

Q/ZXZQAI-001-2023

中信证券股份有限公司企业标准

Q/ZXZQAI-001-2023

中信证券人工智能模型开发标准

Development Standard for Artificial Intelligence Model in CITICS

2023-08-25 发布

2023-08-30 实施

中信证券股份有限公司 发布

目 次

目 次.....	I
前 言	III
引 言	IV
中信证券人工智能模型开发标准.....	1
1 范围	1
2 规范性引用文件.....	1
3 术语和定义	1
3.1 训练集	1
3.2 验证集	1
3.3 测试集	2
3.4 模型评估标准	2
3.5 超参数	2
3.6 特征工程	2
3.7 模型部署	2
3.8 模型上线	2
3.9 灰度上线	2
3.10 影子上线	3
3.11 批处理	3
4 需求管理	3
4.1 需求对接	3
4.2 需求评估	3
4.3 需求设计	4
4.4 需求维护	4
5 数据工程	5
5.1 数据采集	5
5.2 数据验证	5
5.3 数据加工	6
5.4 数据建模	7
6. 模型开发	7
6.1 模型构建	7
6.2 模型训练	9
6.3 模型验证	9
6.4 模型调优	10
7. 模型仓库	10
7.1 构建模型组	10
7.2 入库模型	11
7.3 变量管理	11
7.4 发布模型	12
7.5 评估模型	12
7.6 预测模型	13
8. 模型服务	13
8.1 审核模型与服务	13
8.2 部署服务模型	13

8.3 上线服务模型	14
8.4 管理在线服务	14
参考文献	15

前 言

本标准依据 GB/T 1.1-2020《标准化工作导则 第一部分：标准化文件的结构和起草规则》给出的规则起草。

本标准由中信证券股份有限公司提出。

本标准由中信证券股份有限公司归口。

本标准起草部门：中信证券股份有限公司信息技术中心。

本标准主要起草人：方兴，岳丰，刘殿兴，徐峻峰，钱岳，陈成。

引 言

人工智能（**Artificial Intelligence**，简称 **AI**）作为一项前沿技术，正在深刻地改变着人们的生活和工作方式。人工智能模型的开发是实现人工智能应用的关键环节之一，对于提高企业的竞争力和创新能力具有重要意义。

为了规范和统一中信证券内部人工智能模型的开发流程，确保模型的质量和可靠性，提高开发效率，降低开发成本，特制定本《中信证券人工智能模型开发标准》。本标准旨在为公司内部人工智能模型开发人员提供一套统一的开发标准和最佳实践，以确保模型的可维护性、可扩展性和可复用性。

本标准适用于公司内部所有人工智能模型的开发工作，包括数据准备、模型训练、模型评估和模型部署等环节。通过遵循本标准，能够更加高效地开发出高质量的人工智能模型，为企业的业务发展提供有力支持。

中信证券人工智能模型开发标准

1 范围

本标准规定了中信证券股份有限公司人工智能模型开发的需求管理、数据工程和模型开发、模型评估和模型仓库、模型服务的规范。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是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

