盘古大模型

帮助面板 3.7.0

文档版本	01
发布日期	2025-09-17

华为云计算技术有限公司

商标声明


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 数据资产

解释说明

数据资产是指在平台中被纳入管理、存储并可供使用的数据集。

数据资产包括以下两种形式：

- **用户自行发布的数据集**。

  用户可以通过“数据工程 > 数据发布 > 发布任务”功能将数据集发布为数据资产。发布的数据集支持查看详细信息、编辑属性、删除以及发布至AI Gallery等操作。

- **从AI Gallery订阅的数据资产**。

  除了用户自行发布的数据集，平台还提供了从AI Gallery中订阅数据资产的功能。AI Gallery提供了模型、数据集、AI应用等AI数字资产的共享，为企业级或个人开发者等群体，提供安全、开放的共享及交易环节。

发布数据资产至 AI Gallery

1. 单击数据资产（状态为“未发布到Gallery”）操作列的“发布到Gallery”，对数据资产进行发布。
2. 在“发布到AI Gallery”页面填写AI Gallery资产名称，选择可订阅区域约束与可看范围，输入资产描述，单击“确定”，发布数据资产至AI Gallery。
3. 数据资产列表页将显示发布数据资产的状态：
   - 如果状态为“发布中”，表示该资产正在同步至AI Gallery，请耐心等待。
   - 如果状态为“发布成功”，表示该资产已同步至AI Gallery，可单击操作列“查看发布信息”以查看该资产的发布信息。
   - 如果状态为“发布失败”，表示该资产未成功同步至AI Gallery，可单击“发布到Gallery”重新对数据资产进行发布。

从 AI Gallery 订阅数据资产

1. 单击右上角“订阅数据”，在“从AI Gallery订阅”页面选择需订阅的数据资产，单击“下一步”。
2. 填写资产名称与资产描述后，单击“确定”实现数据资产的订阅。
3. 数据资产列表页将显示订阅数据资产的状态：
   - 如果状态为“订阅中”，表示该资产正从AI Gallery同步中，请耐心等待。

- 如果状态为“订阅成功”，表示该资产已从AI Gallery订阅成功，可单击操作列“查看订阅信息”以查看该资产的订阅信息。
- 如果状态为“订阅失败”，表示该资产未成功从AI Gallery订阅，可单击“重新订阅”重新从AI Gallery订阅数据资产。

管理数据资产

1. 用户自行发布的数据集可执行以下操作：
   - 查看基本信息。在“任务发布”页签，单击具体数据资产，可查看资产的配比详情、数据详情等基本信息。
   - 发布至Gallery。在“任务发布”页签，单击操作列的“发布至Gallery”，可发布数据资产至AI Gallery。
   - 查看发布信息。在“任务发布”页签，单击“查看发布信息”，查看该资产的发布信息（该操作需提前发布该数据资产至AI Gallery）。
   - 编辑属性。在“任务发布”页签，单击操作列的“更多 > 编辑属性”，可编辑数据资产的名称、描述以及资产可见性。
   - 删除。在“任务发布”页签，单击操作列的“更多 > 删除”，可删除当前数据资产。
   - 取消发布。在“任务发布”页签，单击操作列的“更多 > 取消发布”，可将已发布至AI Gallery的数据资产取消发布（该操作需提前发布该数据资产至AI Gallery）。
2. 从AI Gallery订阅的数据资产可执行以下操作：
   - 查看订阅信息。在“AI Gallery”页签，单击具体数据资产或操作列的“查看订阅信息”，查看该资产的名称描述等订阅信息。
   - 编辑属性。在“AI Gallery”页签，单击操作列的“更多 > 编辑属性”，可编辑数据资产的名称、描述以及资产可见性。
   - 删除。在“AI Gallery”页签，单击操作列的“更多 > 删除”，可删除当前数据资产。
   - 重新订阅。如果订阅失败，在“AI Gallery”页签，单击操作列的“重新订阅”，可重新从AI Gallery订阅所需数据资产。

