L70

Q

上海九方云智能科技有限公司

企业标准

Q/ JFY 087-2023

金融智能合规算法模型技术标准

Financial Intelligence Compliance Algorithmic Modelling Technical Standards

2023-04-7 发布　　2023-04-08 实施

上海九方云智能科技有限公司 发布

目　　次

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。

本文件由上海九方云智能科技有限公司提出。

本文件由上海九方云智能科技有限公司归口。

本文件起草单位：上海九方云智能科技有限公司。

本文件主要起草人：李畅、於伟、王晓霞、杨志笑。

引　言

2017年7月，国务院印发的《新一代人工智能发展规划》明确指出要大力发展“智能金融”，创新智能金融产品和服务，发展金融新业态，鼓励金融行业应用智能客服、智能投研、智能风控等先进技术的发展。这一政策的出台代表着人工智能技术在金融场景中的持续深入与赋能。

金融合规工作是金融机构、企业长期稳定发展的内在需求和重要保证，同时也是风险预知和控制的一个关键要素。《中央企业合规管理办法》等法律法规强调加强合规智能化建设，要求通过信息化手段优化管理流程，记录和保存相关信息。随着我国金融管理向更高水平迈进，合规智能化建设对金融合规管理的支撑作用将更加突出。

当前，我国金融机构、企业合规管理建设还处于初级阶段，无法完全满足金融行业合规风险控制需要，相关违法违规事件仍时有发生。同时金融监管趋严、金融机构与企业合规成本大幅提升。

在此现状下，必须将金融合规智能化建设作为金融机构、企业合规管理活动的重要一环，金融合规需要与人工智能有机融合，帮助金融机构、企业实现质检审核、内部协同数字化，管理决策可视化，风控智能化，降低金融机构、企业的合规成本，持续提高合规管理有效性和准确性，全面赋能企业安全经营，维护金融市场安全，防范系统风险。

金融智能合规算法模型技术标准

1 范围

本文件规定了面向金融行业的智能合规算法模型技术所需满足的总体原则、主要功能、数据处理、模型训练、构建集成、模型服务、运营监控、模型迭代、模型管理等技术标准。

本文件适用于开展金融业务的组织机构，以及提供智能合规风控算法模型服务的机构建设、运行和优化智能合规风控模型。

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。

GB/T 27910—2011 金融服务 信息安全指南

JR/T 0200—2020 金融科技创新风险监控规范

JR/T 0263—2022 机器学习金融应用技术指南

JR/T 0258—2022 金融领域科技伦理指引

JR/T 0201—2020 金融科技发展指标

JR/T 0200—2020 金融科技创新风险监控规范

JR/T 0199—2020 金融科技创新安全通用规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能 Artificial Intelligence

利用计算机和机器模仿人类思维的问题解决和决策制定能力。

3.2

深度学习 Deep Learning

通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据分布式特征表示的一种学习方法。

[来源：JR/T 0221—2021，术语和定义3.1]

3.3

智能合规 Intelligent Compliance

利用人工智能和数据分析等技术来辅助和改进组织或个人遵守法规、政策、法律、规章和行业标准的过程。

4 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

5 总体原则

金融智能合规算法模型技术的总体原则一般包括以下内容。

a)依法合规性。严格遵守与人工智能金融应用相关的法律规范及标准，坚守智能金融的合规底线。

b)安全可控性。保持智能合规算法模型服务的安全性、可靠性、可控性，具备可追溯、可信赖、可审计的能力。

c)隐私保护性。须确保尊重用户隐私和数据保护。

d)公平透明性。确保智能合规算法模型决策的透明性，确保算法设定公平、合理、无歧视。

e)可追责性。须建立责任机制，确保智能合规算法模型研发及应用的可追责性。

6 主要功能与技术流程

6.1 主要功能

金融智能合规算法模型技术核心功能如下：

通过对全部自有内容生产流程增加AI预审步骤，识别潜在违规点、违规类型并进行主动拦截，大幅提升人员审核效率，保障内容合规。主要包括以下类型：

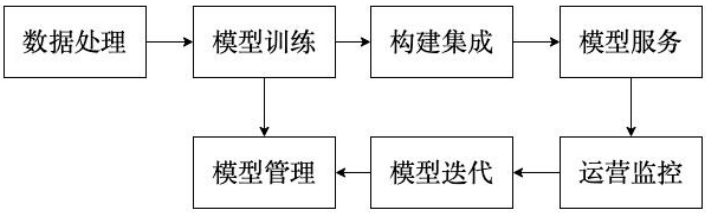
a)图文素材内容审核

b)视频素材内容审核

c)音频素材内容审核

6.2 技术流程

金融智能合规算法模型技术一般性流程见下图。



7 数据处理

7.1 概述

数据处理是将原始数据进行加工、转换和标注，以生成适用于模型训练的数据集，从而为模型训练过程提供高质量数据。数据处理作为实现智能合规算法模型的初期环节，构成了模型训练的基础，高质量的数据预处理有助于产生更加精确的模型。

