海南农商银行智能OCR平台人工智能算法

金融应用信息披露报告

一、重要提示

本报告是依据《人工智能算法金融应用信息披露指南》对披露人工智能算法应用情况的要求，就海南农商银行基于人工智能算法提供的金融产品和服务情况进行说明，并通过我行官网进行披露。

二、基本情况

（一）机构简介

海南农村商业银行是在原海南省农村信用社联合社和19家市县法人农信社（农商银行）基础上，采取新设合并方式组建而成，是全国首家按照全省统一法人模式改革成立的地方法人银行，其前身可溯至1951年琼山区美兰椰子头信用社，拥有74年扎根守土历史。新组建的海南农村商业银行，注册资本金220亿元，总行内设19个部门，下辖总行营业部和18家一级支行，440余家营业网点，员工7000余人。

（二）人工智能算法应用情况

产品名称：智能OCR平台

服务内容：行外客户办理开户等业务需对客户证件进行OCR识别，以辅助客户进行信息的录入；行内用户办理财务发票报销、信贷客户财务报表识别等业务，辅助用户进行信息录入，以提升办理效率。

三、人工智能算法金融应用风险治理情况

（一）组织保障

海南农商银行按照《海南农村商业银行股份有限公司信息化项目建设实施管理办法》要求，成立智能 OCR 平台建设项目组，项目组由项目经理和项目成员组成，其中项目成员包括科技人员和业务人员共 5 人。项目组按照科学规划、严谨设计、周密部署、精心组织的原则，扎实有效推进项目建设。项目负责落实人工智能算法相关数据治理与隐私保护、模型风险和法规遵从与行业标准等方面的组织保障措施，研究、解决项目实施遇到的问题和困难，稳步推进项目进程。

（二）实施情况

智能 OCR 平台建设项目组组织项目组科技人员和业务人员对智能 OCR 平台人工智能技术应用风险进行专项治理，根据法规遵从与行业标准要求，对人工智能技术在金融产品智能服务过程中，在技术应用方面、数据安全方面、系统风险方面以及科技伦理等方面进行风险控制设计与风险排查，确保系统人工智能技术应用安全可控。

四、人工智能算法技术应用信息披露

（一）算法组合信息披露情况

算法组合信息披露是对金融产品和服务所集成的全部人工智能模型算法组合使用情况的整体说明，算法组合信息披露情况见下表 1。

表 1 算法组合信息披露情况

披露分类	披露项	信息披露内容
算法组合信息披露	算法组合单	算法组合清单包含 3 种人工智能算法:卷积神经网络算法、OCR 文字检测算法、OCR 文字识别算法。
	算法组合使用 的开发框架	算法组合使用 PyTorch、TensorFlow 开源框架, Python3、C++开发语言。
	算法组合使用 的开源软件	算法组合使用 OpenSSL 、 OpenCV 、Fastapi、Boost、 Protobuf、Crypto++、Rapidjson 等开源软件库。
	算法组合调度 机制	OCR 识别算法将传入的图片进行一系列预处理操作,通过平滑滤波器、中值滤波器等方法去除图像中的噪声并校正图像中的倾斜角度,然后通过 OCR 文字检测算法检测出图像中的所有的文本行,再将各个文本行通过 OCR 文字识别算法识别出文字内容,最后将文字和坐标整合 起来反馈给图片传入端。
	算法组合触发条件	传入图片文件触发算法组合。
注: 1.PyTorch 是一个由 Facebook 发布的用于机器学习和深度学习的开源深度学习框架。 2.TensorFlow 是一个由 Google 的 Google Brain 团队开发的开源机		

