z-score 异常检测）的时序衍生特征；结合信贷风控领域知识，设计具有业务解释性的复合特征；探索有效的特征交互与组合方法，捕捉变量间的非线性关系；用嵌入式特征选择进行初步筛选利用包裹式方法结合最终模型性能进行特征子集优化。

配套数据情况：

样本量：训练集 8 万条（违约样本 8500 条)，测试集 2 万条（违约样本 300 条)，无合成数据，均为真实脱敏业务数据。

时间区间：2021 年 1 月-2025 年 5 月。

赛题 4：基于小样本与数据血缘的银行业务数据智能分类分级系统

赛题背景：赛题来源于苏州农商银行。长期以来，银行的各个业务系统（如信贷系统、电子渠道）和管理系统（如 CRM 系统）积累大量敏感数据。当前银行内部的数据分类分级主要依赖人工规则与静态标签，面临的重要困境包括新增业务场景难以建立有效分类分级模型；数据经过其他系统的接口调用、加工处理和流转后，血缘关系模糊、难以进行精确的数据流向追踪；人工维护响应滞后，导致审计风险暴露率高。本赛题的核心诉求在于：构建自动化系统，利用有限的样本数据与跨链路血缘关系，实现数据资产的精准分类分级与动态更新，从而实现数据安全的差异化安全防护和合规管理。

赛题要求：

总体目标：开发自动化的银行业务数据智能分类分级系统。

技术任务：

① 构建支持小样本情况下的分级分类模型。使得在数据样本不足的情况下（例如，每类数据的样本亮≤500 条），能达到准确率≥90%。

② 解析数据库表结构和 ddl 脚本、数据 ETL 调度脚本、调度作业依赖关系、API 调度接口等信息，提取相关实体（表/字段/API）的流转血缘关系，构建关系图谱。生成图谱构建工具（支持 Neo4j/GraphDB 等导入）。

③ 开发基于机器学习/人工智能的分类分级动态调整模型，自动优化数据分类分级策略。当数据的血缘拓扑变化时，能自动优化数据分类分级策略，准确率达到 90%以上。

配套数据情况：

数据资源：

样本量：3000 表 10 万个字段及样本数据（含表结构和 1 万条已分类分级结果。

数据库数据：包括表及字段结构、表样本数据。

时间区间：2025 年 1 月-2025 年 6 月。

工具支持：

提供数据库测试环境、数据分类分级领域的国家标准规范。

赛题 5：基于知识图谱实现寿险黑产团伙挖掘的研究

赛题背景：赛题来源于利安人寿保险股份有限公司。保险公司长期面临专业化、组织化的骗保团伙（黑产团伙）威胁。黑产团伙作案具有隐秘性高、作案手段更新快等特点，但传统反欺诈手段依赖人工规则和单点数据分析，需构建一套由多方数据组成的知识图谱平台，实现团伙作案模式实时识别，提升新型骗保案件挖掘效率，提升团伙作案准别率。本赛题旨在解决寿险黑产团伙作案手段多样且持续更新，人工反诈手段识别效率不高的现实问题。

赛题要求：

总体目标：构建可解释且能够动态识别黑产团伙并提供证据链的知识图谱系统。

技术任务：

① 统一时空维度，综合建模结构化、非结构化数据（如就诊记录、理赔记录、医疗机构信息等）。

② 构建具有动态模式识别能力，自适应发现新特征的反欺诈识别模型，当黑产团伙手法升级，模型也能够快速自动学习，识别新型作案手法。针对周期性新增的动态数据（如每季度新增关联手机号、银行卡，案件），能够快速准确识别异常子图突变。

③ 使反欺诈识别模型具有可解释性，即能够提供黑产团伙证据链。研发证据链可视化引擎，自动生成涉案人员/事件/资金/医疗机构的关联时序图和关系网络图，实现 5 度关系快速响应的知识图谱查询。

④ 能够支持但不限于通过规则引擎动态调整模型策略，降低短时低信噪比场景误判率。够识别非结构化数据中的表单、图片等数据，并将有价值的数据导入图谱模型中。在数据类别比例差异大的情况下，通过对模型的优化，提高小类别数据的识别准确性，解决少数类别信息淹没、评估指标失效等问题。