GB/T 41867-2022 信息技术人工智能术语

GB/T 41479-2022 信息安全技术 网络数据处理安全要求

GB/T 35274-2023 信息安全技术 大数据服务安全能力要求

GB/T 42755-2023 人工智能 面向机器学习的数据标注规程

AIPIA/P 0008—2022 人工智能研发运营一体化(Model/MLOps)能力成熟度模型 第一部分：开发管理

AIPIA/P 0009—2022 人工智能研发运营一体化(Model/MLOps)能力成熟度模型 第二部分：模型交付

3 术语和定义

3.1 训练集

指用于训练人工智能模型的数据集，包含大量的样本数据和对应的标签，用于模型学习和参数调整。

3.2 验证集

指用于验证人工智能模型性能的数据集，用于调整模型的超参数，选择最佳的模型配置。

3.3 测试集

指用于评估人工智能模型性能的数据集，用于模型的最终评估，验证模型在未见过的数据上的泛化能力。

3.4 模型评估标准

指用于评估人工智能模型性能的指标或标准，包括准确率、精确率、召回率、F1、LIFT值等，用于衡量模型的预测能力和效果。

3.5 超参数

指在训练人工智能模型时需要手动设置的参数，如学习率、批量大小、隐藏层节点数等，用于调整模型的性能和泛化能力。

3.6 特征工程

指在人工智能模型开发中，对原始数据进行处理和转换，以提取出对模型训练和预测有意义的特征，提高模型性能和准确性的一系列规范和方法。

3.7 模型部署

确保人工智能模型在生产环境中高效、稳定地部署和运行的规范和流程。

3.8 模型上线

确保人工智能模型在开发、测试和部署过程中符合公司内部的技术要求和规范。

3.9 灰度上线

部署模型作为服务后端并对部分请求流量进行处理，其余流量仍通过正式上线的模型进行处理。

3.10 影子上线

在不影响正式模型的前提下，在模型提供在线服务的过程中复制正式上线模型的全部流量。

3.11 批处理

按照一定的顺序和规则将一组计算机指令集合起来，然后一次性提交给计算机系统执行，而无需用户进行干预。通过自动化和简化重复性的任务，提高计算机系统的效率和可靠性。

4 需求管理

4.1 需求对接

4.1.1 沟通需求

沟通需求是指对业务需求进行深入理解、分析，并明确其范围。这一过程要考虑数据和系统支持等因素，以评估需求的可行性，并基于评估结果设计场景方案。

4.1.2 需求确认

在需求确认阶段，对业务需求进行全面理解，包括从业务角度充分理解需求的目标、要求以及相关背景信息。在这个过程中，要考虑需求所解决的问题、问题的背景、需求的目标和适用场景等因素，最终要明确业务成功的标准，以便在后续的建模工作中能够有效展开，并在应用和效果评估阶段进行评估。

通过需求对接，即与业务方沟通需求并进行需求确认，确保双方对需求的理解一致，减少后续的误解和问题。

4.2 需求评估

需求可行性评估是至关重要的环节，其目标在于全面了解需求的业务价值以及确定其可行性。在这个过程中，需要详细考虑多个方面，包括：

- **业务价值分析：**对需求的业务价值进行深入分析，理解其对业务目标的贡献和影响。
- **资源要求评估：**确定实现需求所需的资源，包括数据的可获取性、适用的算法和所需的算力等。

- **约束条件考虑：**考虑时间、质量、成本预算等方面的约束条件，确保项目能够在限定的条件下实现。
- **风险评估：**评估项目实施过程中可能出现的风险，并制定应对策略。
- **安全性和易用性：**考虑模型应用过程中的安全性需求以及用户的易用性体验。
- **其他因素：**根据具体情况，还可能需要考虑其他因素，如法规合规性等。

通过充分的可行性评估，确保项目在开始阶段就有清晰的认知和计划，更好地实现业务目标并降低项目风险。

4.3 需求设计

4.3.1 详细方案设计

方案设计旨在将业务需求转化为具体的技术方案，构建出能够解决业务问题的模型策略和解决方案，包含方案设计编写和方案设计管理两个部分。

4.3.2 方案设计编写

方案设计编写是基于模型场景设计方案，进行详细的方案构建工作。在这个阶段，根据需求对接和可行性评估的结果，将业务需求转化为技术术语，并定义问题的输入、输出要求，以及成功的衡量标准。通过清晰而精确的方案设计编写，确保在后续的模式开发中有明确的指导方向。

4.3.3 方案设计管理

方案设计管理涉及对已编写的方案进行生命周期的管理和跟踪。这包括了方案的发布者、内容、状态、优先级、参考数据等相关配置项的管理。通过对方案设计的有效管理，确保设计方案的一致性、完整性和可追踪性，保证项目团队在开发和交付过程中能够始终遵循规划和方向。

