

盘古大模型
3.0.0

快速入门

文档版本 01
发布日期 2024-12-02



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目录

1 使用盘古预置 NLP 大模型进行文本对话.....	1
2 调用盘古 NLP 大模型 API 实现文本对话.....	3
3 使用盘古加工算子构建单轮问答数据集.....	5
4 创建盘古图片类数据集标注任务.....	8
5 创建盘古 NLP 大模型 SFT 任务.....	10
6 创建盘古多语言文本翻译 workflow.....	21

1 使用盘古预置 NLP 大模型进行文本对话

场景描述

此示例演示了如何使用盘古能力调测功能与盘古NLP大模型进行对话问答。您将学习如何通过调试模型超参数，实现智能化对话问答功能。

准备工作

请确保您有预置的NLP大模型，并已完成模型的部署操作，详见《用户指南》“开发盘古NLP大模型 > 部署NLP大模型 > 创建NLP大模型部署任务”。

操作流程

1. 登录ModelArts Studio大模型开发平台，进入所需空间。
2. 单击左侧“能力调测”，进入“文本对话”页签，选择服务与人设，参数设置为默认参数，在输入框输入问题，单击“生成”，模型将基于问题进行回答。

图 1-1 使用预置服务进行文本对话



3. 可以尝试修改参数以查看模型效果，示例如下：

- a. 将“核采样”参数调小，如改为0.1，保持其他参数不变，单击“重新生成”，再单击“重新生成”，可以看到模型前后两次回复内容的多样性降低。

图 1-2 “核采样”参数调小后生成结果 1

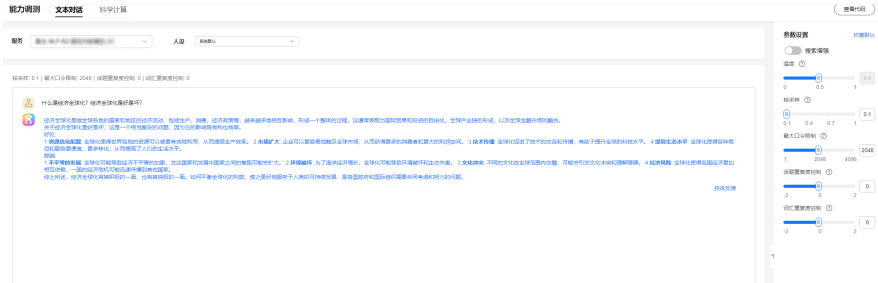
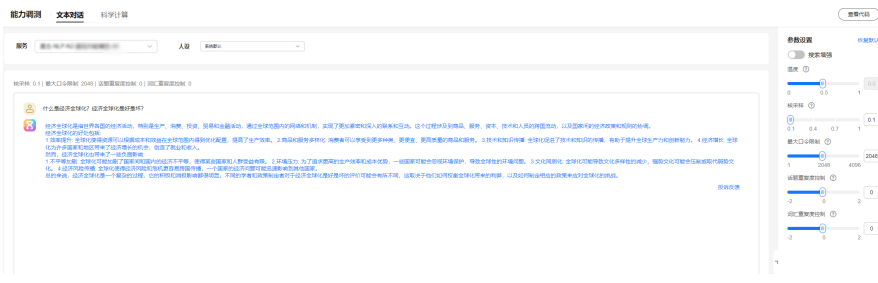


图 1-3 “核采样”参数调小后生成结果 2



- b. 将“核采样”参数调大，如改为1，保持其他参数不变，单击“重新生成”，再单击“重新生成”，可以看到模型前后两次回复内容的多样性提高。

图 1-4 “核采样”参数调大后生成结果 1

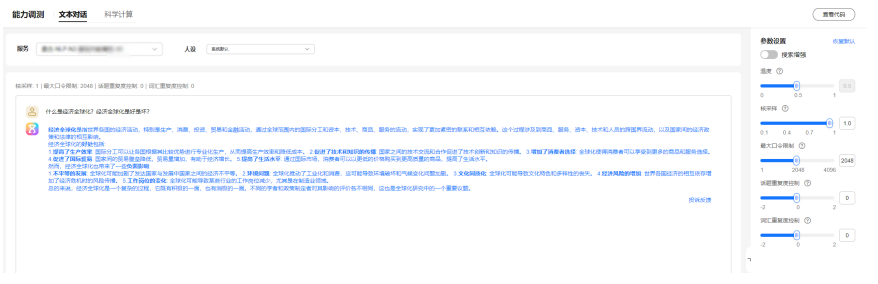
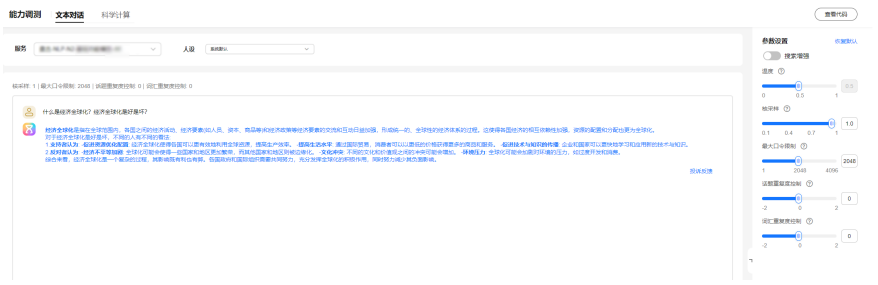


图 1-5 “核采样”参数调大后生成结果 2



2 调用盘古 NLP 大模型 API 实现文本对话

场景描述

此示例演示了如何调用盘古NLP大模型API实现文本对话功能。您将学习如何通过API接口发送请求，传递对话输入，并接收模型生成的智能回复。通过这一过程，您可以快速集成NLP对话功能，使应用具备自然流畅的交互能力。

准备工作

调用盘古NLP大模型API实现文本对话前，请确保您已完成NLP大模型的部署操作，详见《用户指南》“开发盘古NLP大模型 > 部署NLP大模型 > 创建NLP大模型部署任务”。

操作流程

1. 登录ModelArts Studio大模型开发平台，进入所需空间。
2. 获取调用路径及部署ID。单击左侧“模型开发 > 模型部署”，选择所需调用的NLP大模型，单击“调用路径”，在“调用路径”弹窗获取调用路径及部署ID。

图 2-1 获取调用路径和部署 ID



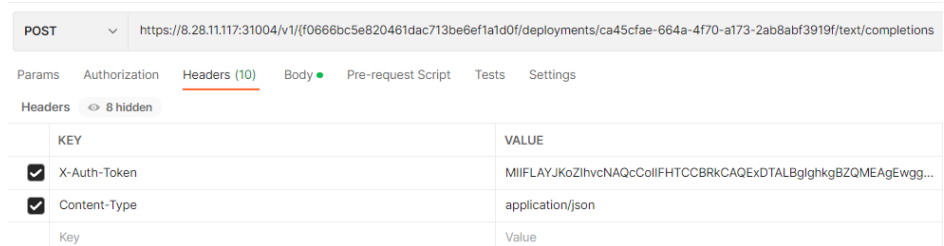
3. 获取项目ID。在页面右上角“我的凭证”，在“API凭证”页面可获取项目ID。

图 2-2 获取项目 ID



4. 获取Token。参考《API参考》文档“如何调用REST API认证鉴权”章节获取Token。
5. 在Postman中新建POST请求，并填入API请求地址。
6. 参考图2-3填写2个请求Header参数。
 - 参数名为Content-Type，参数值为application/json。
 - 参数名为X-Auth-Token，参数值为获取的Token值。

图 2-3 填写 NLP 大模型 API



在Postman中选择“Body > raw”选项，参考以下代码填写请求Body。

```
{
  "messages": [
    {
      "content": "介绍下长江，以及长江中典型的鱼类"
    }
  ],
  "temperature": 0.9,
  "max_tokens": 600
}
```

7. 单击Postman界面“Send”按钮，发送请求。当接口返回状态为200时，表示NLP大模型API调用成功。

3 使用盘古加工算子构建单轮问答数据集

场景描述

此示例演示了如何使用加工算子轻松构建单轮问答数据集。数据集的加工算子是一种灵活的数据预处理工具，能够帮助您将原始数据转化为所需的格式。通过使用加工算子，您可以提取、转换、过滤原始数据，生成适合大模型训练的数据集。