配套数据情况：

样本量：训练集 8 万条（违约样本 2000 条），测试集 2 万条（违约样本 300 条），均为脱敏业务数据。

时间区间：2022 年 6 月-2025 年 6 月。

赛题 6：基于极不平衡样本的客户金融服务产品多步推荐模型

赛题背景：赛题来源于无锡农村商业银行。社保卡发卡行的百万级个人客户中，绝大多数持有业务为社保卡即储蓄账户，而储蓄之外的其他产品，持有客户样本占比低，除无差别营销外，只能从已持有相关业务的客群中进行挖掘，未能从客户特征方面形成有效的拓新模型或策略，新建设产品后难以参考历史营销数据将新产品融入推荐策略。本赛题旨在解决社保卡发卡行客户众多但其他业务较少，产品推荐效果不足的现实问题。

赛题要求：

总体目标：构建具备多步推荐能力的推荐模型。

技术任务：

① 开发能够考虑个体在社会中的发展逻辑以及金融业务间的属性差异的推断性预测模型。基于客户在各个产品上的行为表现，结合客户自身的社会属性，给出一个预测模型，并且支持在给出不同的反馈结果后给出第二步推荐产品。

② 定义出有效的营销机会标签，使得在数据支持很少的情况下给出有效的差异化推荐并针对客户响应情况给出下一步推荐产品。基于客户的属性标签和尽量少的营销反馈数据组合定义出有效营销标签。

③ 使模型具有冷启动功能，在推出新的金融产品时，能够将新产品融入推荐策略中。解析新金融产品的属性，建设产品间的互相替代、补充、相反的关系，在给出新产品属性时，能够基于存量数据圈选出推荐客户群。

配套数据情况：

样本量为百万级，时间区间 2024 年全年，真实脱敏数据（连续型指标做分段处理）。

给出客户基本信息表，量级百万级。给出样本客户 2024 年全年业务事件明细，流水表中业务类型区分信贷、财富、支付、渠道，具体事件区分为开立事件、交易事件、关闭事件、开通绑定事件，给到关联产品，并独立给出产品的属性清单（包含信贷/财富、价格、期限、目标客群覆盖率、是否新产品等），其中开立事件代表新业务，标记为成功样本。

赛题 7：基于银行流水数据的黑灰产中介挖掘模型构建与应用

赛题背景：赛题来源于江阴农商银行。黑灰产中介团伙通过操控大量银行账户或支付工具，以“分散转入-集中转出”“高频小额交易”“快速资金归集”等模式从事洗钱、诈骗分账、非法集资等犯罪活动，这些活动隐蔽性高，难以排查。本赛题旨在通过构建自动化、智能化的流水交易分析方案，精准识别中介团伙的核心账户与资金链路，提升金融机构对黑灰产的主动防御能力。

赛题要求：

总体目标：基于流水数据，识别黑灰产中介的核心账户与资金链路。

技术任务：

① 交易网络规模庞大（百万级节点），且包含大量正常交易噪声，传统社区发现算法难以高效提取高关联性的黑灰产团伙子图，存在计算效率低、误判率高的问题。开发基于改进 Louvain 算法或并行化标签传播（Label Propagation）的团伙检测方法。

② 针对黑灰产资金交易中存在的动态行为模式演化问题（如"快进快出"向"长周期洗钱"转变），构建一个融合多维度特征的动态风险评估模型，在保证实时性的前提下，准确识别核心中介账户，避免误判。通过融合动态行为分析与网络拓扑特征的综合方法，实现对黑灰产账户的精准检测。

③ 直观展示资金流向、团伙层级，辅助人工研判，避免“黑箱”结论。开发交互式可视化系统，支持动态过筛、关键链路高亮等功能。

配套数据情况：

数据来源：

真实交易流水记录，已脱敏处理。

样本量与字段：

账户表（Account.csv）：万条记录，含 account_id, register_ip, device_id, phone 等字段。交易表（Transaction.csv）：万条记录，含 txn_id, from_account, to_account, amount, timestamp 等字段。

