



# 中华人民共和国金融行业标准

JR/T 0287—2023

## 人工智能算法金融应用信息披露指南

Guidance on information disclosure for financial applications based on  
artificial intelligence algorithms

2023 - 11 - 08 发布

2023 - 11 - 08 实施

中国人民银行 发布



## 目 次

前言 .....	II
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 信息披露原则 .....	2
5 信息披露形式 .....	2
6 信息披露内容 .....	3
附录（资料性） 某银行人工智能算法金融应用信息披露报告示例 .....	6

## 前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国人民银行科技司提出。

本文件由全国金融标准化技术委员会（SAC/TC 180）归口。

# 人工智能算法金融应用信息披露指南

## 1 范围

本文件提供了人工智能算法在金融领域应用过程中的信息披露原则、信息披露形式和信息披露内容等要素。

本文件适用于金融机构使用人工智能算法提供金融产品和服务时规范地开展算法信息披露活动，增强人工智能算法的可解释性和透明度，维护金融消费者合法权益。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

JR/T 0221—2021 人工智能算法金融应用评价规范

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**信息披露** information disclosure

金融机构使用人工智能算法开展金融产品和服务时，以独立信息披露报告、产品服务说明书等灵活形式披露本机构人工智能算法金融应用的相关信息，从而助力提升人工智能算法金融应用透明度和安全性的行为。

### 3.2

**金融产品** financing product

金融机构为满足客户某种需求并使客户获取确定或可能收益，向市场提供的金融服务。

[来源：GB/T 32319—2015，3.2，有修改]

### 3.3

**第三方软件产品** third parties' software products

由金融机构以外的其他公司提供的软件产品。

### 3.4

**开源软件** open source software

一种可以获取源代码的软件。

注：1. 开源软件的著作权持有人通过开源许可证将软件的复制、修改、再发布和商业应用等权利向公众开放。

2. 开源软件通常为免费获取使用，但存在商业版本、服务定制收费等形式，不等同于免费软件。

### 3.5

#### 超参数 hyperparameter

在模型训练过程之前设置的一种可调节参数。

注：超参数不属于训练数据，相对于在模型训练过程中学习得到的模型参数，超参数一般基于先验知识，由人工控制调节。

### 3.6

#### 算法推理 algorithm inference

将训练得到的模型应用到新的未标记实例来进行预测的过程。

### 3.7

#### 算法重构 algorithm refactoring

在不改变算法原有功能的基础上，通过调整代码优化算法非功能特性的过程。

## 4 信息披露原则

### 4.1 依法合规

信息披露内容合法、合规，信息发布渠道必须符合国家及行业相关管理要求和标准规则等。

### 4.2 真实准确

披露的信息宜以客观事实或者具有事实基础的判断和意见为依据，如实反映客观情况，避免存在虚假记载或不实陈述，引用的数据和资料注明来源。披露的信息宜使用准确和贴切的语言，避免含有任何失实、夸大、误导性的陈述。

### 4.3 及时一致

当使用人工智能算法的金融应用产品和服务发生发布、更新、停止、关闭（退出）或造成重大影响风险事件时，及时发布信息披露报告。信息披露保持披露口径的一致性。信息披露内容同时采用中文文本与外文文本时，保持文本内容一致，理解发生歧义时，以中文文本为准。

### 4.4 完整可读

信息披露内容尽可能完整，在必须满足国家及行业相关管理要求前提下，以足够简明、清晰和完整的方式披露信息，保障人工智能金融应用信息披露内容的可读性和完整性。

## 5 信息披露形式

### 5.1 信息披露条件

当使用人工智能算法的金融产品和服务首次发布、发生重大风险事件（因算法滥用、算法歧视造成大范围争议或负面影响事件等）、重大变更（建模算法变更、重要特征发生变化以及可能对相关方权益造成影响的其他情况等）、服务关闭（退出）时，发布人工智能算法金融应用相关信息，确保信息披露的及时性和一致性。