4.4 需求维护

4.4.1 持续反馈

持续反馈是指在整个项目周期中与业务方保持紧密沟通，及时收集业务变化和新的需求，确保项目始终与实际需求保持一致。

4.4.2 需求迭代

需求迭代是指根据业务方的反馈和变化，对需求进行周期性的调整和更新，包括对模型算法、数据集和参数的调整和优化，确保项目逐步适应业务变化，提高交付的价值和质量。

4.4.3 需求版本管理

对于需求的变更和迭代，需要进行版本管理。通过版本管理、记录和追踪需求的变更历史，方便查找和回溯，确保在需求迭代过程中的代码和文档的一致性和可追溯性。

5 数据工程

5.1 数据采集

数据采集是指从各种数据源中获取数据的过程。

5.1.1 数据接入

数据接入是指通过各种方式（如 API、爬虫等）将数据从外部数据源导入到系统中。

5.1.2 数据存储

数据存储是指将采集到的数据存储合适的数据存储介质中，如数据库、数据仓库等。

5.2 数据验证

数据验证是指对采集到的数据进行验证和纠错的过程。

5.2.1 数据纠错

数据纠错是指对数据中的错误进行修正，如修复缺失值、处理异常值等。

5.2.2 质量控制

质量控制是指对数据的质量进行评估和控制，确保数据的准确性、完整性和一致性。

5.3 数据加工

数据加工是指对采集到的数据进行加工处理的过程。

5.3.1 数据标注

数据标注是指对数据进行标记或注释，包括分类标注、实体标注、情感标注等，为模型训练提供有标签的数据。

数据标注的格式要求：

- **分类标注：**对数据进行分类标注，将每个样本标记为不同的类别。标注格式为：样本数据 - 类别标签。
- **实体标注：**对数据中的实体进行标注，标记实体的起始位置和结束位置。标注格式为：样本数据 - 实体起始位置 - 实体结束位置 - 实体类型。
- **情感标注：**对数据进行情感标注，标记每个样本的情感倾向。标注格式为：样本数据 - 情感倾向。

例如：

分类标注：句子 1 - 类别 A

实体标注：句子 2 - 5 - 8 - 实体 A

情感标注：句子 3 - 正面情感

5.3.2 数据处理

数据处理是指对数据进行清洗、转换、整合等操作，为后续模型训练和应用提供高质量的数据。

数据处理的格式要求：

- **清洗：**对数据进行去除噪声、处理缺失值、处理异常值等操作，确保数据的质量和完整性。
- **转换：**对数据进行转换，如将文本数据转换为数值型数据、进行特征工程等操作，以便模型能够处理。
- **整合：**将多个数据源的数据进行整合，如合并不同表格的数据、整合不同格式的数据等，以便模型能够综合利用。

5.4 数据建模

5.4.1 数据分析探查

在模型选择阶段，数据分析探查是非常重要的步骤。通过对数据统计分析和可视化分析，了解数据的分布相关性和特征，为后续的模型选择提供依据。

- **统计分析：**通过统计方法对数据进行分析，包括计算数据的均值、方差、中位数等统计指标，以及计算不同特征之间的相关系数，了解数据的整体情况和特征之间的关系。
- **可视化分析：**通过可视化手段将数据呈现出来，包括绘制直方图、散点图、箱线图等，直观地展示数据的分布情况和特征之间的关系，发现数据中的规律和异常。

5.4.2 特征工程相关

特征工程是指对原始数据进行特征选择、特征提取和特征构建的过程，目的是提取出对问题有用的特征，为后续的模型训练提供合适的输入。

- **特征选择：**根据数据分析的结果，选择对问题有用的特征。特征选择可以通过统计方法、机器学习方法或领域知识来进行，减少特征的维度和冗余，提高模型的效果和效率。
- **特征提取：**对原始数据进行变换和组合，提取出新的特征。特征提取可以通过数学方法、统计方法或机器学习方法来进行，提取出更具代表性和区分性的特征。
- **特征构建：**根据问题的特点和领域知识，构建新的特征。特征构建可以通过数学方法、统计方法或领域知识来进行，提取出更具代表性和区分性的特征。

6. 模型开发

6.1 模型构建

模型构建是指根据业务需求和数据特征，选择适当的算法和模型架构进行模型的设计和搭建。

6.1.1 模型选择

在模型构建阶段，根据问题的性质、数据的特征以及计算资源的可用性来确定适用的模型类型和结构，满足具体业务需求。以下是一些需要考虑的因素：

- **问题性质：**包括问题的类型（如分类、回归、聚类等）、问题的复杂度以及问题的约束条件。根据问题的性质，选择合适的模型类型，如神经网络、决策树、支持向量机等。
- **数据特征：**对数据的特征进行分析，包括数据的维度、数据的分布、数据的稀疏性等。根据数据的特征，选择适合处理这些特征的模型结构，如卷积神经网络适用于处理图像数据，循环神经网络适用于处理序列数据等。
- **计算资源：**考虑可用的计算资源，包括计算能力、存储容量以及计算时间等。不同的模型类型和结构对计算资源的需求不同，根据可用的计算资源来选择适合的模型。