器学习框架。

3.Python3 为解释型的编程语言。C++为编译型的编程语言。

4.OpenSSL 是一个开源的加密库,提供了广泛的安全功能,主要用于数据传输中的加密和解密,包括 SSL 和 TLS 协议。

5.OpenCV 是一个开源的计算机视觉和机器学习软件库。

6.FastAPI 是一个现代、快速(高性能)的 Web 框架。

7.Boost 是一个用于 C++的开源库集合,提供了很多对 C++标准库的扩展。

8.Protobuf 是由 Google 开发的一种轻巧的数据交换格式,用于序列化结构化数据。

9.Crypto++是一个开源的 C++加密库,提供了多种加密算法,包括对称加密、非对称加密、散列函数和消息认证码等。

10.Rapidjson 是一个快速的小型 JSON 解析器和生成器,用 C++编写。

（二）算法逻辑信息披露情况

算法逻辑信息披露是对组合中的每个算法对象逐一说明算法机理，算法逻辑信息披露情况见表 2。

表 2 算法逻辑信息披露情况

披露分类	披露项	披露项
算法逻辑信息披露	算法功能说明	a）卷积神经网络算法：由多组多维矩阵组成的多层神经网络，通过多个可训练的滤波器和偏置项进行矩阵计算操作得到特征图，并通过一系列的加权、加偏置等算法处理可以使其对图像的特征提取能力较于传统的图像特征提取算

		法有显著提升。 b）OCR 文字检测算法：对于普通场景（如行驶证），我们将文字检测转换为对关键字目标或关键条目的检测；对于复杂场景，由于证件目标在图像中所占比例过小，直接提取微小候选目标会导致一定的定位精度损失。为了保证高召回和高定位精度，我们采用由粗到精的策略进行检测。 c）OCR 文字识别算法：引入背景（Blank）类别以吸收相邻字符的混淆性。整体网络结构分为三层：卷积层、递归层和翻译层；其中卷积层提取特征；递归层既学习特征序列中字符特征的先后关系，又学习字符的先后关系；翻译层实现对时间序列分类结果的解码。
	算法推理过程说明	a）卷积神经网络算法推理过程：输入图像后，经多组多维矩阵构成的多层神经网络，通过可训练的滤波器和偏置项进行矩阵计算，生成特征图，再经加权、加偏置等算法处理，完成图像特征的提取。 b）OCR 文字检测算法推理过程：采用由粗到精的策略，先进行大致区域的检测，再对候选区域进行精细定位，以保证高召回和高定位精度。 OCR 文字识别算法推理过程：输入经检测得到的文字区域图像，卷积层提取字符特征；递归层学习特征序列中字符特征的先后关系以及字

		符本身的先后关系；翻译层对递归层输出的时间序列分类结果进行解码，最终得到识别结果。
	算法推理结果说明	a）卷积神经网络算法推理结果：输出图像的特征图，其特征提取能力较传统图像特征提取算法有显著提升，为后续的图像检测等任务提供了高质量的特征表示。 b）OCR 文字检测算法推理结果：通过由粗到精的策略，实现了对目标的高召回和高定位精度的检测，为文字识别提供了准确的区域定位。 OCR 文字识别算法推理结果：输出准确的文字识别结果，能够有效区分相邻易混淆字符，实现对输入文字区域的精准识别，将图像中的文字内容转化为可编辑、可理解的文本信息。
	算法技术路线选择说明	基于目前先进的卷积神经网络（CNN），循环神经网络及其变体（RNN，LSTM，GRU）结合条件随机场（CRF）技术和词嵌入（Word Embedding）技术构建基础的算法库。为了保证高召回和高定位精度，采用由粗到精的策略。
	算法技术成熟度说明	卷积神经网络（Convolutional Neural Network，CNN）算法在 OCR（Optical Character Recognition，光学字符识别）领域能够自动从图像中提取有效的特征，避免了传统 OCR 技术中需要手动设计特征的繁琐过程。这种自动学习的能力大大提高了 OCR 的准确性和泛化能力。基于 CNN 的 OCR 技术能够处理不同字体、大小、