时间区间：

2022 年 1 月-2024 年 1 月数据，包含工作日/节假日交易波动。

二、人工智能：

赛题 8：非平衡数据下的电诈交易挖掘

赛题背景：赛题来源于南京银行。电信网络诈骗（以下简称“电诈”）已成为威胁客户资金安全的主要风险之一。在金融反欺诈领域，银行面临两大核心挑战：时间差滞后与样本极端不平衡。无法捕捉早期潜伏期（账户正常活动）与试探期（小额测试交易），即无法在事前发现异常交易。并且，电诈分子占比极低（约 0.1%），导致实时规则命中的准确率较难提高，误报较高。本赛题旨在解决电诈分子在客户样本中占比极低，反诈监测准确率低的现实问题。

赛题要求：

总体目标：构建一套动态智能反欺诈监测体系，能够在极低黑样本比例和无时间标注的条件下，精准识别电诈账户及其异常交易的时间区间，并提供欺诈行为的关键特征（如“短时间内跨多省份 IP 登录+集中转出至同一陌生账户”等），支撑风控人员决策。

技术任务：

① 基于少量标注样本，构建能够识别异常交易起止时间的人工智能模型。利用 20%标注样本（含异常交易起止时间），构建人工智能模型，预测异常交易起止时间，要求预测时间误差≤±1 天。

② 使模型具备融合交易流水、行为埋点等特征数据的能力，构建跨渠道时序关联特征，解决少数样本被淹没的问题。构建异常交易识别特征库，从交易流水（金额、频率）和行为埋点（登录 IP、设备指纹）等大类数据中提取多粒度特征（小时级聚合统计+滑动窗口波动检测），利用如图神经网络（GNN）等建模操作识别关联性。

配套数据情况：

交易流水数据：包含 10 万个正常账户和 1,000 个电诈账户的完整交易记录（时间戳、金额、对手方等）。

行为埋点数据：登录设备、IP、操作路径等秒级行为日志。

标注样本：200 个黑样本（20%）已由业务专家手工标注异常交易起止时间（精确到天）。

时间区间：2024 年至 2025 年，均为真实数据。

赛题 9：基于 LLM 的账务智能体-财枢智擎

赛题背景：赛题来源于苏州银行。银行在账务检索和账务核对存在三个关键的效率问题：非财会专业新人入门难；账务检索时会计管理人员根据分录反向穿透到账号交易明细有难度；若账务出现总分不平，因可能跨多个系统的原因，财务人员无法快速定位是哪笔交易。本赛题的核心在于实现交互式账务术语问答、账务多层检索问答、账务问题定位分析以提高账务管理人员工作效率。

赛题要求：

总体目标：构建一套基于大语言模型对账务数据特征训练后的智能体。

技术任务：

① 使用知识获取与构建技术搭建账务知识库，依据 LLM 平台，构建知识库问答智能体。做到能够精准的描述账务体系的基础组成，以及其他关键术语的概念及作用。能够通过文本对话的方式指导刚接触账务的员工快速学习行内账务相关知识，降低员工频繁查阅不同文档造成时间成本

② 针对超过 20 轮复杂对话中的信息衰减问题，构建可持续跟踪的业务意图理解。开发记忆强化网络，通过注意力权重动态筛选关键信息，解决长对话中的信息衰减问题。构建干扰信息过滤模块，自动识别并忽略与业务无关的闲聊内容。能够通过对话的方式获取到多维度的信息，降低会计使用人员在维度信息获取上的沟通成本。

③ 使用推理与决策技术根据总分不平科目，检索出对应的日期、机构、币种、金额，再进一步往上追溯该科目相关的发生明细，进而推理出总分不平的可能性原因。根据总分核对异常结果，能够进行深度思考并通过实际数据给出可能造成的原因，降低会计人员定位总分核对异常的分析时间成本。