7.2 数据处理

数据处理应该包含以下过程：

1）对原始数据进行数据清理、数据转换和数据增强等操作，旨在解决数据异常、数据缺失以及数据冗余等问题，以提升数据的质量。

原始数据通常可分为结构化数据和非结构化数据（如文本、图像、音频等）。对结构化数据的处理涵盖了去重复、处理无效值和填充缺失值等步骤。对非结构化数据的处理包含以下内容：

a)对文本数据的处理包括降低词频、添加生僻词等；

b)图像数据的处理包括旋转、翻转、裁剪等；

c)视频数据的处理则包括帧的抽取等；

d)音频数据的处理则包括去噪等。

这些步骤有助于优化原始数据，使其适合用于模型训练和分析，从而提高了数据的质量。

2）需要对数据进行合规业务标注，作为模型训练样本数据。

3）需要支持大批量数据的接入和处理，具备一定的自动化数据处理能力。

8 模型训练

8.1 概述

模型训练是自动构建、优化和调整计算模型的过程，旨在使其能够准确执行任务、提取数据信息和实现复杂推理。它包括选择优化算法、参数调整、迭代训练、验证和部署。在金融合规场景中，模型训练效率和质量至关重要。

模型开发过程应包括以下内容：

a) 优化算法选择及参数调整。人工智能的优化算法是一类用于改进机器学习和深度学习模型的算法，以便它们能够更好地执行任务、提高性能或适应不断变化的数据。这些算法的主要目标是在模型的参数空间中搜索并找到最佳参数组合，从而最小化或最大化一个特定的性能指标或目标函数。

b) 迭代训练。在确定优化算法以及其参数后，模型将进行长时间、高资源占用的训练过程。应以合理的训练策略对模型进行训练。

c) 验证和部署。在训练完成后，需要对模型进行部署以便于用户使用。

8.2 优化算法

对于优化算法，应该考虑：

a) 优化算法的选择。常见的优化算法包括：梯度下降算法、牛顿法、遗传算法、模拟退火算法、蚁群优化算法等。

b) 对于深度学习，可选择下列算法之一：随机梯度下降算法、动量法与Nesterov加速、AdaGrad、RMSprop、AdaDelta、Adam。

c) 优化算法的参数调整。常见的参数选择方法包括手动搜索、网格搜索和随机搜索等。

8.3 迭代训练

迭代训练非常耗时，对于迭代训练，应该按照流程：

a) 确保模型的可行性，包括但不限于模型结构是否正确，梯度传播是否正常等。

b) 确保计算资源与模型训练相匹配。

c) 同时对不同参数的模型进行训练。

d) 监控训练情况，保证梯度正常，损失下降正常。

8.4 模型验证

模型验证需要在给定的指标下进行，并对训练得到的模型进行评估。模型验证应包括

a) 评估指标

b) 模型比较规则

c)合规业务评测

9 构建集成

9.1 概述

构建集成是指对训练后的模型，包括模型结构，模型权重与模型配置等进行必要的打包、封装和集成，在经过测试后，形成简单可用的交付物的过程。通常情况下，交付物的使用者（如其他技术人员）往往并不具备人工智能算法的相关专业知识，因此，构建集成对金融智能合规算法模型技术的落地、应用与分发非常重要。