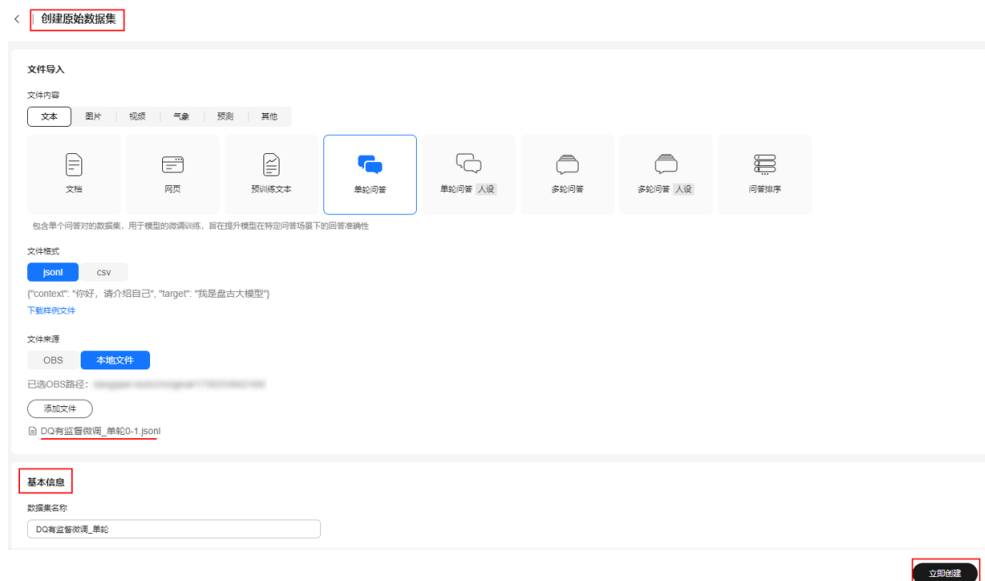
准备工作

请提前准备数据并上传至OBS服务，上传步骤请详见[通过控制台快速使用OBS](#)。

操作流程

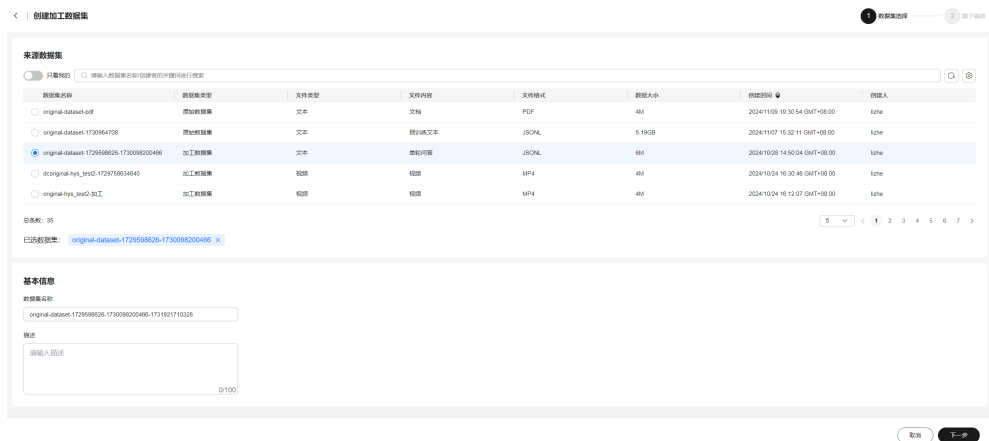
1. 登录ModelArts Studio大模型开发平台，进入所需空间。
2. 选择左侧“数据工程 > 数据获取”，单击右上角“创建原始数据集”。
3. 在“创建原始数据集”页面，选择“文本 > 单轮问答”，选择文件格式、文件来源并添加文件，填写数据集名称及描述，单击“立即创建”。

图 3-1 创建原始数据集



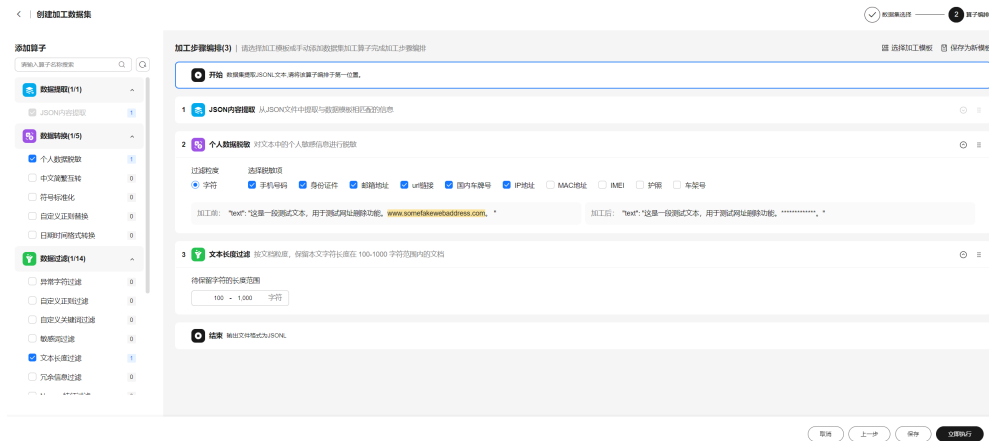
4. 创建成功的数据集的任务状态为“成功”，单击操作列的“上线”按钮，将该数据集上线，用于后续加工操作。
5. 选择左侧“数据工程 > 数据加工”，单击右上角“创建加工数据集”。
6. 在“来源数据集”分页，选择“文件内容”为“单轮问答”的数据集，填写数据集名称和描述，单击“下一步”。

图 3-2 选择数据集



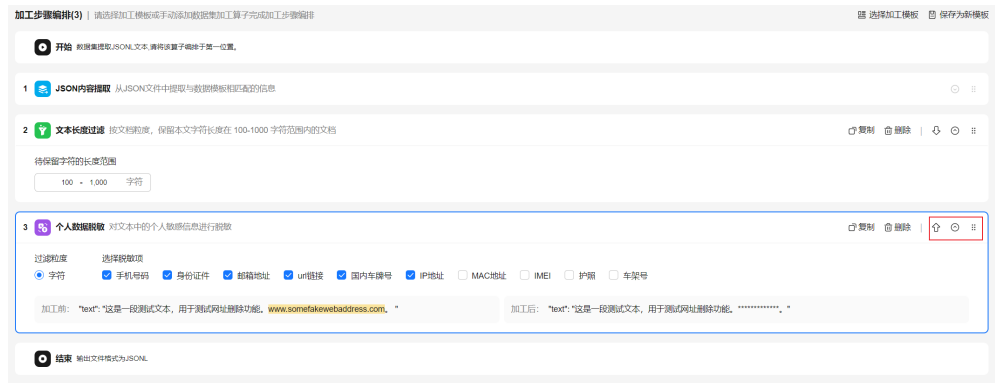
7. 在加工步骤编排页面展示了预先设置好的开始、结束步骤。在左侧“添加算子”分页可选择合适的算子，如个人数据脱敏、文本长度过滤等。导入的数据集格式为“JSONL”，因此默认添加了JSON内容提取算子。