6.1.2 模型搭建

在实际的模型开发过程中，使用各种工具和框架来搭建机器学习和深度学习模型，如 TensorFlow、PyTorch、Scikit-Learn 等。这些工具和框架提供了丰富的函数和接口，方便构建各种类型的模型。

在机器学习和深度学习模型的搭建过程中，需要考虑以下几个方面：

- **工具和框架：**在模型开发过程中，选择适用于特定任务的工具和框架至关重要。例如，TensorFlow 和 PyTorch 是深度学习领域常用的框架，而 Scikit-Learn 适用于传统机器学习。正确选择工具和框架大大提高模型搭建的效率和性能。
- **模型结构：**根据任务的性质和数据的特征来设计合适的模型结构。对于机器学习模型，模型的结构设计涉及选择适当的算法和模型类型，这包括线性模型、决策树、支持向量机、随机森林等模型；对于深度学习模型，模型的结构设计包括确定神经网络的层数、每层的神经元数量以及它们之间的连接方式。深度学习模型通常包括输入层、隐层 hidden layer 和输出层，其中隐层的添加提高模型的复杂性，更好地捕获数据的复杂关系。
- **激活函数：**在机器学习和深度学习模型中，选择适当的激活函数来引入非线性变换。常用的激活函数包括 ReLU、Sigmoid、Tanh 等，根据任务和模型的需求选择合适的激活函数提高模型的表达能力和学习能力。
- **损失函数：**在模型的训练过程中，定义一个损失函数来衡量模型的预测结果与真实结果之间的差异。常用的损失函数包括均方误差、交叉熵和对数损失等。选择合适的损失函数指导模型的学习过程。
- **优化算法：**在模型的训练过程中，选择一个合适的优化算法来更新模型的参数，以最小化损失函数。常用的优化算法包括梯度下降、随机梯度下降和 Adam 等。选择合适的优化算法提高模型的收敛速度和准确度。

6.2 模型训练

模型训练是指利用标注数据对模型进行参数优化，使其能够从数据中学习并提取有用的特征。

6.2.1 数据集划分

在模型训练前，将数据集划分为训练集、验证集和测试集。训练集用于模型参数的学习，验证集用于模型的调优和选择，测试集用于评估模型的性能。

6.2.2 模型训练

通过使用训练集对模型进行训练，不断调整模型参数，以最小化损失函数。训练过程中，可以采用批量梯度下降、随机梯度下降等优化算法，提高模型的收敛速度和准确度。

6.2.3 模型训练策略

在模型训练过程中，需要制定合适的训练策略，以提高模型的性能和泛化能力。以下是一些常用的训练策略：

- **数据增强：**通过对训练数据进行随机变换、旋转、缩放等操作，增加训练样本的多样性，提高模型的鲁棒性和泛化能力。
- **学习率调整：**在训练过程中，逐渐降低学习率，以使模型在接近收敛时更加稳定，避免过拟合。
- **正则化：**通过添加正则化项，限制模型的复杂度，防止过拟合。常用的正则化方法包括 L1 正则化和 L2 正则化。
- **早停法：**在验证集上监测模型的性能，当模型在验证集上的性能不再提升时，停止训练，以避免过拟合。
- **批量归一化：**在每个批次的数据上进行归一化操作，加速模型的收敛速度，提高模型的稳定性。