配套数据情况：

数据规模：分户余额数据约 2 亿条，会计分录数据约 1 亿条，交易明细数据约 700 万条，总分核对结果数据约 60 万条。

数据类型：纯文本格式的结构化数据。

时间区间：2025 年 6 月 1 号-2025 年 6 月 10 号。

赛题 10：基于大模型的数据标准自动化对标

赛题背景：赛题来源于苏州银行。银行内部各系统（如核心、财务、信贷、CRM 等）的数据字典与企业级数据标准的对标工作，仍依赖于项目经理的人工操作。确保各系统中的数据字段与企业级数据标准中的定义，如字段含义与数据标准的定义、口径、业务规则等一致，需要建立映射关系（这个过程称之为对标)。本赛题旨在解决企业级数据标准统一的现实问题，提高对标效率。

赛题要求：

总体目标：构建自动化对标工具，快速完成数据字典与企业级数据标准的对标任务，减少人工操作时间、提升对标准确率，为数据资产管理和数据分析提供更高质量的基础数据。

技术任务：

① 构建一套智能化的数据标准自动对标工具，使模型具备持续学习和适应变化的能力。将数据标准中的规则（通常是文档形式）转化为程序化的规则，并处理规则中的模糊性或主观性（如“近似匹配”或“业务理解偏差”)。建立数据标准变更的版本控制机制，跟踪标准定义的演变历史，实现增量学习能力，使模型在不遗忘旧知识的前提下学习新标准。

② 优化系统架构实现高效集成与性能平衡。设计微服务架构，将语义理解、规则应用、术语检索等功能解耦。实现模型服务化，通过容器化部署和自动扩缩容满足性能需求。

③ 进行领域特定的模型训练与优化，实现商业银行数据字典字段与数据标准的自动化比对、偏差识别，最终实现自动建立映射关系。采用领域自适应技术(Domain Adaptation)，在通用语言模型基础上进行增量预训练。设计多任务学习框架，同时学习术语对齐、数据类型识别、业务规则提取等关联任务。

配套数据情况：

数据标准样本量：684 个，均为真实脱敏业务数据。

已对标数据字典样本量：训练集 95387 条，测试集 23608 条。

时间区间：2020 年 12 月-2024 年 12 月（无实时更新)。

赛题 11：基于大模型保险文档精准召回系统——面向寿险产品知识库的语义搜索能力检验

赛题背景：赛题来源于利安人寿保险股份有限公司。企业内部知识库通常包含数千份产品文档，并以 PDF 存储，包含图片、表格、文本结构，但业务人员需实时查询产品信息时存在召回失效、文档解析失败、信息割裂、时效滞后等问题。本赛题旨在解决保险业务人员需实时查询内部知识库，但根据语义内容进行的查询结果不精准、有遗漏的现实问题。

赛题要求：

总体目标：结合大语言模型开发能够支持精准查询企业知识库内容的工作流和智能体。

技术任务：

① 开发 AI 智能体，提高语料召回率，确保在用户提出问题时，能够尽可能多地找到相关的语料。构建查询重构方案（例："快返型产品" → "生存金 5 年内返还的寿险条款"）；通过测试集的相关内容，提取保险术语本体（如"养老社区保险=CCRC 保险"）。

② 提高智能体理解用户口语表达的意图的能力，提高问答的准确性和针对性。构建查询意图自助优化系统，解决业务人员口语查询（如"能住养老院的保险"）与文档规范表述（如"CCRC 资格权益"）之间的语义鸿沟。