图 3-3 加工算子编排



8. 单击各算子右上角的图标可进行拖动，调整算子的编排顺序。

图 3-4 调整算子编排顺序



9. 算子编排完成后，单击“立即执行”，返回“数据加工”页面，可查看当前数据加工任务的状态。

4 创建盘古图片类数据集标注任务

场景描述

此示例演示了如何快速创建图片Caption数据标注任务。您将学习如何设置任务参数、配置标注要求并进行标注。

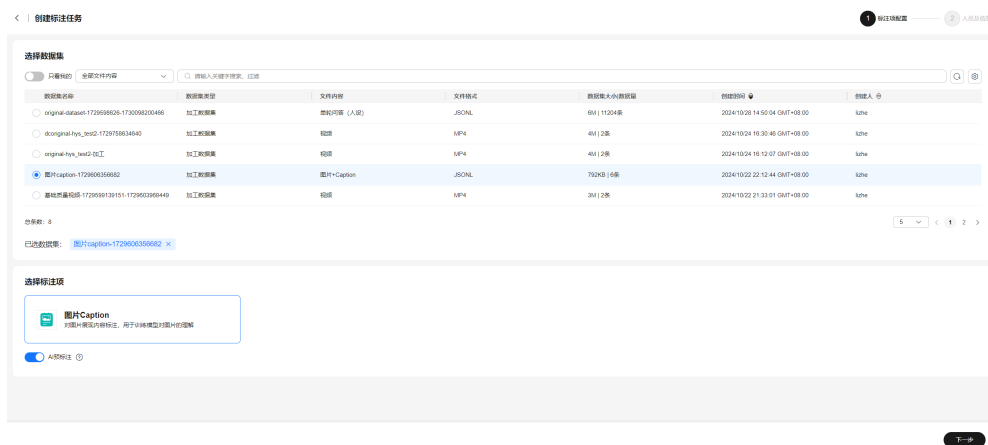
准备工作

请提前准备数据并上传至OBS服务，上传步骤请详见[通过控制台快速使用OBS](#)。

操作流程

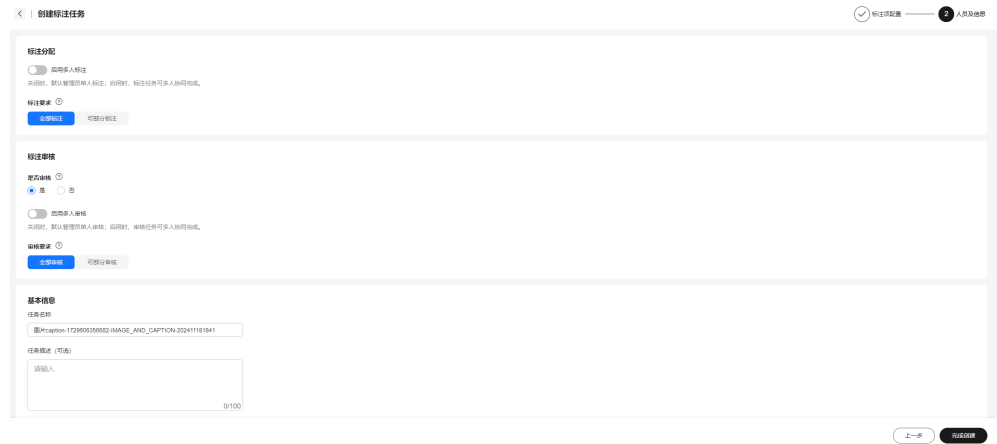
1. 登录ModelArts Studio大模型开发平台，进入所需空间。
2. 选择左侧“数据工程 > 数据获取”，单击右上角“创建原始数据集”。
3. 在“创建原始数据集”页面，选择“图片 > 图片+Caption”，选择文件格式、文件来源，填写数据集名称及描述，单击“立即创建”。
4. 创建成功的数据集的任务状态为“成功”，单击操作列的“上线”按钮，将该数据集上线，用于后续标注操作。
5. 选择左侧“数据工程 > 数据标注 > 标注管理”，单击右上角“创建标注任务”。
6. 在“来源数据集”分页，选择“文件内容”为“图片+Caption”的数据集，选择标注项，可开启“AI预标注”功能，单击“下一步”。

图 4-1 选择标注数据集



7. 在“创建标注任务”页面，可配置标注分配、标注审核和标注任务的基本信息，单击“完成创建”。

图 4-2 配置标注分配、标注审核、基本信息



8. 当标注任务状态为“已创建”时，在操作列单击“上线”按钮。
9. 进入“数据标注 > 标注作业”页面，单击标注任务操作列的“标注”按钮，进入“标注”页面。
10. 在“标注”页面的右下角展示了AI预标注的结果，可参考其内容进行标注。

图 4-3 标注页面



11. 一条数据标注完成后，单击“提交”按钮可继续标注剩余数据，直至所有数据标注完成。

5 创建盘古 NLP 大模型 SFT 任务

场景描述

此示例演示了如何从头创建SFT（有监督微调）训练任务。通过该示例，您将了解以下内容：

- 如何将数据导入平台并进行数据加工、标注和评估操作。
- 如何创建SFT训练任务并配置训练参数，以提升文本理解和生成的质量。
- 如何执行模型的压缩和部署操作。

准备工作

- 请提前准备文本类数据，文本类数据集格式要求请详见《用户指南》“使用数据工程准备与处理数据集 > 数据集格式要求 > 文本类数据集格式要求”。
- 上传文本类数据至OBS服务，上传步骤请详见[通过控制台快速使用OBS](#)。

操作流程

盘古NLP大模型SFT任务创建流程见[表5-1](#)。

表 5-1 盘古 NLP 大模型 SFT 任务创建流程

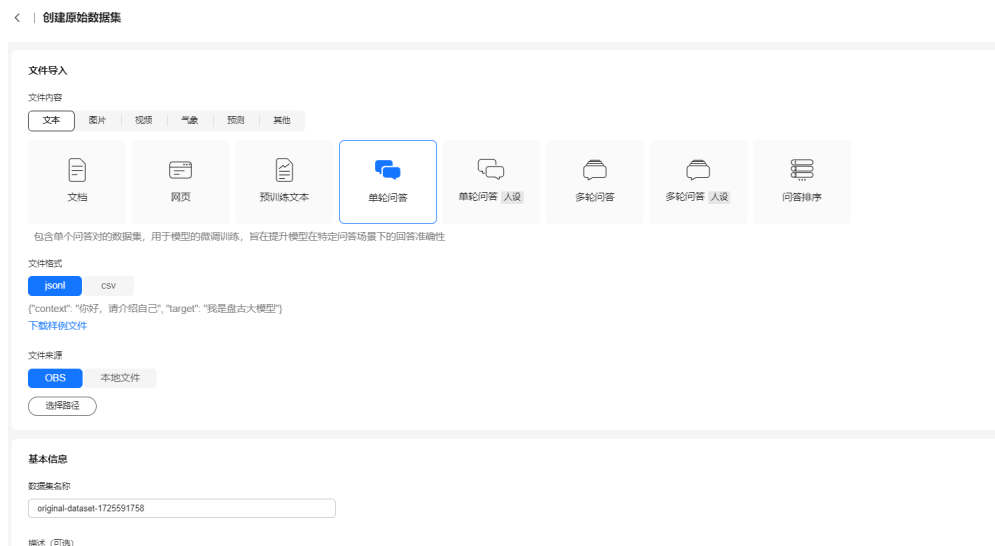
操作步骤	说明
步骤1：导入数据至盘古平台	本样例场景实现将存储在OBS的文本数据导入至盘古平台，并上线为原始数据集。
步骤2：加工文本类数据集	本样例场景帮助用户利用数据集加工算子处理原始数据集。
步骤3：标注文本类数据集	本样例场景帮助用户高效完成数据标注任务，提升标注数据的可靠性和可用性。
步骤4：评估文本类数据集	本样例场景帮助用户利用数据集评估标准评估和优化数据质量。

操作步骤	说明
步骤5: 发布文本类数据集	本样例场景实现将处理好的数据集发布为模型训练可用的数据集。
步骤6: 训练NLP大模型	本样例场景实现NLP大模型的训练操作。
步骤7: 压缩NLP大模型	本样例场景实现NLP大模型的压缩操作。压缩是指通过减少模型的参数量或计算复杂度，在尽量保持模型性能的前提下，减小其存储需求和推理时间，从而提升模型的部署效率，尤其在资源受限的环境中具有重要意义。常见的压缩方法包括剪枝、量化、知识蒸馏等。
步骤8: 部署NLP大模型	本样例场景实现NLP大模型的部署操作。