6.3 模型验证

模型验证是指通过验证数据集对训练好的模型进行评估，确保模型的泛化能力和性能。

6.3.1 模型检查

在模型验证阶段，对模型进行检查，验证其效果和正确性。主要从模型本身能否正常运行的角度出发，检查模型的输入输出是否符合预期，是否存在异常情况。

6.3.2 模型调整

根据模型检查的结果，对模型进行调整和优化。可以通过调整模型的参数、数据预处理、特征选择等方式，改善模型的性能和效果。

6.4 模型调优

模型调优是指根据验证结果对模型进行调整和优化，提高模型的准确性和稳定性。

6.4.1 参数调优

在模型调优阶段，需要对模型的参数进行调整，提高模型的性能和泛化能力。通过调整学习率、正则化参数等超参数，优化模型的训练过程。

6.4.2 数据调优

在模型调优过程中，还可以从训练数据的维度进行调整。通过增加样本数量、减少噪声、平衡数据分布等方式，改善模型的训练效果和鲁棒性。

7. 模型仓库

7.1 构建模型组

模型组构建是对模型进行分类，实现对项目中的模型进行集中、分组管理。

7.1.1 模型组类别划分

根据模型类别构建模型组，模型组的类别划分为机器学习模型组、深度学习模型组和预训练模型组。其中机器学习模型基于二分类、多分类和回归问题训练的模型；深度学习模型基于图像分类、

目标检测、图像分割、文本分类、文本生成等任务训练的模型；预训练模型为用户在系统之外训练的模型，可导入模型仓库实现模型管理。

7.1.2 模型组的命名

模型组的命名要求在项目内唯一，以业务场景+任务类型的模式进行标准化命名；要求在简介部分描述模型组用途等信息；并为模型组设置标签，实现模型组快速检索。

7.2 入库模型

对已完成训练的模型执行入库操作，实现模型的统一管理。

7.2.1 标准化开发模型

基于数据智能平台自动建模和工作流进行标准化开发的模型，根据模型指标进行模型筛选，可直接提交至模型仓库。

7.2.2 外部导入模型

以模型文件形式直接导入模型仓库的，机器学习模型要求为 DCPipeline 模型文件、DeepTables 模型文件、PMML 模型文件、ONNX 模型文件和基于 Sklearn 的 pkl 模型文件，深度学习模型要求为 DCPipeline 模型文件、ONNX 模型文件、基于 Keras 的 h5 模型文件、基于 Tensorflow 的 pb 或 ckpt 模型文件和基于 PyTorch 的.pth 模型文件，预训练模型要求为基于 TensorFlow2 的 ckpt、h5 或 SaveModel 模型文件和基于 PyTorch 的.pth 模型文件。

7.3 变量管理

变量用于描述模型的输入和输出信息，一个模型组中的模型必须具有相同的目标变量，新入库模型要求模型的特征变量少于或等于目标模型组的特征变量。

7.3.1 变量生成

模型组的变量在第一个模型入库后生成，用户可自定义添加变量。模型未发布时，可对变量进行修改，已发布模型变量不可修改。

7.3.2 变量添加

模型组的变量可以通过以下三种方式实现添加：

- **自定义变量：**用于创建单一特征。
- **数据集获取：**将指定的数据集的列名作为变量。
- **文件获取：**解析用户上传的 csv 文件，将列名作为变量名。

7.4 发布模型

模型仓库中的模型可将模型发布到服务。服务中模型可部署、上线，对外提供模型服务。

7.4.1 发布要求

模型不可以同时发布到多个服务，当模型被发布到某个服务后，处于已发布状态，要求先从服务中移除该模型，再进行重新发布；模型发布到已有服务时，要求模型只能发布到与其具有相同特征的服务中；在单个项目下，不同模型组只要特征相同即可发布到同一服务中；模型发布时，可以将筛选的模型发布至存量服务，也可以新增服务。

7.4.2 发布模式

模型组包含自动发布与手动发布模式。自动发布要求设置一个自动发布规则，当模型组入库时，若满足自动发布规则，则会将该模型发布至指定服务；手动发布要求每次仅将满足筛选条件的模型发布到指定服务。

7.5 评估模型

模型的评估指标是衡量模型好坏的重要标准，模型仓库中的评估功能，实现一个模型组中的不同模型可以在相同基准下进行评估与对比。模型评估根据不同任务类型，提供多维度模型评估指标，以量化模型效果。

7.5.1 标准化开发模型

基于数据智能平台自动建模和工作流进行标准化开发的模型，在入库时保留模型评估结果。

7.5.2 外部导入模型

以模型文件形式直接导入模型仓库的，要求进行重新评估。

7.6 预测模型

利用模型预测功能对数据集进行结果预测，要求数据集其包含的字段需要与模型组的变量一致。机器学习模型，支持上传文件或使用数据集进行预测；深度学习模型仅支持通过数据集进行预测。