		旋转角度、光照条件等复杂情况下的文字，并且在不同应用场景中展现出良好的泛化性。
	算法重构条件说明	为了进一步提高复杂场景下的识别能力和处理效率，对卷积神经网络算法进行重构，引入自适应预处理模块和注意力增强机制。在识别过程中，系统能自动调整预处理参数以适应不同背景和字体的图像，同时通过注意力机制关注图像中的关键文字区域，实现更精准的文字提取和识别。
	算法假设条件说明	OCR 识别是保证银行业务在需要对图片信息进行提取的时候，给出图片信息的计算机所能接收的反馈信息给到前端业务进行反馈。
	算法使用限制说明	目前 OCR 识别的服务对象为图片提取文字业务类型，支持中英文文字提取。
	算法参数及超参数说明	算法输入参数主要包括图片和类型。 算法输出参数主要包括图片的 kv 健值等。

（三）算法应用信息披露情况

算法应用信息披露是与人工智能算法金融应用场景相关信息的说明，避免因对算法应用的错误理解而误导客户，信息披露情况见表 3。

表 3 算法应用信息披露情况

披露分类	披露项	披露项

算法应用信息披露	算法应用场景	OCR 主要应用于我行客户开户、信息审核等业务场景。
	算法应用目的	业务凭证及单据识别，以进行辅助录入审核。
	算法应用服务范围	向我行二代零售信贷系统、支付网关、手机银行、微信银行、海南农商微服务等业务系统提供 OCR 识别的接口调用服务。
	算法应用服务前提	业务系统要传入文件到 OCR 识别算法模型。
	算法应用获得渠道	OCR 识别的算法模型应用从供应商下载私有化部署包。
	算法应用潜在风险和防护措施	OCR 算法的识别精度会被图像的质量影响，图像质量差的识别率会相对较低，通过在手机银行等前端应用不断尝试图片压缩率，以及对性能不佳模型进行及时的迭代训练，并在重要交互环节通过屏幕文字确认、界面提示等方式降低客户使用风险。
	算法应用必要性	利用 OCR 识别算法实现文件识别和辅助录入，可以显著提升业务效率。
	算法应用预期效果	利用 OCR 识别算法实现文件识别和辅助录入，将审核类业务的效率提升 30%以上。

（四）算法数据信息披露情况

算法数据信息披露是对算法使用的数据来源、数据采集、数据质量控制以及数据与场景的关联性进行充分说明，

信息披露情况见表 4。

表 4 算法数据信息披露情况

披露分类	披露项	披露项
算法数据信息披露	算法数据与金融应用相关性说明	OCR 识别主要是基于传入的文件进行识别，辅助提升录入效率，算法训练数据也是基于相应的卡证票据。
	算法数据来源说明	算法调优、测试数据集主要来源于我行历史数据集。
	算法数据采集说明	算法模型涉及的调优数据、测试数据、推理数据均已经数据所有者（控制者）授权，采用去标识化进行脱敏处理，脱敏后无法直接关联真实客户身份。
	算法数据质控说明	具备健全严格的算法数据质量管理制度，覆盖数据采集、数据清洗、数据整理、数据标注、数据集构建各个环节，保障算法数据的完整性、一致性，数据分布的合理性、无偏性，数据样本的充足性，数据操作的规范性、合规性。
	算法组合使用的第三方软件产品	不涉及

（五）算法主体信息披露情况

算法主体信息披露是对人工智能算法金融应用服务提供者建立的算法管理相关机制（安全保障、风险防范、伦理治 理等机制）的说明，信息披露情况见表 5。

表 5 算法主体信息披露情况

披露分类	披露项	信息披露内容
算法主体信息披露	算法主体建立安全保障机制说明	我行建立健全数据使用管理办法、敏感信息内控管理制度，明确各环节责任分工，为智能 OCR 平台提供算法安全保障和数据安全保障。智能 OCR 场景模型部署环境属生产环境，且有操作日志保存，保障推理过程安全。生产环境模型有应急处理机制，遇到任何上线失败情况可立刻回退。
	算法主体建立风险防范机制说明	我行建立产品创新管理办法、数据使用管理办法，形成 “事前审核准入、事中监控预警、事后处置问责” 的风险防范及补偿机制，作为智能客服的责任主体确保客户权益得到有效保障。