2 创建导入任务

解释说明

用户将存储在OBS服务中的数据导入ModelArts Studio平台后，数据将以“原始数据集”的形式被统一管理。

操作步骤

1. 在“创建导入任务”页面，选择“数据集类型”、“文件格式”和“导入来源”。
2. 选择需导入的数据，单击“确定”。
3. 填写“数据集名称”和“描述”。
4. 可选择填写“扩展信息”，包括“数据集属性”、“数据集版权”：
   - 数据集属性。可以给数据集添加行业、语言、自定义信息。
   - 数据集版权。主要用于记录数据集的版权信息，确保数据的使用合法合规，便于用户清晰地了解数据集来源和版权授权。
5. 单击右下角“立即创建”完成数据的导入操作。

3 创建加工任务

解释说明

数据加工是数据工程中的核心环节，旨在通过使用数据集加工算子对数据进行预处理操作，以确保数据符合模型训练的标准和业务需求。

当前支持加工的数据集类型为：文本类、视频类、图片类、气象类。

操作步骤

1. 在“创建加工任务”页面，选择需要加工的数据集。
2. 单击“下一步”，进入“加工步骤编排”页面，左侧列表为当前数据集可选择的加工算子。
   a. 在左侧“添加算子”分页勾选所需算子。
   b. 在右侧“加工步骤编排”页面配置各算子参数，可拖动算子以调整执行顺序。
   c. 在编排过程中，可单击右上角“保存为新模板”将当前编排流程保存为模板。后续创建新的数据加工任务时，可直接单击“选择加工模板”进行使用。

      若选择使用加工模板，将删除当前已编排的加工步骤。
3. 加工步骤编排完成后，单击“下一步”进入任务配置页面。
   – 参考表3-1填写资源配置参数。除了如下参数配置外，支持用户自定义参数配置。

表 3-1 参数配置

参数名称	参数说明
numExecutors	Executor的数量，默认值2。 numExecutors * executorMemory最小值为4，最大值为16。

参数名称	参数说明
executorCores	每个Executor进程使用的CPU内核数量，默认值2。 numExecutors * executorMemory最小值为4，最大值为16。executorCores和executorMemory的比例需要在1:2~1:4之间。
executorMemory	每个Executor进程使用的内存数量，默认值4。 executorCores和executorMemory的比例需要在1:2~1:4之间。
driverCores	驱动程序进程使用的CPU内核数量，默认值2。 driverCores和driverMemory的比例需要在1:2~1:4之间。
driverMemory	驱动程序进程使用的内存数量，默认值4。 driverCores和driverMemory的比例需要在1:2~1:4之间。

- 自动生成加工数据集，启用后任务运行成功自动生成加工数据集，可用于下游数据集发布；如关闭需在加工任务列表操作生成。

  需要填写数据集名称、描述和扩展信息（扩展信息为可选项，主要包括行业、语言和自定义信息）。

4. 单击右下角“启动加工”，将启动加工任务。

4 创建数据指令

解释说明

数据指令通常由用户定义的输入变量（如带{{}}符号的参数）和Prompt提示词指令组成，旨在通过自动化的方式根据用户的需求生成或处理数据。

当前数据指令支持数据合成任务，支持的数据集类型为：文本类。

操作步骤

本章将以“生成主题散文”的场景为例，详细介绍数据合成指令的配置步骤。

1. 在创建数据指令页面，进行指令配置。
   a. 选择变量标识符为“双大括号{{}}”。
   b. 输入指令为“请以{{topic}}为主题，写一篇字数不超过{{num}}的散文。”
   c. 单击“识别”，再单击“确定”。
2. 按照表4-1进行变量配置。

   表 4-1 数据指令变量配置

变量类型	变量名称	数据类型	变量描述
输入变量	topic	string	主题
	num	string	字数
输出变量	output	string	散文

   其中，输出变量的“变量描述”字段为大模型理解的内容，需仔细填写。
3. 调测。
   - 在“调测 > 模型”中，选择指令所需的模型，可自定义配置超参数值。
   - 在“调测 > 输入”中，可通过给变量赋值来查看效果。例如，输入topic变量值为秋天、num为100。