步骤 1: 导入数据至盘古平台

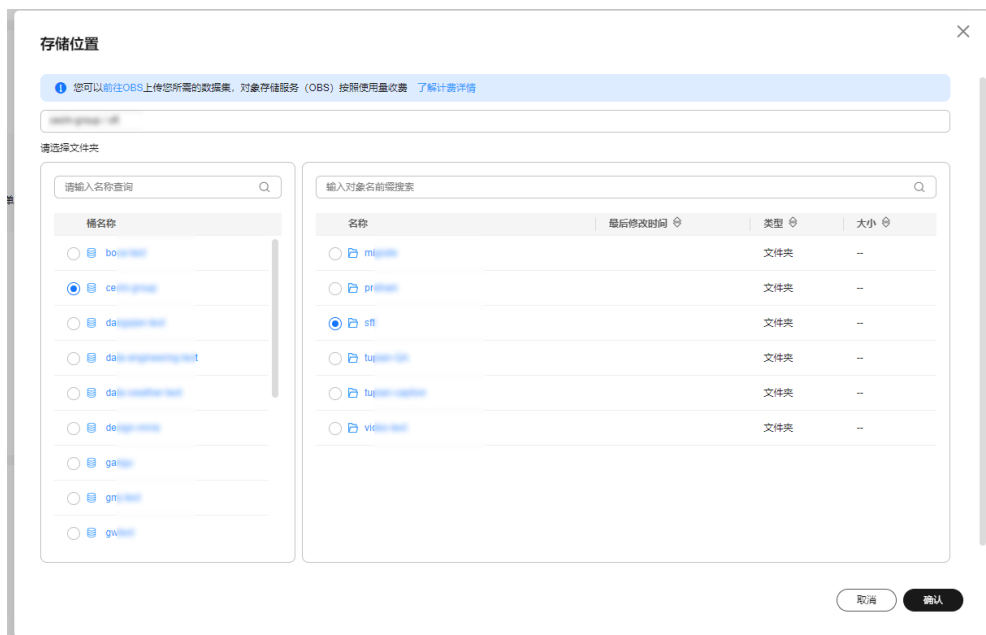
1. 登录ModelArts Studio大模型开发平台，进入所需空间。
2. 选择左侧“数据工程 > 数据获取”，单击右上角“创建原始数据集”。
3. 在“创建原始数据集”页面，选择“文本 > 单轮问答”，选择文件格式、文件来源，填写数据集名称及描述，单击“立即创建”。

图 5-1 创建原始数据集



4. 单击“选择路径”，在“存储位置”弹窗中选择需导入的数据，单击“确认”。

图 5-2 选择导入的数据



5. 数据集信息设置完成后，填写“数据集名称”和“描述”，并设置“拓展信息”。

拓展信息包括“标签设置”与“数据版权”：

- 标签设置。通过标签设置，可以给数据集添加行业、语言、标签信息。
- 数据版权设置。训练模型的数据集除用户自行构建外，也可能会使用开源的数据集。数据版权功能主要用于记录和管理数据集的版权信息，确保数据的使用合法合规，并清晰地了解数据集的来源和相关的版权授权。通过填写这些信息，可以追溯数据的来源，明确数据使用的限制和许可，从而保护数据版权并避免版权纠纷。

图 5-3 设置数据版权

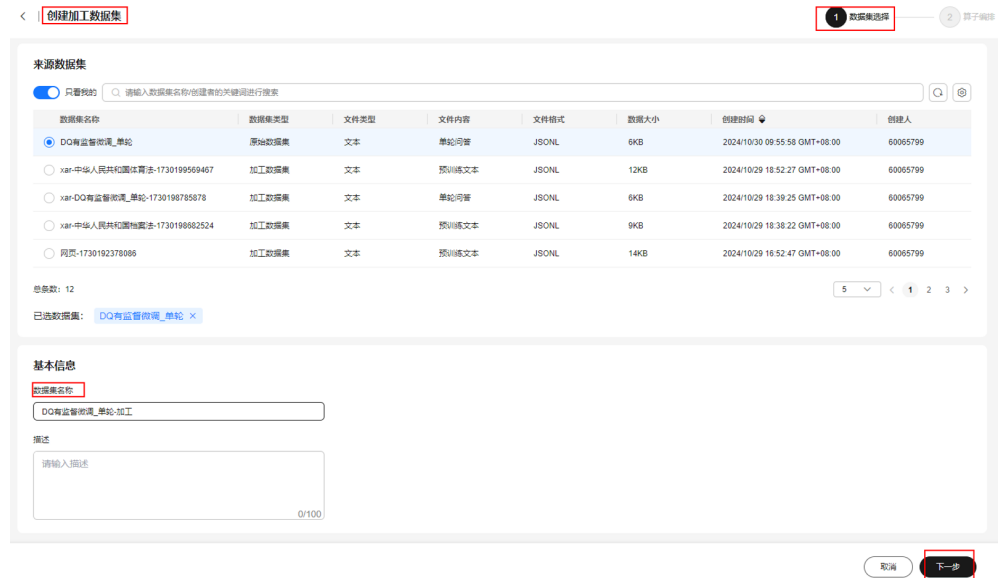


6. 单击页面右下角“立即创建”完成原始数据集的创建操作。创建成功的数据集的任务状态为“成功”，单击操作列的“上线”按钮，将该数据集上线，用于后续加工操作。

步骤 2：加工文本类数据集

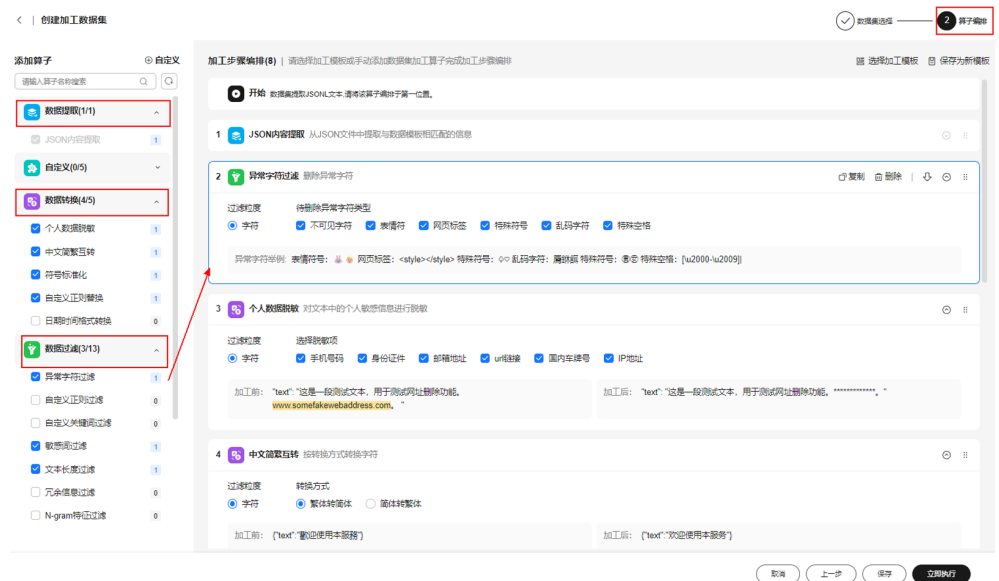
1. 选择左侧“数据工程 > 数据加工”，单击右上角“创建加工数据集”。
2. 在“创建加工数据集”页面，选择步骤1中创建好的原始数据集，填写数据集名称及描述，单击“下一步”。