	算法主体建立伦理治理机制说明	我行成立产品创新委员会和产品创新办公室，审议本行涉及业务创新、技术创新的管理制度，审议数字化转型过程中的模式创新、机制创新等方案和制度。
	算法主体建立信息披露组织实施保障措施说明	建立健全严格的算法数据质量管理制度，覆盖数据采集、数据清洗、数据整理、数据标注、数据集构建各个环节，保障算法数据的完整性、一致性，数据分布的合理性、无偏性，数据样本的充足性，数据操作的规范性、合规性。同时建立完善的信息披露内部管理制度和流程，明确信息披露的责任部门和工作要求，确保信息披露及时、准确、完整。

（六）算法变更信息披露情况

人工智能算法金融应用采用在线实时服务，由于人工干预程度越少，服务及其策略的调整可能更加频繁快速，可能引发风险性活动，应运用与之匹配的实时监测技术并及时披露相关调整以及变更信息，信息披露情况见表6。

表6算法变更信息披露情况

披露分类	披露项	披露项
算法变更信息披露	算法变更版本号	a）卷积神经网络算法，版本 2.1.5，更新时间 2023-8-8。 b）OCR 文字检测算法，版本 1.0，更新时间 2023-8-8。 c）OCR 文字识别算法，版本 10.0，更新时间 2023-8-8。
	算法变更原因说明	
	算法变更影响说明	
	算法变更生效时间	
	算法变更实施说明	按照我行相关实施管理办法及变更交付流程等要求，开展与算法更新、软件更新的类型、内容和程度相适宜的验证与确认活动，将风险管理、可追溯分析贯穿于更新全程，形成涉及人工智能算法的产品用户接受测试报告、上线变更评审、应急处置及回退方案和上线验证报告等记录以供核查。
	算法变更保障措施	从技术架构实现、功能完备性、数据一致性、性能稳定性等方面，对算法变更前后进行差距分析并予以记录，并针对变更前后的性能差距采取人工优化、人工填写数据等方式进行必要补救。

（七）算法审计信息披露情况

人工智能算法金融应用涉及数据、算力、场景等多种要素的深度融合与交互，一定程度上增加了算法金融应用的风险。为保证金融产品和服务的业务连续性，降低技术风险、操作风险，有必要对人工智能算法金融应用进行合规性审计，并及时将算法审计活动所依据的政策、法规、标准等信息以及审计结果进行说明,信息披露情况见表 7。

表 7 算法审计信息披露情况

披露分类	披露项	信息披露内容
算法审计信息披露	算法审计依据	依据《人工智能算法金融应用信息披露指南》、《海南农村商业银行股份有限公司开源技术应用管理实施细则》、《海南农村商业银行股份有限公司信息化项目建设实施管理办法》等政策法规和内部制度开展算法审计工作。
	算法审计信息	1.上线前，利用算法对抗鲁棒性测试工具进行自动化算法审计，审计活动覆盖算法安全漏洞扫描、算法逻辑漏洞扫描等方面；所有严重和一般问题已完成整改。 2.上线前，依据相关管理制度进行算
	审计内容说明	
	算法审计结论	
	算法审计频率	

		法审计，审计活动覆盖算法公平性、算法可解释性、算法鲁棒性等方面。 3.上线后，依据相关内部制度开展算法跟踪审计，审计活动覆盖算法运行情况、算法安全风险事件、算法合规情况等方面。 通过算法审计活动，确认OCR识别算法组合满足算法上线条件，在运行过程中符合国家及行业相关规范和我行内部管理制度要求，未发现重大违法违规问题和安全风险隐患。

五、重大事项

报告期内，海南农商银行无人工智能算法金融应用业务相关的重大诉讼、仲裁事项。

报告期内，海南农商银行未发生人工智能算法金融应用业务相关的重大案件、重大差错等情况。