### 5.2 信息披露方式

金融机构信息披露可采取企业自我声明、信息披露报告、产品服务说明书等方式，通过官方网站等线上渠道披露人工智能算法金融应用相关信息。

### 5.3 信息披露维度

金融机构信息披露以提供的金融产品和服务为维度，对其所使用或集成的人工智能算法均可开展算法信息披露工作。

## 6 信息披露内容

### 6.1 概述

人工智能算法金融应用信息披露内容分为算法组合信息披露、算法逻辑信息披露、算法应用信息披露、算法数据信息披露、算法主体信息披露、算法变更信息披露、算法审计信息披露7个方面。信息披露内容可参考附录中提供的人工智能算法金融应用信息披露报告示例模板。

**注：**本文件作为人工智能算法金融应用信息披露工作的指南，以指导金融机构有序开展信息披露工作为目的。金融机构在依法合规、保障安全的前提下，可根据披露信息的上下文、应用场景的敏感性以及内部管理需求，平衡好信息披露和隐私保护的关系，自主开展信息披露活动，在取得客户和社会公众信任的同时切实保障用户个人信息和数据安全。国家及行业主管部门有相关管理要求的，应遵照其要求开展有关工作。

### 6.2 算法组合信息披露

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，包括但不限于以下内容。

- a) 对金融应用中人工智能算法组合所使用的开发框架进行说明，例如开发框架的名称、版本、主流编程语言。
- b) 对金融应用中人工智能算法组合所使用的开源软件进行说明，例如开源软件的名称、用途、版本、开源许可证、来源等。
- c) 列出金融应用中人工智能算法组合信息，例如语音识别算法、意图识别算法、对话决策算法、语音合成算法等。
- d) 对金融应用中人工智能算法组合的调度机制进行说明，例如调度顺序、调度算法及其实现功能等。
- e) 对金融应用中人工智能算法组合的触发条件进行说明，例如触发金融应用的场景、动作、参与方信息等。

### 6.3 算法逻辑信息披露

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，包括但不限于以下内容。

- a) 对金融应用中人工智能算法功能进行说明，例如语音识别、图像识别功能及功能描述等。
- b) 对金融应用中人工智能算法推理进行说明，例如描述推理（或推理执行）的过程、推理（或推理执行）的结果等。
- c) 对金融应用中人工智能算法技术路线进行说明，例如技术路线成熟度、技术路线的选择考虑等。
- d) 对金融应用中人工智能算法重构条件进行说明，例如算法重构的触发条件等。
- e) 对金融应用中人工智能算法假设条件进行说明，例如算法进行开发运行的前提条件、算法进行开发运行的假设等。

- f) 对金融应用中人工智能算法参数及超参数进行说明，例如算法主要的输入参数、算法主要的输出参数等。

#### 6.4 算法应用信息披露

算法应用信息披露是对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，包括但不限于以下内容。

- a) 对金融应用中人工智能算法应用场景进行说明，例如算法应用金融业务领域（营销、运营、客服、交易、风控等）说明、算法应用渠道（移动客户端软件、网页、自助设备等）说明、算法应用在金融服务流程中的作用等。
- b) 对金融应用中人工智能算法应用目的进行说明，例如算法应用带来的业务增益（降本增效、提升客户体验等），应用该算法的使用依据、前期调研结果和其他理由等。
- c) 对金融应用中人工智能算法应用风险进行说明，例如算法应用上线前的风险评估结果、算法应用风险防范措施等。
- d) 对金融应用中人工智能算法应用使用限制进行说明，例如算法的适用对象、适用场景等使用限制条件等。
- e) 对金融应用中人工智能算法应用使用效果进行说明，适度披露算法应用评估指标，例如准确率、召回率、误差等。

#### 6.5 算法数据信息披露

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金融应用场景的关联性进行充分说明，包括但不限于以下内容。