4. 单击“生成”可查看生成结果，调测完成后，单击“立即创建”，创建该数据指令。

5 创建合成任务

解释说明

数据合成利用预置或自定义的数据指令对原始数据集进行处理，并根据设定的轮数生成新的数据。

当前，数据合成功能支持合成单轮问答、单轮问答（人设）、问答排序、偏好优化DPO、偏好优化DPO（人设）类型的文本类数据。

操作步骤

1. 在“创建合成任务”页面，选择需要合成的数据集、“合成内容”与“单条数据合成次数”。
2. 如果合成前的数据集与合成后的数据集结构相同，可选择开启“将源数据集整合至合成后数据”，在所有合成轮数运行完成后，将生成的数据与原始数据集合并，单击“下一步”。
3. 进入“合成步骤编排”页面，在左侧“添加指令”页面可选择预置指令或自定义指令。
   - 预置指令。平台为用户提供了多种预置指令如表5-1，便于用户执行合成任务。

表 5-1 预置数据指令清单

指令分类	指令名称	指令描述
生成问题	问题改写为更低难度	该指令可以通过用户输入的问题，使大模型按要求生成一个难度更低、更为简单的问题。
	问题改写为更高难度	该指令通过用户输入的问题，使大模型按要求生成一个难度更高、更为复杂的问题。
	基于提问生成作答要求	该指令根据输入的问题，使大模型泛化一个相应问题的作答要求，该要求与原问题内容不直接相关。该指令可与根据作答要求回答问题的指令进行编排，实现风格多样回答的合成。

指令分类	指令名称	指令描述
	根据样例生成相似问题_few-shot	该指令通过用户输入的多个问题样例，生成一个或多个与样例风格相匹配的新问题。
	根据文本生成问题	根据用户输入的上下文，生一个问题。可用于文本生成QA对的合成编排
	问题改写	改写问题，生成更复杂的问题，可用于指令泛化
生成回答	回答改写	根据用户指定人设，改写回答的风格，不改变回答内容。可与人设泛化指令编排，实现问答对泛化
	根据文本生成回答_遵循要求	根据用户指定的指令要求和问题，根据输入的上下文生成相应回答。可与指令泛化进行编排，实现事实遵循类问答对泛化
	问题生成回答	根据提问，生成回答
	根据文本生成回答_扮演指定人设	根据用户指定的人设和问题，根据输入的上下文生成相应回答。可与人设泛化指令编排，实现事实遵循类问答对泛化。
	问题生成回答_扮演指定人设	根据用户指定的人设和问题，生成相应回答。可与人设泛化指令编排，实现问答对泛化。
	根据文本生成回答	根据用户输入的上下文和问题，生成相应回答。可实现事实遵从的合成
生成问答对	文本生成问答对_判断题	该指令能够从用户提供的参考文本中构建出一个判断题，同时给出其正确回答。
	文本生成问答对_填空题	该指令能够从用户提供的参考文本中构建出一个填空题，同时给出其正确回答。
	文本生成问答对_问答题	该指令能够从用户提供的参考文本中构建出一个问题，同时给出其相应回答。
	根据文本抽取问答对_金融场景	根据用户输入的金融类文档进行问答对的抽取。
生成人设	根据问题生成人设	根据用户输入的问题生成一个人物设定。
其他	BadCase问题泛化	该指令通过用户提供的badcase问题和回答，利用大模型生成在类似情景下可能犯错的攻击性问题。用户可指定生成的攻击性问题个数，个数不超过10。
	根据答案推导解题思路	指令通过用户输入的问题和回答，利用大模型生成包含相应解题思路的回答。