图 5-4 创建加工数据集



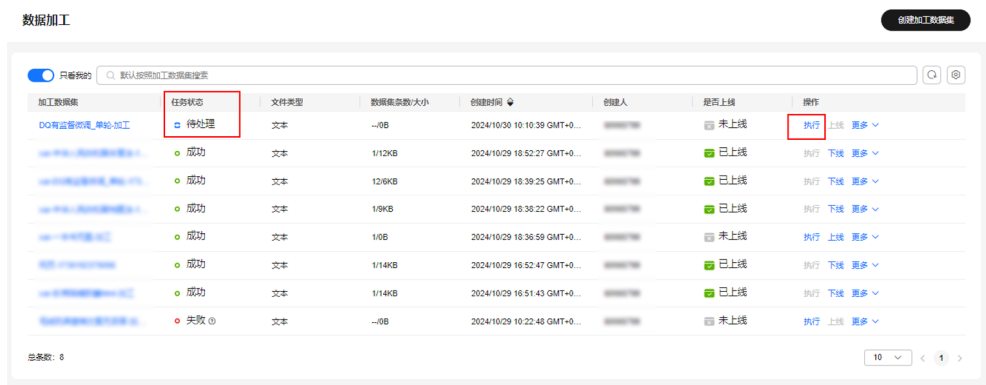
3. 在左侧算子列表中添加相应的数据加工算子，单击“立即执行”。

图 5-5 加工算子编排



4. 单击“执行”，开始执行加工任务，加工成功的数据集的任务状态为“成功”，单击“上线”，用于后续标注操作。

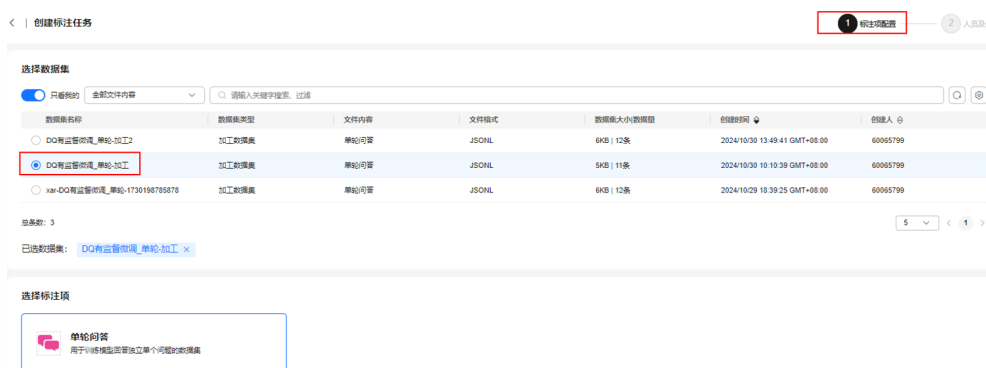
图 5-6 上线加工后的数据集



步骤 3: 标注文本类数据集

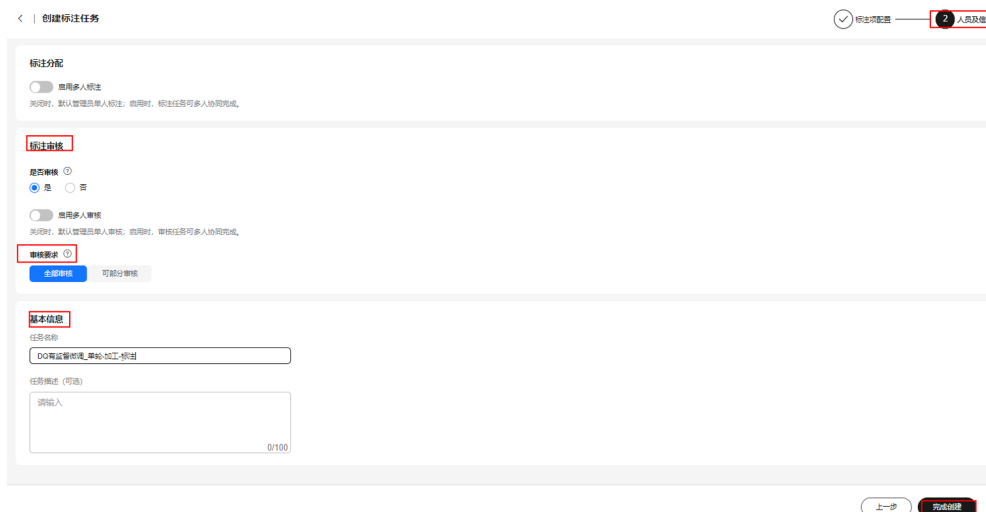
1. 选择左侧“数据工程 > 数据标注 > 标注管理”，单击右上角“创建标注任务”。
2. 在“创建标注任务”页面，选择**步骤2: 加工文本类数据集**中加工好的数据集，选择标注项“单轮问答”，单击“下一步”。

图 5-7 选择加工后的数据集



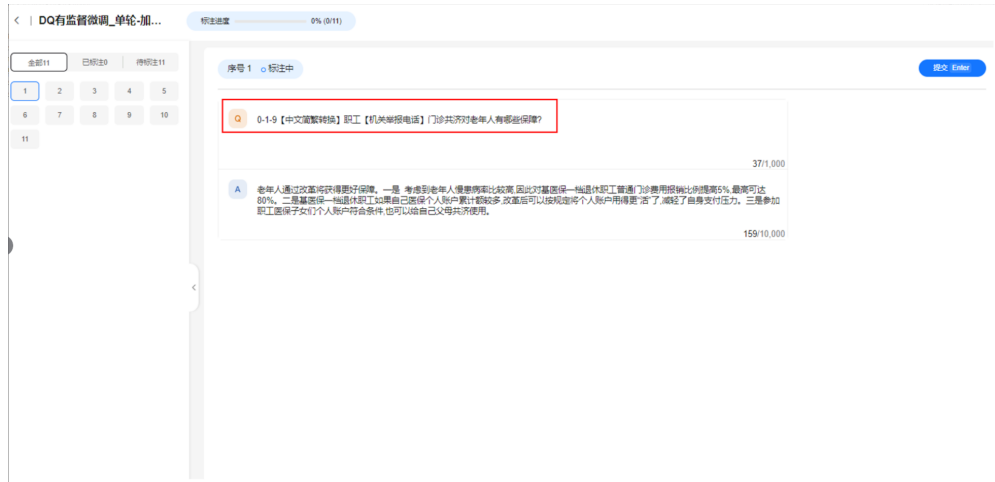
3. 配置“标注审核”与“基本信息”，单击“完成创建”。

图 5-8 配置标注审核、基本信息



4. 选择左侧导航栏的“数据工程 > 数据标注 > 标注作业”，单击数据集操作列的“标注”，进入数据集的标注页面。
5. 在标注页面后，需要逐一确认问题（Q）及答案（A）是否正确，如果问题或答案不正确，可以对其进行二次编辑。

图 5-9 标注页面



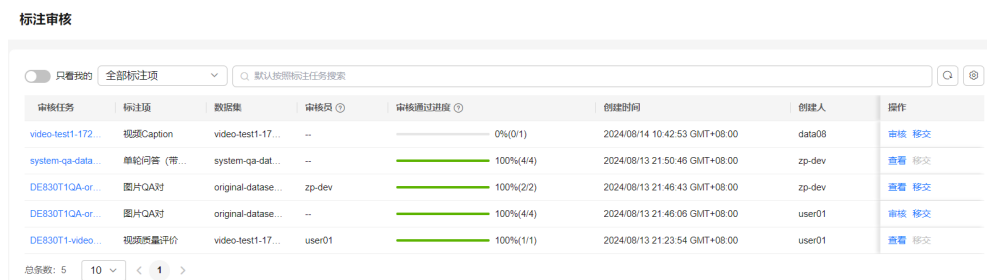
6. 一条数据标注完成后，单击“提交”按钮可继续标注剩余数据。所有数据标注完成后，页面会出现标注任务成功的提示。

图 5-10 标注任务完成提示



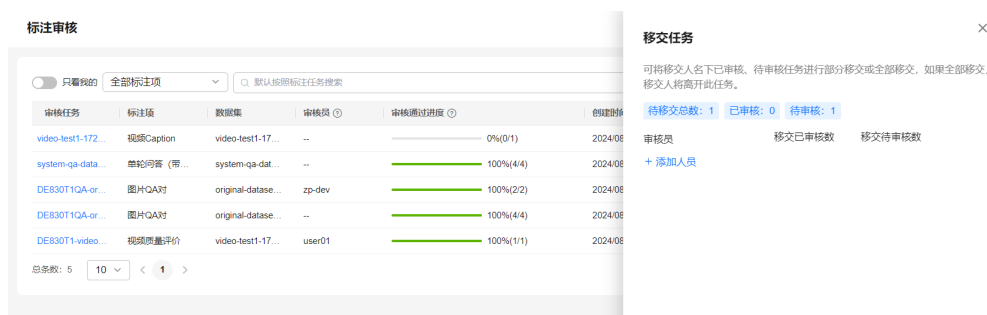
7. 创建数据集标注任务时，如果设置了启用标注审核，在完成标注后可以在“标注审核”页面审核标注结果。

图 5-11 标注审核页面



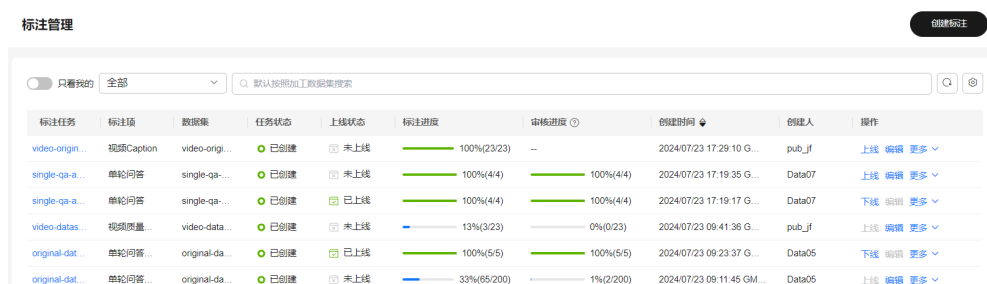
- 在“标注审核”页面，单击操作列“审核”可进入审核页面审核数据。如果需要将该审核任务移交给其他人员，可以单击操作列“移交”设置移交人员以及移交的数量。