构建集成流程应包括以下内容：

a) 工程设计。在开始设计集成前，应对该过程进行详细的调研、规划与设计。

b) 流程自动化。对流程实现自动化，以提高开发、测试、发布的效率。

c) 反馈机制。在交付物交付后，应与用户保持联系，给予用户反馈问题的通道。

9.2 工程设计

工程设计需要考虑以下方面：

a) 需要考虑的内容为交付物类型。

b) 交付物使用方式。

c) 项目周期等。

9.3 流程自动化

自动化流程应包括以下方面：

a) 对模型权重、配置等的管理

b) 新代码的发布

c) 对新代码的测试

d) 迭代后的交付物进行快速打包、封装和分发。

9.4 反馈机制

反馈机制应能实现以下方面：

a) 基于反馈快速进行调整。

b) 保证模型的高可用性和时效性

10　模型服务

10.1　概述

模型服务是指将构建集成后的模型进一步部署至业务环境，使其能够更为直接的方式为用户所使用。此外，模型服务也包括对服务进行的相关管理与维护，使其具有高效性，鲁棒性，时效性。在金融领域中，人工智能服务的形态需要高度易用并健壮。

模型服务应该考虑：

a) 服务管理。模型服务应该便于管理。

b) 服务形态。金融人工智能模型的模型服务需要与用户群体相适应。

10.2　服务管理

服务管理包括：

a) 对服务进行维护与更新。

b) 服务管理的维护需要考虑服务的资源需求、负载均衡、运行情况监控等

c) 服务管理的更新包括选择服务的更新策略，对不同应用场景、不同形态、不同受众的服务使用相适应的更新策略。

10.3　服务形态

服务形态应考虑：

a) 对于技术人员，可使用较为复杂的形态，如API、提供框架支持（PyTorch、Tensorflow）等，并给予充实、详细的文档。

b) 对于一般用户，宜提供较为友好的形态，如图形界面，语音交互等方式。

11　运营监控

11.1　概述

运营监控是指持续地跟踪和评估模型在生产环境中的表现，从而及时发现和解决问题，确保模型的效果稳定和可靠，为业务决策提供重要支持。

11.2　运营监控的核心能力

运营监控的核心能力通常包括：

a) 自动化与实时能力：运营监控需要具备自动化和实时监控的能力，能够持续监控模型的运行状态和性能，及时发现问题并采取相应措施。

b) 可追溯能力：记录历史监控数据和事件，确保监控过程的可追溯性，以便后续分析和比对，也为模型改进提供依据。

c) 预警分析处置能力：建立健全的预警机制，能够及时发现异常情况并触发预警，配备分析和处置能力，能够自动或部分自动地分析和处理异常情况，减少人工干预的成本和时间。

d) 连续迭代支持能力：运营监控需要与模型迭代的训练工作高效对接，为模型的持续训练和在线更新提供支持，确保模型在实时性要求较高的场景中持续优化和改进。

11.3 运营监控的主要维度

运营监控的主要维度通常包括：

a) 模型预测性能监控：模型上线后应该建立一个完善的指标体系，包括模型准确性、召回率、精确率等。定期的对这些指标进行监测和分析，从而保障模型性能的可信度。

b) 模型业务效果监控：通过定义关键指标、数据采集与记录、对比分析等手段，监控业务目标直接相关的指标，及时发现模型效果的问题，并采取相应措施保障业务目标的达成。

c) 数据流入和流出监控：持续地监测持续流入和流出的数据流及其特征，确保数据的及时性、准确性和完整性以符合模型输入要求，确保模型输出的可靠性。

d) 基础设施和环境监控：持续地监测模型运行和生产环境的变化，包括软件版本、基础设施、网络通信、负载均衡等，并及时对其进行处理和调整。

e) 服务异常监控与处理：持续地监测和分析异常情况，例如数据缺失、模型不稳定、系统故障等，并且有相应的处理机制和流程，以确保问题能够得到及时解决。

f) 建立监控报警机制：建立即时的应急报警机制，采用自动化报警方式及时告知相关人员，例如邮件、短信、微信等。

12 模型迭代

12.1 概述

模型迭代是在市场条件不断变化的动态环境中，通过模型再训练的动态迭代，可以实现模型的不断优化和更新，以适应新的数据和环境变化，从而维持模型的准确性和鲁棒性。模型迭代可以降低金融应用中业务目标变化和数据分布变化带来的风险，提高模型性能和新鲜度。