- a) 对金融应用中人工智能算法数据与金融应用相关性进行说明，例如数据与金融应用关联关系、关联程度等。
- b) 对金融应用中人工智能算法使用数据来源进行说明，例如产生数据的主体、数据产生方式等。
- c) 对金融应用中人工智能算法使用数据的采集过程进行说明，例如获取数据采集授权的方式、实施采集活动的主体、采集数据的方法等。
- d) 对金融应用中人工智能算法使用数据的质量控制进行说明，例如负责质量控制的主体、质量控制的方法流程、质量控制的评估结果等。
- e) 对金融应用中人工智能算法使用数据全生命周期的安全与隐私保护机制进行说明。
- f) 对金融应用中人工智能算法所接入的具备收集和使用个人信息功能的第三方软件进行说明，例如第三方软件产品的名称、开发商、用途、功能、收集和使用的个人信息类型等。

#### 6.6 算法主体信息披露

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，包括但不限于以下内容。

- a) 对金融应用中算法主体建立的安全保障机制进行说明，例如针对算法依赖库安全、算法模型安全、算法运行环境安全、算法数据安全、算法攻击防范等建立的有关机制和采取的措施。
- b) 对金融应用中算法主体建立的风险防范机制进行说明，例如算法上线机制、算法退出机制、算法使用风险说明、应急处理机制、风险赔偿机制等。
- c) 对金融应用中算法主体建立的伦理机制进行说明，例如公平与包容机制、透明和可解释性机制、隐私与数据保护、伦理问题的组织机制和采取的主要措施等。
- d) 对金融应用中算法主体建立的信息披露组织实施保障措施进行说明，例如信息披露目标、信息披露组织架构、信息披露内部管理制度、信息披露流程等。



## 6.7 算法变更信息披露

人工智能算法金融应用以自动化服务为主，由于人工干预少，服务及其策略的调整更加频繁快速，可能引发风险性活动，有必要运用与之匹配的实时监测技术，及时披露相关调整变更信息，包括但不限于以下内容。

- a) 对金融应用中人工智能算法变更所使用的开发框架进行说明，例如开发框架的名称、版本、主流编程语言等。
- b) 对金融应用中人工智能算法变更的情况进行披露，说明算法变更具体原因，涉及的变更内容及其与上一版本的差异，例如版本号、变更原因、变更内容等。
- c) 对金融应用中人工智能算法变更的生效时间及详细的变更影响进行披露，例如算法变更对本身应用系统的影响、对上下游关联系统的影响、对客户的影响等。
- d) 对金融应用中人工智能算法变更的实施说明进行披露，例如算法变更具体的实施时间、实施地点、实施方式、实施步骤、实施时长、实施结束后的验证方式等。
- e) 对金融应用中人工智能算法变更的保障措施进行披露，例如算法变更实施中的异常情况、变更实施后验证的异常情况、其他不可控因素等各类异常情况下的保障措施。

## 6.8 算法审计信息披露

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，人工智能算法在金融场景中应用可能存在一定的潜在风险。为保持金融产品和服务的业务连续性，降低技术风险、操作风险，依法依规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，包括但不限于以下内容。

- a) 对金融应用中算法审计类型进行披露，算法审计类型分为第三方外部审计和内部审计。其中，针对第三方外部审计说明人工智能算法应用合规性评估服务提供方名称及其资质背景、评估时间、评估结论及评估报告有效期等。
- b) 对金融应用中算法审计依据进行披露，例如人工智能算法金融应用所遵循的国家标准、行业标准或国家及行业管理要求等。
- c) 依据JR/T 0221—2021的相关要求对金融应用中算法审计基本信息进行披露，例如人工智能算法在安全性、可解释性、精准性、性能等测评指标的测试结果。
- d) 对金融应用中算法审计内容进行披露，例如人工智能算法在数据安全、个人信息保护等方面的评价内容、重要评价结果和结论等。
- e) 对金融应用中算法审计结论进行披露，例如人工智能算法应用过程中的违法违规行、国家及行业管理部门公开发布的惩处决定情况和整改情况。
- f) 对金融应用中算法审计频率进行披露，例如人工智能算法的内部安全审计流程、周期以及相关审计人员等信息。

附 录  
(资料性)

某银行人工智能算法金融应用信息披露报告示例

本附录以示例的形式提供人工智能算法金融应用信息披露报告内容。

示例：

1 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，就某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。