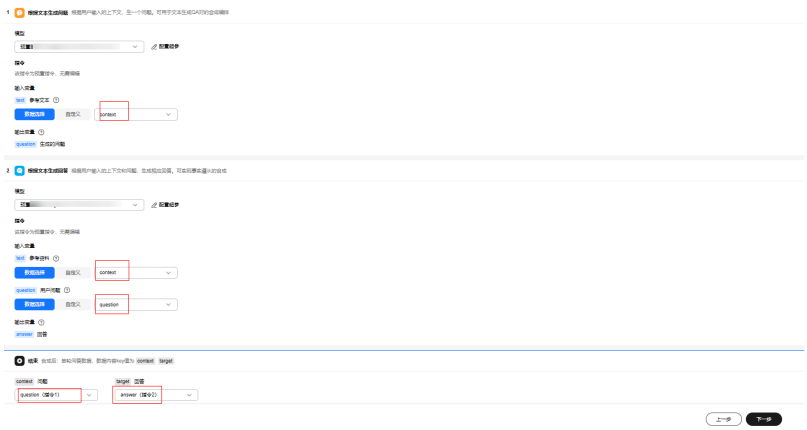
指令分类	指令名称	指令描述
	指令泛化	根据用户指定风格，进行指令泛化。可与指定要求类的问答对生成相关指令编排，实现问答对泛化

- 自定义指令。平台支持编排用户自定义指令，可在“数据工程 > 数据加工 > 合成任务”中单击“管理合成指令”进行创建。

4. 指令选择完成后，配置指令参数。

   如图5-1，展示了预训练文本类数据集的合成指令参数配置示例，该合成任务实现利用预训练文本生成问答对。

   图 5-1 预训练文本类数据集合成指令参数配置示例

5. 指令编排完成后，单击右上角“启用调测”，可以对当前编排的指令效果进行预览。
6. 指令调测完成后，单击“下一步”，填写如下信息。
   - 自动生成加工数据集：启用后任务运行成功自动生成加工数据集，可用于下游数据集发布；如关闭需在加工任务列表操作生成。
   - 填写数据集和描述。
   - 填写扩展信息（可选）：包括行业、语言和自定义信息。
7. 单击“创建并启动”，平台将启动合成任务。

6 创建标注任务

解释说明

数据标注是数据工程中的关键步骤，旨在为无标签的数据集添加准确的标签，从而为模型训练提供有效的监督信号。

标注数据的质量直接影响模型的训练效果和精度，因此高效、准确的标注过程至关重要。

数据标注功能支持创建标注任务、标注数据集（标注作业）、审核标注后的数据集（审核作业）与管理标注任务（任务管理）。其中，不同角色权限支持的功能及展示的前端界面略有差异，详见**表6-1**。

表 6-1 不同角色支持的数据标注任务权限清单

角色名称	创建标注任务	标注作业任务	审核作业任务	任务管理任务
超级管理员	√	√	-	√
管理员	√	√	-	√
标注管理员	√	√	-	√
标注作业员	-	√	-	-
标注审核员	-	-	√	-

当前支持标注的数据集类型为：文本类、视频类、图片、音频类。

创建文本类标注任务

1. 在“创建标注任务”页面，选择需要标注的文本类数据集，并选择“标注项”。选择标注项时，不同类型的数据文件对应的标注项有所差异，可基于页面提示进行选择。

   其中“单轮问答”、“单轮问答（人设）”标注项支持“AI辅助标注”功能，若开启该功能，需要选择已部署的NLP服务作为AI辅助标注模型。

2. 标注要求可设置以下两种方式的“标注要求”：

- 选择“全部标注”：要求标注人员需要对全部的数据进行人工标注后才可提交标注结果。
- 选择“可部分标注”：允许标注人员在确认AI预标注满足要求后，直接使用AI预标注功能完成数据集的标注并提交标注结果。

3. 可选择开启“多人作业”功能，开启后，可选择多人协同完成作业，并增加审核功能可供选择。参考表6-2配置标注分配与审核。

表 6-2 标注分配与审核配置

参数类型	参数名称	参数说明
标注分配	标注员	添加标注人员与数量。
标注审核	是否审核	• 否，标注后不进行审核操作。 • 是，审核员会检查标注员的标注内容，若发现问题，审核员可注明原因并驳回标注数据，标注员需重新标注。
	审核员	添加审核人员与数量。
	审核要求	• 全部审核：要求审核员对全部数据，逐条进行人工审核，才能完成审核任务。 • 可部分审核：审核员在审核一部分数据后，发现标注质量均很高，则可以一键提交剩余待审核数据，默认审核通过，即可完成审核任务。