③ 使智能体能够综合多个文档信息，提供全面准确且连贯的回答。消除单产品信息分散于条款、精算报告等多文档导致的碎片化召回，实现跨文档语义整合。

④ 针对产品库高频更新导致的过期文档误召回问题（如停售产品仍被检索），建立时效性精准打分系统。根据语料的时效性对其进行合理排序，使智能体的回答优先参考最新语料。

配套数据情况：

数据规模：知识库数据 PDF 文件 1000 份（包含养老、年金、健康险等主力产品线，含迭代版本），时间范围为：2022-2024 年。

监管依据文件：制度文本若干

赛题 12：基于领域小模型与通用推理大模型的非结构化数据智能分类模型构建与应用

赛题背景：赛题来源于江南农村商业银行。银行业务场景中，每天都要处理海量非结构化数据，但是目前主要依靠人工分类或基础的文本分析工具处理这些数据。人工分类不仅耗时耗力，还容易因主观判断偏差导致分类错误；基础文本分析工具则多基于关键词匹配，面对合同中复杂的法律条款表述、客户咨询时口语化的语义表达，无法准确识别文本内涵。本赛题旨在解决银行非结构化数据智能分类问题，缩短数据处理周期，降低人工审核成本，提升风险识别的及时性和准确性。

赛题要求：

总体目标：构建基于领域小模型与通用推理大模型的非结构化数据智能分类模型，提取银行各类凭证材料中的非结构化信息，并实现高效分类。

技术任务：

① 构建领域小模型+OCR+布局分析模块，解析合同、票据的文本与版面结构，自动提取关键特征。突破人工编写正则表达式或关键词规则的局限，解决复杂语义及开放版式集合分类难题。

② 在少量标注数据情况下，通过领域专用小模型对样本特征的精细提取，结合通用推理大模型的泛化能力，实现零样本泛化。

③ 最终构建能够支持新增类别以及对未知文档分类的非结构化数据分类模型。弥补大语言模型无法直接处理非结构化文档图像、复杂表格版式的缺陷，提升多模态大模型细粒度特征感知能力，降低大模型高算力带来的使用成本问题。

配套数据情况：

样本量：已脱敏的凭证数据集，包含 500 条样本，训练集 400 条，测试集 100 条，无合成数据。

时间区间：2020 年 1 月 2024 年 12 月（覆盖近 5 年历史数据，无实时更新）。

数据类型：文本类：贷款合同、财报文本等；图像类：票据扫描件、证件图像等；混合类：图文混合报表、合同 PDF 等。

三、图像识别：

赛题 13：基于计算机视觉的授信审批印章真伪鉴别模型构建与运用

赛题背景：赛题来源于江苏银行。商业银行授信审批（特别是网贷业务）的关键环节中，申请人提交的证明材料（如银行流水单、合同文本等）常包含需核验的印章信息。商业银行一般通过验印系统进行印章验证，但是，当扫描件因亮度不均、角度倾斜或清晰度下降时，系统极易将真实印章误判为“不一致”，并且系统对基于图像处理的伪造手段（如局部像素篡改、对抗样本攻击、图层叠加）一般无识别能力。本赛题旨在解决银行验印系统的关键技术难题。

赛题要求：

总体目标：构建智能化、自动化的印章真伪鉴别解决方案，实现精准定位图片中的印章区域，智能判别其真伪，输出量化的真伪概率，与预留印鉴进行对比并输出量化的一致性概率。

技术任务：

① 基于深度学习的目标检测算法，训练模型精准识别流水图像中的印章位置，标注出印章区域的坐标信息。在分辨率不一、版式多样的各类（如银行流水等）扫描件中，定位印章位置并裁剪。基于深度学习的目标检测算法（如 YOLO 系列、Faster R-CNN），训练模型精准识别流水图像中的印章位置，标注出印章区域的坐标信息。

② 设计多尺度图像预处理模块（自适应对比度增强+去模糊算法），提升低分辨率/模糊图像中的印章特征。针对伪造手段的强对抗性，突破特征提取瓶颈，输出可量化的印章真伪概率值。

③ 构建几何校正模块，消除扫描畸变对匹配的影响，判定印章与另一印章是否为同一印章。设计抗干扰相似度算法，在环境光亮度/扫描拍摄角度场景保持鲁棒性，能够在环境、光照、色彩不同的情况下，准确判定印章与另一印章是否为同一印章。针对给定的扫描印章，判定其与给定的印章是否一致，输出可量化的印章一致性概率值。