本报告正文及摘要已经公司内部审议通过，并通过企业网站、行业组织的信息发布平台等渠道进行披露。

2 基本情况

2.1 机构简介

某银行是国家金融管理部门批准成立的商业银行之一。某银行核心战略是通过科技驱动，坚持普惠原则，为金融服务实体经济贡献力量。

2.2 人工智能算法应用情况

产品名称：金融产品智能销售服务

服务内容：在金融产品营销场景中，挖掘潜在客户并通过智能语音机器人自动模拟真人客服拨打电话向客户推荐金融产品，实现多轮人机互动。

3 人工智能算法金融应用风险治理情况

3.1 组织保障

某银行成立金融科技管理委员会，负责研究金融科技管理架构，分析人工智能算法金融应用业务发展和风险管理状况，制定一定时期的人工智能算法技术应用的发展政策、管理制度和相关技术性要求，审议人工智能技术应用相关业务。

3.2 实施情况

某银行信息金融科技管理委员会工作情况：共召开 3 次委员会会议，对金融产品智能销售服务等人工智能技术应用项目进行审议。

某银行信息金融科技管理委员会组织的人工智能技术应用风险专项治理工作中，对金融产品智能销售服务在技术应用方面、数据安全方面、系统风险方面以及伦理失范方面进行全面排查，相关问题均已通报给归口部门进行整改，风险管理部门跟踪并核实整改情况。目前已对所有风险问题完成整改。

4 人工智能算法技术应用信息披露—金融产品智能销售服务

4.1 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	算法组合清单	算法组合清单包含 5 种人工智能算法：语音识别算法、多轮对话意图识别算法、意图识别算法、对话决策算法、语音合成算法。
	算法组合使用的开发框架	算法组合使用 Tensorflow2.0 的开发框架、Python3 的开发语言。
	算法组合使用的开源软件	算法组合使用 opencv、tensorRT、numpy、matplotlib 的开源软件库。
	算法组合调度机制	语音识别算法将客户语音转成文本送入对话决策算法，对话决策算法依次调用多轮对话意图识别算法和意图识别算法进行识别并选择应答话术，通过语音合成算法转成语音进行播报。
	算法组合触发条件	选定营销目标客户，主动拨打电话触达客户，客户接通后触发算法组合。

注：1. Tensorflow2.0 为开源的人工智能开发框架。

2. Python3 为解释型的编程语言。

3. opencv 为跨平台计算机视觉和机器学习软件库。

4. tensorRT 为高性能深度学习推理工具。

5. numpy 为开源的数值计算扩展。

6. matplotlib 为绘图工具。

#### 4.2 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	信息披露内容
算法逻辑信息披露	算法功能说明	<p>a) 语音识别算法：将自然语音转换为文字序列。</p> <p>b) 多轮对话意图识别算法：完成文字序列的词法分析、句法分析、语义分析等。</p> <p>c) 意图识别算法：理解客户话语以及意图。</p> <p>d) 对话决策算法：基于策略决策对话输出。</p> <p>e) 语音合成算法：将文本序列转化为模拟语音。</p>