7.6.1 机器学习模型

机器学习模型可以选择项目中的结构化数据集，或者上传结构化数据 csv 文件。选择的数据集或者上传的文件要求包含的字段与模型组变量一致。

7.6.2 深度学习模型

深度学习模型要求上传结构化数据 csv 文件进行模型预测，上传的文件要求包含的字段与模型组变量一致。

8. 模型服务

8.1 审核模型与服务

模型服务将训练任务得到的模型部署为服务，并以多种方式对外提供预测能力。

8.1.1 模型审核

在模型与服务均为审核通过状态方能对模型进行部署、上线。在模型服务列表中，要求经审核无误后，将待审核模型移动到通过，即完成模型审核。

8.2 部署服务模型

部署的模型可作为在线服务的后端提供预测能力。

8.2.1 部署要求

单项目中要求部署模型的上限为 4 个。在模型部署时，要求为待部署模型调整合适资源、容器镜像后在进行提交。

8.2.2 部署状态

部署成功后，可查看模型部署状态等信息，具体如下：

- **状态以及日志：**部署成功或失败以及具体日志信息。
- **API 调试：**对部署成功的模型，提供 API 服务，可用于内部测试，调试 API 服务状态。
- **模型对比：**对部署成功以及已上线模型提供图形化评估对比，实现模型选优、上线、迭代。

8.3 上线服务模型

模型部署成功后，可以将其上线，以服务的形式对内外部应用提供预测和调用。

8.3.1 正式上线

通过该方式部署的模型作为服务后端，接收客户端发送的请求数据并返回预测结果。单个服务中首个上线的模型为正式上线。

8.3.2 灰度上线

通过该方式部署的模型作为服务后端并对部分请求流量进行处理，其余流量仍通过正式上线的模型进行处理。新版本模型上线时，要求使用灰度上线，实现服务后端的平稳迁移。

8.3.3 影子上线

在不影响正式模型的前提下，在模型提供在线服务的过程中复制正式上线模型的全部流量。

8.4 管理在线服务

针对在线服务，支持多种调用方式，并要求进行监控、评估以及版本管理。

8.4.1 服务调用

通过在线服务对外提供算法模型的推理能力，为适应不同应用场景，支持多种服务调用方式：

- **同步调用：**通过 HTTP 方式提供模型预测服务，适合小批量、调用准实时、频次不高的场景。
- **同步调用：**基于 HTTP2.0 通信协议，具有更高传输效率与性能，适合低延迟、高频次场景。
- **异步调用：**以消息中间件为中介，为消息队列中的数据提供预测服务，适合大批量、是实现要求较高的场景。
- **批处理：**对存储在文件系统中的数据进行预测，适合批量、非实时处理的场景。

8.4.2 系统监控

为保障服务稳定运行，要求监控内容如下：

- **容器级监控：**通过容器为模型提供运行时环境，要求监控容器的资源利用率，该指标是监控服务健康持续运行的重要指标。
- **服务级监控：**要求监控对服务调用成功率、响应时长等参数。

8.4.3 服务评估

模型具有时效性，通过服务评估监测模型是否衰退。评估功能可以基于工作流实现，且可以对评估结果导出进行跨时间维度对比。

8.4.4 服务变更

通过上线版本记录服务中的模型变更情况。要求上线版本记录服务中模型的变更情况，支持版本回滚。

参考文献

- [1] JR/T 0221—2021 《人工智能算法金融应用评价规范》
- [2] GB/T 41867-2022 《信息技术人工智能术语》
- [3] GB/T 41479-2022 《信息安全技术 网络数据处理安全要求》
- [4] GB/T 35274-2023 《信息安全技术 大数据服务安全能力要求》

- [5] GB/T 42755-2023 《人工智能 面向机器学习的数据标注规程》
- [6] AIIA/P 0008—2022 《人工智能研发运营一体化(Model/MLOps)能力成熟度模型 第一部分：开发管理》
- [7] AIIA/P 0009—2022 《人工智能研发运营一体化(Model/MLOps)能力成熟度模型 第二部分：模型交付》