图 5-12 移交任务页面



- 进入审核页面后，可通过单击“通过”或“不通过”按钮逐一一对数据进行审核，直至所有数据审核完成，期间可对不满足要求的数据进行驳回，驳回后将分给标注人员重新标注。
- 数据集审核无问题后，可在“标注管理”页面对数据集进行上线。

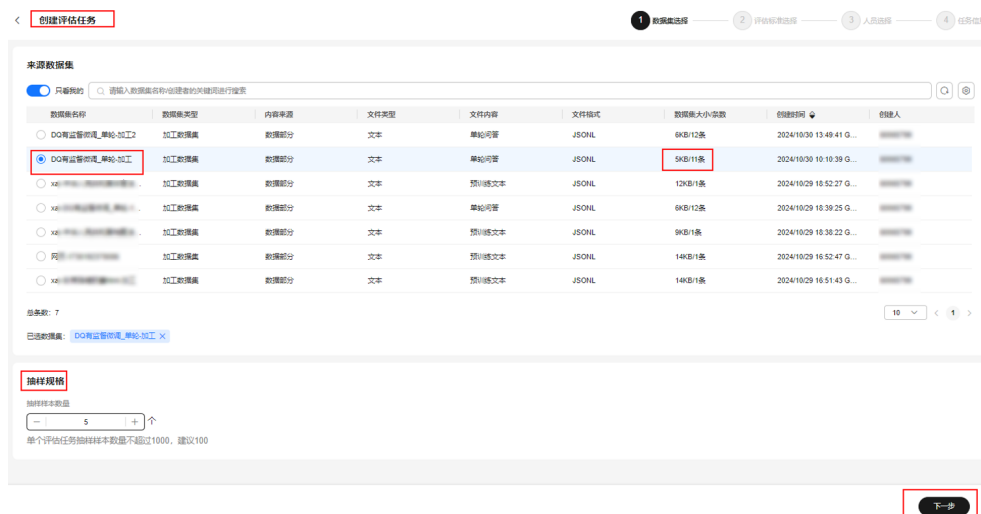
图 5-13 标注管理页面



步骤 4：评估文本类数据集

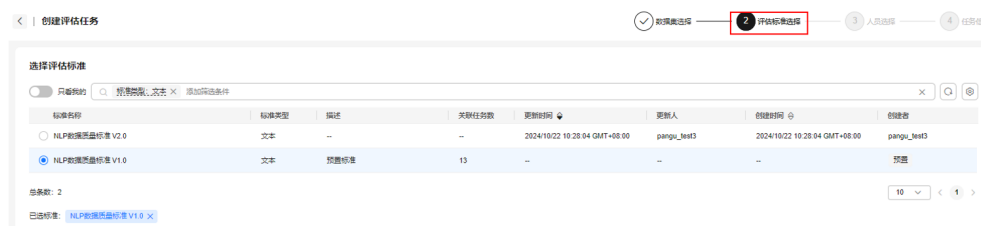
- 选择左侧“数据工程 > 数据评估 > 评估任务”，单击右上角“创建评估任务”。
- 在“创建评估任务”页面，选择数据集，设置抽样规格，单击“下一步”。

图 5-14 创建评估任务



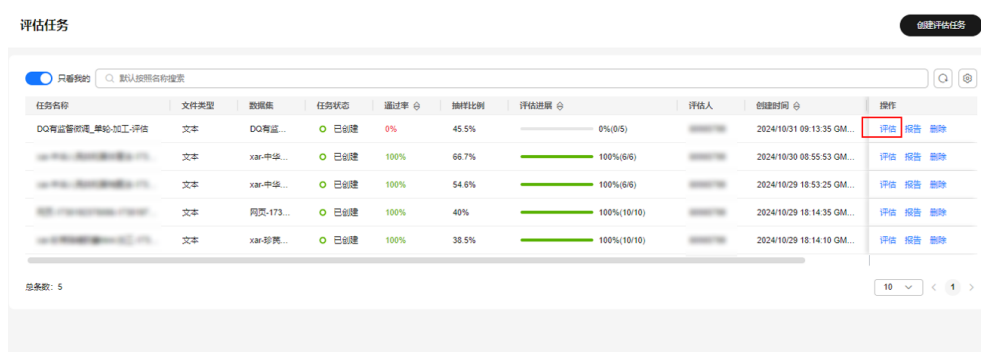
3. 根据数据集类型选择评估标准，此处选择“NLP数据质量标准V1.0”，单击“下一步”。

图 5-15 选择评估标准



4. 选择评估人员，单击“下一步”。填写“任务名称”，单击“完成创建”，在数据集操作列单击“评估”。

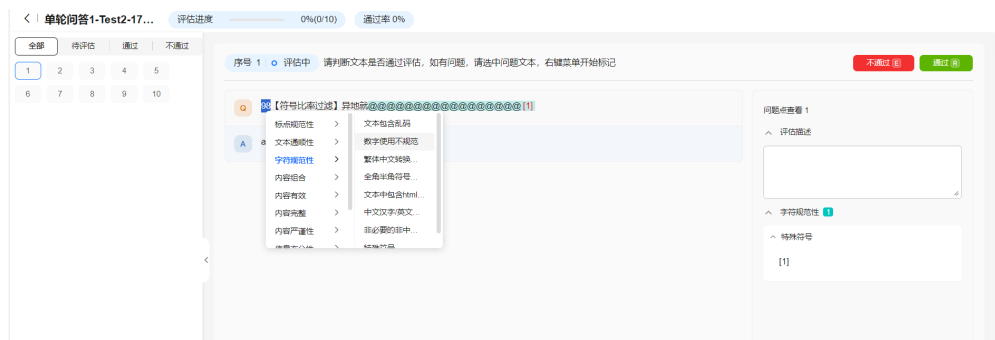
图 5-16 评估任务列表



5. 在评估页面，可参考评估项对当前数据的问题进行标注，且不满足时需要单击“不通过”，满足则单击“通过”。

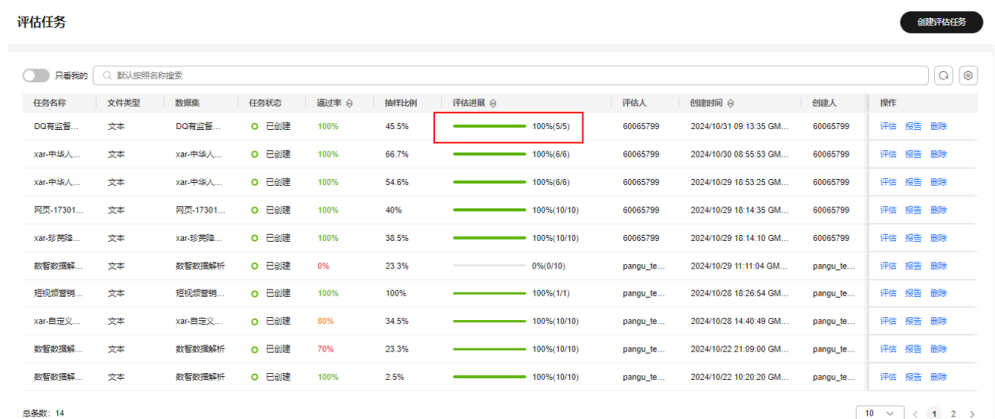
对于文本类数据集而言，可选择问题内容后，单击鼠标右键进行数据问题的标注。

图 5-17 标记数据集问题



- 全部数据评估完成后，评估状态显示为“100%”，表示当前数据集已经评估完成，可以回退到“评估任务”页面，查看，单击操作列“报告”按钮，获取数据集质量评估报告。