表 2 算法逻辑信息披露情况（续）

披露分类	披露项	信息披露内容
算法逻辑信息披露	算法推理过程说明	多轮对话意图识别算法和意图识别算法均为意图识别模块，对客户表达的自然语言进行语义理解，多轮对话意图识别算法模块主要是基于客户对话文本进行文本分类，得到概率最高的意图类别，意图识别算法主要是基于客户对话文本进行文本相似度计算，与已有意图识别算法库中的文本进行语义匹配，得到最相似的标准语义。  对话决策算法模块将多轮对话意图识别算法和意图识别算法 2 个模块返回的结果进行比较判别，得到得分最高的意图类别或者标准语义，从而得到对应的机器人话术。
	算法推理结果说明	
	算法技术路线选择说明	多轮对话意图识别算法模块采用基于预训练语言模型的文本分类模型。 意图识别算法采用的是粗召回和精排的算法路线。
	算法技术成熟度说明	文本分类是自然语言处理领域的常用成熟技术，在开源模型中广为应用。 粗召回采用的是文本向量召回，通过无监督向量抽取模型进行向量表征提升，从而提升粗召回的精准度；精排采用文本分类模型，通过粗召回缩小分类模型输入类别数，通过文本分类进一步筛选意图类别。粗召回和精排是意图识别算法中较为常用的成熟技术。
	算法重构条件说明	为了进一步提升机器人对话的自然度和拟人化，对原有算法链路进行重构，增加语义打断模块，在播报过程中客户插话时进行实时判断客户是在表达何种意图，实时停止播报并根据客户意图选择新的话术进行播报。
	算法假设条件说明	智能语音机器人保证输出的内容符合某银行贷款或理财等产品的属性，在客户对产品的任何信息有疑问时，需给出相应解答，因此假设意图识别算法意图库覆盖各业务产品的所有产品信息。
	算法使用限制说明	目前，智能语音机器人服务对象为语音听力无障碍人士，语言支持普通话。
	算法参数及超参数说明	算法输入参数主要包括客户询问、上一轮对话文本、业务场景号。算法输出参数主要包括意图身份、意图对应话术身份、话术语音身份等。

#### 4.3 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	信息披露内容
算法应用信息披露	算法应用场景	智能语音机器人主要应用于某银行个人贷款、理财产品的主动营销活动中。
	算法应用目的	利用人工智能算法实现个性化产品推荐、自动人机对话等功能。
	算法应用服务范围	智能语音机器人服务于某银行移动手机银行所有注册客户，提供基于手机银行的个人贷款、理财产品的查询、咨询和促销服务。该服务适用于采用中国标准普通话的客户，并不支持粤语、客家话等特殊情况。
	算法应用服务前提	智能语音机器人是通过某银行移动手机银行移动端应用在客户手机端安装使用，需获取客户手机端应用程序安装权限，以及通信录、摄像头及麦克风等访问权限。其应用服务的前提是客户手机银行可正常注册使用。
	算法应用获得渠道	智能语音机器人应用可通过移动设备供应商提供的应用中心下载安装。
	算法应用潜在风险和防护措施	智能语音机器人采用人工智能技术，其服务质量受客户口音、外部环境噪声等影响，可能影响语音识别率。某银行通过丰富场景化的训练语料，持续提升算法模型的适应性提高特殊情况的识别率，并对重要交互环节通过屏幕文字确认、语音提示等方式降低客户使用风险。
	算法应用必要性	利用人工智能算法实现以客户为中心的个性化、精准化产品营销，提供基于数据的决策服务，降低营销成本。
	算法应用预期效果	预期客户名单命中率可在原有基础上提升 60%、客户服务获得率可在原有基础上提升 90%、营销精准度可在原有基础上提升 80%。

## 4.4 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	信息披露内容
算法数据信息披露	算法数据与金融应用相关性说明	智能语音机器人主要采集客户语音数据，用于营销问答。
	算法数据来源说明	算法训练、调优、测试数据集主要来源于某银行历史电话客服语音数据集，为增加方言兼容性，部分调优数据集采购于第三方数据服务机构提供的定制化数据集以及来自开源社区的数据库。

表4 算法数据信息披露情况（续）

披露分类	披露项	信息披露内容
算法数据信息披露	算法数据采集说明	智能电销场景所有算法模型涉及的训练数据、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	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。
	算法组合使用的第三方软件产品	算法组合使用智能数据分析第三方软件产品，主要用于数据统计分析。

#### 4.5 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	某银行建立健全人工智能技术应用安全管理制度、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，为智能电销提供算法安全保障和数据安全保障。智能电销场景所有算法模型涉及的训练数据均进行脱敏处理，不存在传输、存储环节的篡改和泄漏，模型训练和部署环境均属于安全开发和生产环境，且有操作日志保存，保障训练和推理过程安全。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滚。
	算法主体建立风险防范机制说明	某银行建立健全人工智能技术应用安全管理制度、数据使用管理办法、敏感信息内控管理制度，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，作为智能电销的责任主体确保客户权益得到有效保障。