12.2 模型迭代的关键问题

12.2.1 重新训练模型的时机

重新训练模型的时机要考虑的主要因素包括：

a) 监控指标：设定监控模型性能的指标和当指标低于或超过时重新训练模型的预设阈值，可能包括模型的准确率、回报率、风险指标等。

b) 环境变化：输入数据或环境由于市场条件变化、新产品推出或监管政策调整等因素引起发生了导致模型的预测性能下降的根本性的变化。

12.2.2 重新训练模型的数据集

重新训练模型的数据集和主要特点如下：

a) 历史数据具有代表性，以便捕捉过去的市场行为和趋势,应覆盖各种市场情况和时间段。

b) 实时数据随金融市场情况随时发生变化，可以让模型更好地适应当前的市场环境和动态。

12.2.3 重新训练模型的方式

重新训练模型的方式和主要特点如下：

a) 利用历史数据在离线环境中重新训练模型，适用于数据量较大、训练过程较繁琐的情况。

b) 离线重新训练可以在计算资源充足的环境下进行，并且不会中断实时应用程序的运行。

c) 利用实时数据对现有模型进行增量训练，快速响应市场变化，连续更新模型提高性能。

d) 在线重新训练通常需要具备对模型的实时监控和评估的能力。

12.3 模型迭代的关键环节

人工智能模型迭代的关键环节通常包括：

a) 监控指标和触发条件设置：明确监控模型性能的指标，如模型的准确率、回报率、风险指标等，并设定触发再训练任务执行的条件。

b) 循环训练管道配置：建立良好配置的模型训练流水线，及时获取模型最新训练配置和流程。

c) 模型输出持续监控机制：通过持续监控模型的输出，并与实际情况进行比较，了解模型在新数据和新场景下的表现。

d) 模型再训练方式选择：根据应用场景的安全性和业务需求的实时性，选择离线重新训练或在线增量学习。

e) 模型更新管道配置：考虑配备模型部署和模型服务发布更新的管道，快速响应市场变化并进行模型迭代。

13 模型管理

13.1 概述

模型管理是从模型的整个生命周期出发，对大量模型资产进行全盘统一管理，全方位了解模型运行状态和使用情况，便于模型的风险管理、合规审计和共享使用，确保智能合规算法模型在金融应用中的可靠性、安全性和有效性。

13.2 模型注册管理

模型注册管理主要包括：

a) 在模型管理平台上进行模型注册，包括模型的基本信息、描述、作者等。确保模型信息的清晰、完整记录。

b) 实施模型权限管理，确保只有授权人员才能进行模型的添加、删除、修改等操作，以保护模型的安全和合规性。

c) 模型管理平台支持导入第三方模型，并对自训练和第三方模型进行统一集中管理。包括对模型文件和镜像的导入、存储和版本管理。

13.3 模型资产管理

模型资产管理主要包括：

a) 应当建立模型档案，包括模型的版本、训练数据集、训练参数、评估指标等信息，方便模型的溯源和管理。

b) 确保模型在开发阶段经过充分的验证和测试，包括数据质量评估、模型性能评估、风险评估等。记录模型在不同数据集上的性能指标、运行效果等，为模型改进和优化提供依据。

c) 编写模型使用文档，详细说明模型的输入输出、使用方法、限制条件等，方便其他人员了解和使用模型。

d) 提供统一管理服务，包括对模型服务的部署、监控和更新等，确保服务的稳定性和高效性。

13.4 模型版本管理

模型版本管理主要包括：

a) 应当提供模型版本记录，包括基本信息、运行环境、输入输出等维度的变更记录。每个变更都应有明确的目的和说明。

b) 支持更新视图，对比不同版本之间的差异并查看详细的变更记录，方便开发人员和审核人员进行审查和决策。

c) 提供版本提交和回滚功能，确保模型的发布过程可控和可追溯，同时能够快速回滚到上一个稳定版本。

13.5 模型安全和合规管理

模型安全和合规管理主要包括：

a) 应当对模型进行安全加固，包括模型文件的加密存储和传输、访问权限的管理、数据隐私保护等，防止模型信息泄露和滥用。

b) 应当遵循金融领域的合规要求，如个人信息保护、反洗钱、反欺诈等，确保模型的应用符合相关法规和政策。

13.6 模型评估和加速管理

模型评估和加速管理主要包括：

a) 应当支持对模型在云端和边缘环境下的效果和性能进行综合评估，以确定模型适用场景。

b) 提供降低模型推理精度和压缩模型复杂度的模型性能提升工具，在相同计算资源下提高预测推理服务的吞吐量。

13.7 模型部署和共享管理

模型部署和共享管理主要包括：

a) 支持模型的云部署和边缘部署，提供面向不同环境的模型镜像或SDK文件，以满足不同环境下的部署需求。

b) 应当具备将模型共享到当前项目组织或全平台的能力，促进团队合作和知识共享。