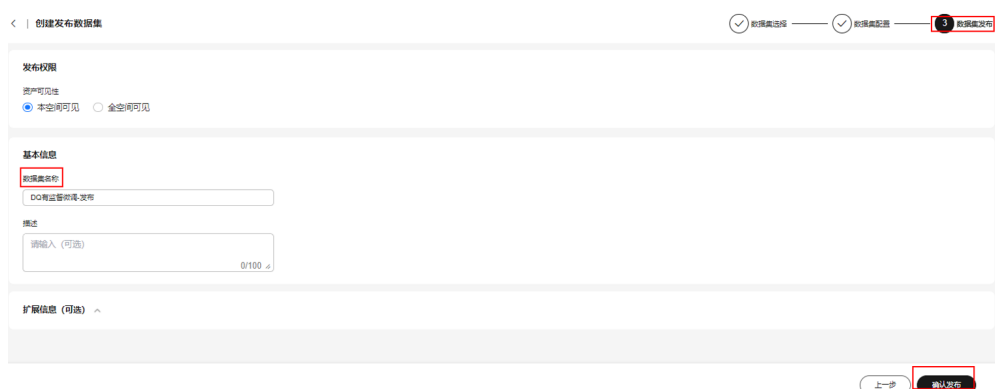
图 5-18 查看评估任务进展



步骤 5：发布文本类数据集

- 选择左侧“数据工程 > 数据发布”，单击右上角“创建发布数据集”。
- 在“创建发布数据集”页面，选择“文本 > 单轮问答”，选择“发布方式”为“单个数据集”，选择评估后的数据集，单击“下一步”。
- 选择“格式配置”为“盘古格式”，单击“下一步”。
- 选择资产可见性，填写数据集名称，单击“确认发布”。

图 5-19 创建发布数据集



- 发布成功的数据集的发布状态为“成功”，可用于后续的训练操作。

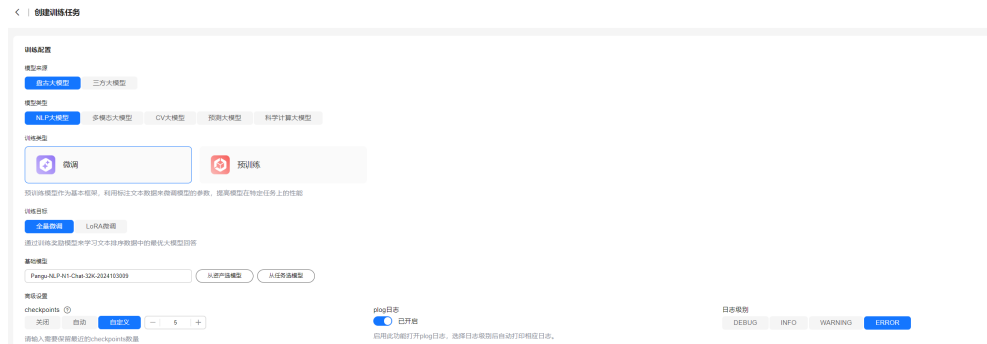
图 5-20 查看数据集发布状态



步骤 6: 训练 NLP 大模型

- 选择左侧“模型开发 > 模型训练”，单击右上角“创建训练任务”。
- 在“创建训练任务”页面，依次选择“盘古大模型 > NLP大模型 > 微调 > 全量微调”。
- 在“基础模型”中，选择“从资产选模型 > 预置”，选择所需NLP大模型，单击“确定”。
- 在“高级设置”中，依次配置checkpoints、plog日志以及日志级别。

图 5-21 创建训练任务页面



- 在“训练参数”中，平台已经预置了默认的模型参数，也可根据需求自行修改。
- 在“数据配置”中选择训练数据集。填写基本信息后，单击“立即创建”。
- 创建好训练任务后，自动返回至“模型训练”页面，模型将自动开始训练，训练完成后的任务状态为“已完成”。

步骤 7: 压缩 NLP 大模型

- 在左侧导航栏中选择“模型开发 > 模型压缩”，单击界面右上角“创建压缩任务”。
- 在“创建压缩任务”页面，单击“从资产选模型”，选择步骤6: 训练NLP大模型步骤中训练好的大模型，选择压缩策略，填写任务名称、压缩后模型名称及描述。
- 参数填写完成后单击“立即创建”创建模型压缩任务。

步骤 8：部署 NLP 大模型

1. 在左侧导航栏中选择“模型开发 > 模型部署”，单击“创建部署”。
2. 在“创建部署”页面，选择“盘古大模型 > NLP大模型”，单击“从资产选模型”，选择[步骤7：压缩NLP大模型](#)步骤中压缩后的模型。
3. 选择部署方式“云上部署”，选择最大TOKEN长度，如“4096”，选择架构类型“ARM”。
4. 在“资源配置”中配置实例数，如“1”，“安全护栏”功能已默认开启。
5. 填写服务名称及描述，单击“立即部署”，启动模型部署。

6 创建盘古多语言文本翻译 workflow

场景描述

该示例演示了如何创建多语言文本翻译 workflow，示例将调用华为云文本翻译 API，详见[文本翻译](#)。

准备工作

- 提前开通“文本翻译”服务。登录[自然语言处理控制台](#)，在总览页面下方开通“文本翻译”服务，单击“开通服务”。

图 6-1 开通文本翻译服务

服务	开通状态	计费状态	套餐包剩余可用额度/总额度	操作
基础版API	-	-	您还未购买套餐包	购买套餐包 一键开通
领域版API	-	-	您还未购买套餐包	购买套餐包 一键开通
定制版API	-	-	-	一键开通
免费版API	-	-	-	一键开通
文本翻译	未开通	-	-	开通服务
文档翻译	未开通	-	-	开通服务

- 提前获取**文本翻译API调用地址**，该服务支持区域为**华北-北京四**。
调用地址的示例如下：
`https://nlp-ext.cn-north-4.myhuaweicloud.com/v1/{project_id}/machine-translation/text-translation`
其中，**project_id**为所属区域的项目ID，获取方式见[获取项目ID](#)。
- 提前获取**文本翻译服务Token**，请参考[认证鉴权](#)。请获取区域为**华北-北京四**的Token。
- 为确保持有可用的NLP大模型，请先完成NLP大模型部署操作，详见《[用户指南](#)》“[开发盘古NLP大模型](#) > [部署NLP大模型](#) > [创建NLP大模型部署任务](#)”。

操作流程

创建盘古多语言文本翻译 workflow 的流程见[表6-1](#)。

表 6-1 创建盘古多语言文本翻译 workflow 流程

操作步骤	说明
步骤1: 创建并配置插件	本样例场景实现文本翻译插件的创建与配置。
步骤2: 创建并配置 workflow	本样例场景实现多语言文本翻译 workflow 的创建与配置。
步骤3: 试运行 workflow	本样例场景实现 workflow 的试运行。

步骤 1: 创建并配置插件

1. 登录 ModelArts Studio 大模型开发平台，进入所需空间。
2. 在左侧导航栏中选择“Agent 开发”，将跳转至 Agent 开发平台。
3. 在 Agent 开发平台，单击左侧导航栏“工作台”按钮，在“插件”页签，单击右上角“创建插件”按钮。
4. 输入插件名称及插件描述，单击“下一步”。

图 6-2 创建插件

创建插件

1 基本信息 — 2 配置信息 — 3 参数信息

* 插件名称

机器翻译

* 插件描述

一个机器翻译插件。 |

9/1,600

5. 填写插件的 URL（**准备工作**中获取的文本翻译 API 调用地址），选择请求方式为“POST”。权限校验选择“用户级鉴权 > Header”，填写密钥鉴权参数名为 X-Auth-Token、密钥来源参数名为 X-Auth-Token，请求头为{"Content-Type": "application/json"}，单击“下一步”。

图 6-3 配置插件信息

创建插件 ×

基本信息 ———— **2 配置信息** ———— 3 参数信息

★ 插件URL ★ 请求方式

权限校验

无需鉴权 用户级鉴权 API Key

秘钥位置

Header Query

参数列表 ^

密钥鉴权参数名	密钥来源参数名	
<input type="text" value="X-Auth-Token"/>	<input type="text" value="X-Auth-Token"/>	<input type="button" value="🗑"/>

[+ 添加参数](#)

请求头

- 配置参数信息，如图6-4。
配置完成后，单击“确定”，完成文本翻译插件的创建。

图 6-4 配置参数信息



步骤 2: 创建并配置 workflow

1. 单击左侧导航栏“工作台”按钮，在“workflow”页签，单击右上角“创建工作流”按钮。
2. 输入 workflow 名称及描述，单击“确定”，进入 workflow 编排页面。