配套数据情况：

样本量：印章页扫描图片共 1 万组，包含造假印章样例 500 组，无印章样例 3000 组，含真实印章样例 6500 组。每组包含扫描图片和预留印章。

赛题 14：银行业务影像质量缺陷智能检测算法探索

赛题背景：赛题来源于南京银行。银行业务的数字化和无纸化转型过程中，大量业务办理依赖于客户提交的影像资料。但是，由于客户拍摄环境复杂、设备差异、操作不当或影像本身问题，导致大量提交的影像存在模糊、缺角、反光、信息不完整、畸变等质量缺陷。本赛题旨在解决当前人工或半自动化的方式审核影像的效率不足、易出错问题。

赛题要求：

总体目标：构建基于深度学习的银行业务影像质量缺陷智能检测算法，实现对影像质量的实时、自动、高精度检测，并在发现缺陷时能立即提醒，从而显著提升影像审核效率，降低运营成本，并优化客户的业务办理体验。

技术任务：

① 开发具有智能影像预处理与特征增强模块的深度学习算法，并且能够自动校正并提取业务影像中的关键信息。实现自适应光照补偿与阴影去除，对过曝、欠曝、存在阴影的影像进行亮度、对比度与色彩的自适应调整，提升影像各区域细节可见性。构建去模糊与去噪网络，有效去除因手抖、对焦不准、低光照等导致的运动模糊、高斯噪声等，同时保留影像关键细节信息。针对拍摄角度不正、透视变形等问题，开发自动图像矫正算法，将影像校正到标准视角，并对由于镜头或拍摄姿态引起的几何畸变进行有效矫正。

② 使开发的深度学习算法能够对无法提取关键信息的影响进行质量标注分类，并贴上对应的质量问题标签。同时实现对多种细粒度影像缺陷（如模糊、反光、缺角、信息缺失、遮挡、污渍、畸变、分辨率不足等）的分类识别与精确像素级定位。结合银行业务影像的特性，引入注意力机制，使模型能够重点关注业务关键信息区域，并对缺陷类型进行更深层次的语义理解，区分“信息不清晰”与“信息不存在”。利用少量标注样本快速泛化到新缺陷类型，提升模型的适应性。

配套数据情况：

影像样本库：已初步收集约 3000 张经过脱敏处理的影像，部分样本包含常见缺陷类型（如模糊、反光、缺角）。

赛题 15：基于图像识别与深度学习的银行信贷业务图片鉴伪质检系统研究

赛题背景：赛题来源于江南农村商业银行。银行信贷业务中，客户提交的身份证、收入证明、房产证等材料的真实性直接关系到信贷风险的控制。然而，当前材料审核主要依赖人工目测，存在造假识别难度大、审核效率低下、误判率高等现实问题。本赛题旨在解决信贷业务图片的真伪识别效率问题。

赛题要求：

总体目标：开发智能化图片鉴伪质检系统，能够自动识别材料中的篡改痕迹，提升审核效率和准确性，降低信贷风险。。

技术任务：

① 开发多类型造假特征检测算法，支持常见造假类型的识别。研究 PS 修图、复印件篡改等造假手段的特征规律，构建翻拍假证识别模型（通过摩尔纹、畸变分析区分原件与翻拍）。设计对抗生成网络（GAN）合成图像鉴别器，支持 10 种以上常见造假类型的识别。

② 融合图像视觉特征（如印章边缘）、文字语义（如合同金额）、格式规则（如发票代码逻辑）实现交叉验证。建立规则库：如“身份证号码与出生日期逻辑校验”，构建跨模态校验引擎。开发图文一致性模型：检测房产证图片与登记文字的区域匹配度。融合注意力机制识别关键字段篡改（如金额、日期）。

③ 优化模型结构，使模型轻量化且支持实时检测配套数据情况。实现单张图片检测时间≤1 秒，支持高并发处理（≥100 张/分钟）。

配套数据情况：

样本量：1000 份图片（含 200 份造假样本）。

数据类型：银行流水（10%）、企业执照（30%）、收入证明（5%）、购销合同（5%）、其他影像（50%）。

造假类型：文字 PS（40%）、公章伪造（20%）、图片 PS（10%）、全图合成（30%）。