表5 算法主体信息披露情况（续）

披露分类	披露项	信息披露内容
算法主体信息披露	算法主体建立伦理治理机制说明	某银行成立人工智能技术应用伦理治理工作组，负责建立伦理治理规章制度并开展伦理治理审查工作，伦理治理的内容包括道德层面、技术层面、法制层面，对于不同国家、不同民族、不同文化、不同性别等都有明确的规范。伦理治理工作组由评审领导小组、风险管理领导小组、金融科技项目管理领导小组和金融科技管理委员会办公室共同构成。
	算法主体建立信息披露组织实施保障措施说明	某银行印发《某银行金融科技管理委员会工作规定》，规范人工智能伦理治理相关工作开展和实施的程序、角色分工以及议事规则。

## 4.6 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	信息披露内容
算法变更信息披露	算法变更版本号	a) 语音识别算法，版本 1.0，更新时间 2023-7-15。
	算法变更原因说明	b) 语音合成算法，版本 1.0，更新时间 2023-7-15。
	算法变更影响说明	c) 多轮对话算法，版本 10.0，更新时间 2023-7-18。
	算法变更生效时间	d) 意图识别算法，版本 10.0，更新时间 2023-8-05。
		e) 对话决策算法，版本 10.0，更新时间 2023-8-05。
算法变更实施说明	f) 更新内容为融合多轮对话算法和意图识别算法，提升整体意图识别准确率，将原来对话决策算法的串行调度合为一次调用。	
	算法变更实施说明	按照《某银行人工智能技术应用安全管理办法》《某银行系统变更管理办法》《某银行系统验收测试管理办法》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，形成算法产品客户接受测试报告、算法变更评审、公司审批、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、算法回退旁路等方式进行必要补救。变更后，算法依旧限制在智能客户场景的应用范围内使用，并提供自动语音播报警示提示信息。

#### 4.7 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	某银行对智能电销开展内部算法审计和外部算法审计。 a) 内部算法审计： ——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。 ——上线前，依据《某银行人工智能技术应用安全管理办法》进行算法审计，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；所有严重和一般问题已完成整改。 ——通过算法审计活动，确认智能电销算法组合满足算法上线条件。 ——上线后，依据《某银行人工智能技术应用安全管理办法》《某银行人工智能技术应用信息披露管理办法》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；所有严重和一般问题已完成整改。 b) 外部算法审计：依据 JR/T 0221—2021《人工智能算法金融应用评价规范》开展外部算法审计，准确率达到 85%，精准率达到 90%，单条响应时间 100 毫秒。
	算法审计信息	
	审计内容说明	
	算法审计结论	
	算法审计频率	

#### 5 重大事项

报告期内，某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，某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。



## 参 考 文 献

- [1] GB/T 32319 银行业产品说明书描述规范
  - [2] JR/T 0227 金融机构环境信息披露指南
  - [3] JR/T 0258 金融领域科技伦理指引
  - [4] 《关于规范金融机构资产管理业务的指导意见》（银发〔2018〕106号文印发）.2018-04-27
  - [5] 《上市公司信息披露管理办法》（中国证券监督管理委员会令第182号发布）.2021-03-18
  - [6] 《关于加强互联网信息服务算法综合治理的指导意见》（国信办发文〔2021〕7号文印发）.2021-09-17
  - [7] 《金融科技发展规划（2022—2025年）》（银发〔2021〕335号文印发）.2021-12-31
  - [8] 《互联网信息服务算法推荐管理规定》（国家互联网信息办公室 中华人民共和国工业和信息化部 中华人民共和国公安部 国家市场监督管理总局令第9号发布）.2022-03-01
  - [9] 《生成式人工智能服务管理暂行办法》（国家互联网信息办公室 中华人民共和国国家发展和改革委员会 中华人民共和国教育部 中华人民共和国科学技术部 中华人民共和国工业和信息化部 中华人民共和国公安部 国家广播电视总局令第15号发布）.2023-05-23
  - [10] 《科技伦理审查办法（试行）》（国科发监〔2023〕167号文印发）.2023-09-07
-