图 6-5 创建工作流



3. 在工作流编排页面，平台已预先编排了开始、大模型与结束组件。


单击组件右上角的 ，可以进行组件的重命名、复制、删除操作。开始和结束组件为必选组件，无法删除。

图 6-6 组件的重命名、复制、删除操作



4. 鼠标拖动左侧“意图识别”组件至编排页面，并连接开始组件和意图识别组件。单击意图识别组件进行配置操作。

图 6-7 连接组件操作



5. 在“参数配置”中，配置输入参数。
 - **参数名称：**默认名称input，为固定值，不可编辑。
 - **取值：**选择“引用 > query”。query为开始组件的输出变量值。
6. 在“分类配置”中，填写相关情景分类。

相关分类为针对该分类的描述语句或者关键词，同时也将作为大模型进行推理和分类的依据。分类数量为2 ~ 5个。

意图识别 [🔗](#)

✕

参数配置

输入参数 ^

参数名称	取值
input	引用 <input type="text" value="query"/>

分类配置 ^

分类1

文本翻译

分类2

其他

添加场景

7. 在“高级配置”中配置预先部署好的模型与prompt提示词。单击“确定”，完成参数配置。

图 6-8 意图识别组件参数配置



- 鼠标拖动左侧“提问者”组件至编排页面，并连接意图识别组件与该组件，单击该组件进行配置操作。提问者组件参数配置见图6-9。

其中，请将“高级配置 > Prompt提示词”中的内容替换为：

请你从目标数据中提取翻译相关参数，参数包括origintext（待翻译文本），from（原始语言），to（翻译目标语言）。目标数据未提供或者有歧义（如存在多个）的变量请保证取值为空''。from的取值范围为ar、de、ru、fr、ko、pt、ja、th、es、en、vi、zh、auto。ar为阿拉伯语，de为德语，ru为俄语，fr为法语，ko为韩语，pt为葡萄牙语，ja为日语，th为泰语，tr为土耳其语，es为西班牙语，en为英语，vi为越南语，zh为中文，auto为不指定由插件自行决定语种。to的取值范围是ar、de、ru、fr、ko、pt、ja、th、es、en、vi、zh。ar为阿拉伯语，de为德语，ru为俄语，fr为法语，ko为韩语，pt为葡萄牙语，ja为日语，th为泰语，tr为土耳其语，es为西班牙语，en为英语，vi为越南语，zh为中文。请注意：不要使用任何工具、不用理会问题的具体含义，并保证你的输出仅有json格式的结果数据，以保证返回结果可以被json.dumps直接解析。你的返回格式格式示例为：{"origintext":"a","from":"b","to":"c"}。

图 6-9 提问器参数配置

参数配置

输入参数 ^

参数名称	取值
<input type="text" value="text"/>	引用 <input type="text" value="query"/>

+ 添加参数

输出参数 ^

参数名称	类型	描述
<input type="text" value="origintext"/>	String	<input type="text" value="待翻译文本"/>
<input type="text" value="from"/>	String	<input type="text" value="原始语言代码"/>
<input type="text" value="to"/>	String	<input type="text" value="翻译目标语言代码"/>

+ 添加参数

问题配置

问题

请输入

高级配置 ^

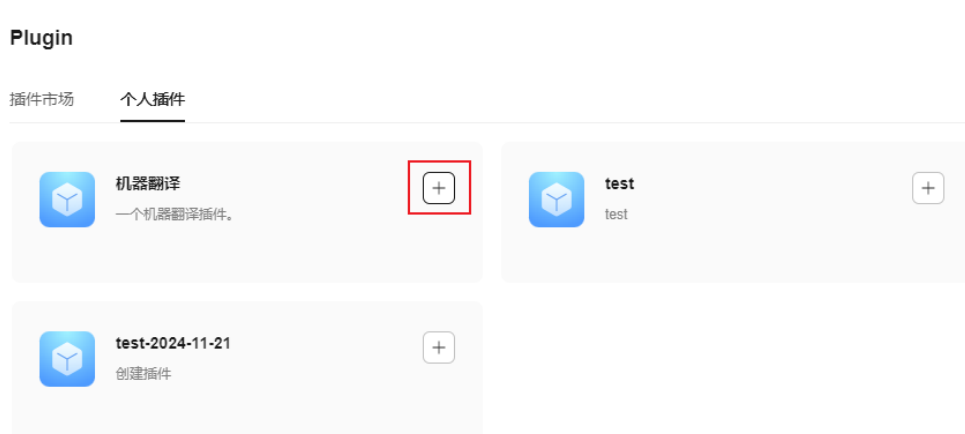
模型选择

Prompt提示词

请你从目标数据中提取翻译相关参数，参数包括origintext（待翻译文本），from（原始语言），to（翻译目标语言）。目标数据未提供或者有歧义（如存在多个）的变量请保证取值为空''。from的取值范围为ar、de、ru、fr、ko、pt、ja、th、es、en、vi、zh、auto。ar为阿拉伯语，de为德语，ru为俄语，fr为法语，ko为韩语，pt为葡萄牙语，ja为日语，th为泰语，tr为土耳其语，es为西班牙语，en为英语，vi为越南

- 鼠标拖动左侧“插件”组件至编排页面，在“Plugin”弹窗中单击“个人插件”，添加步骤6创建的机器翻译插件，连接提问器组件和插件组件，再连接插件组件和结束组件。

图 6-10 添加插件



10. 单击插件组件，按照图6-11，进行参数配置，单击“确定”。

图 6-11 配置文本翻译插件参数



11. 鼠标拖动左侧“大模型”组件至编排页面，连接意图识别组件和该组件，再连接该组件和结束组件。单击该组件进行配置操作。
大模型组件参数配置见图6-13，其中，“模型选择”中需选择已经部署的NLP大模型。

图 6-12 大模型组件参数配置



图 6-13

12. 单击结束组件，配置前置组件的输出参数。

图 6-14 配置结束组件输入参数



13. 配置完成的工作流编排页面见图6-15。

图 6-15 文本翻译 workflow 编排



步骤 3: 试运行 workflow


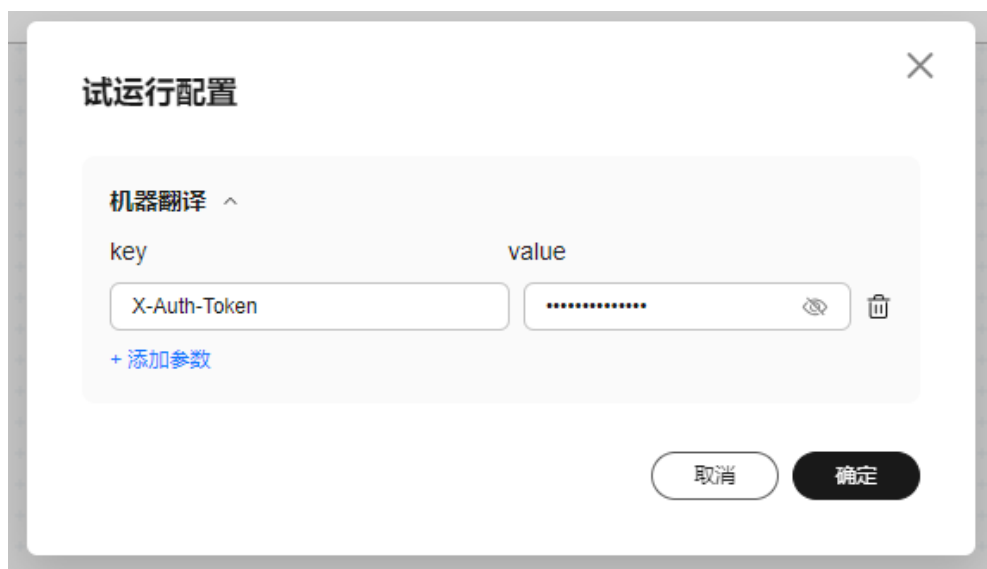
1. 配置文本翻译插件的Token。单击右上角“试运行”，单击 ，再单击“添加参数”，填写文本翻译的鉴权参数X-Auth-Token及value值（[准备工作中](#)中获取的Token），单击“确定”。

图 6-16 配置试运行参数



2. 试运行 workflow。在“试运行”页面，输入对话。
示例如图6-17所示，当用户输入问题时，workflow的意图识别组件将对用户意图进行分类，输出符合用户意图的回答。

图 6-17 试运行 workflow