4. 配置完成后，单击“完成创建”。

创建视频类标注任务

1. 在“创建标注任务”页面，选择需要标注的视频类数据集，并选择“标注项”。选择标注项时，不同类型的数据文件对应的标注项有所差异，可基于页面提示进行选择。
2. 可选择开启“多人作业”功能，开启后，可选择多人协同完成作业，并增加审核功能可供选择。参考表6-3配置标注分配与审核。

表 6-3 标注分配与审核配置

参数类型	参数名称	参数说明
标注分配	标注员	添加标注人员与数量。
标注审核	是否审核	• 否，标注后不进行审核操作。 • 是，审核员会检查标注员的标注内容，若发现问题，审核员可注明原因并驳回标注数据，标注员需重新标注。
	审核员	添加审核人员与数量。

参数类型	参数名称	参数说明
	审核要求	• 全部审核：要求审核员对全部数据，逐条进行人工审核，才能完成审核任务。 • 可部分审核：审核员在审核一部分数据后，发现标注质量均很高，则可以一键提交剩余待审核数据，默认审核通过，即可完成审核任务。

3. 配置完成后，单击“完成创建”。

创建图片类标注任务

1. 在“创建标注任务”页面，选择需要标注的图片类数据集，并选择“标注项”。选择标注项时，不同类型的数据文件对应的标注项有所差异，可基于页面提示进行选择。

   如果选择“图片Caption”标注项，可开启“AI预标注”功能。AI预标注将自动生成标注内容，不会覆盖原始数据集，供标注人员参考，以提高标注效率。
2. 可选择开启“多人作业”功能，开启后，可选择多人协同完成作业，并增加审核功能可供选择。参考**表6-4**配置标注分配与审核。

   表 6-4 标注分配与审核配置

参数类型	参数名称	参数说明
标注分配	标注员	添加标注人员与数量。
标注审核	是否审核	• 否，标注后不进行审核操作。 • 是，审核员会检查标注员的标注内容，若发现问题，审核员可注明原因并驳回标注数据，标注员需重新标注。
	审核员	添加审核人员与数量。
	审核要求	• 全部审核：要求审核员对全部数据，逐条进行人工审核，才能完成审核任务。 • 可部分审核：审核员在审核一部分数据后，发现标注质量均很高，则可以一键提交剩余待审核数据，默认审核通过，即可完成审核任务。

3. 配置完成后，单击“完成创建”。

创建音频类标注任务

1. 在“创建标注任务”页面，选择需要标注的音频类数据集，并选择“标注项”。选择标注项时，不同类型的数据文件对应的标注项有所差异，可基于页面提示进行选择。
2. 可选择开启“多人作业”功能，开启后，可选择多人协同完成作业，并增加审核功能可供选择。参考**表6-5**配置标注分配与审核。

表 6-5 标注分配与审核配置

参数类型	参数名称	参数说明
标注分配	标注员	添加标注人员与数量。
标注审核	是否审核	● 否，标注后不进行审核操作。 ● 是，审核员会检查标注员的标注内容，若发现问题，审核员可注明原因并驳回标注数据，标注员需重新标注。
	审核员	添加审核人员与数量。
	审核要求	● 全部审核：要求审核员对全部数据，逐条进行人工审核，才能完成审核任务。 ● 可部分审核：审核员在审核一部分数据后，发现标注质量均很高，则可以一键提交剩余待审核数据，默认审核通过，即可完成审核任务。

3. 配置完成后，单击“完成创建”。

7 标注数据集

解释说明

数据标注是数据工程中的关键步骤，旨在为无标签的数据集添加准确的标签，从而为模型训练提供有效的监督信号。

标注数据的质量直接影响模型的训练效果和精度，因此高效、准确的标注过程至关重要。

数据标注功能支持创建标注任务、标注数据集（标注作业）、审核标注后的数据集（审核作业）与管理标注任务（任务管理）。其中，不同角色权限支持的功能及展示的前端界面略有差异，详见表7-1。

表 7-1 不同角色支持的数据标注任务权限清单

角色名称	创建标注任务	标注作业任务	审核作业任务	任务管理任务
超级管理员	√	√	-	√
管理员	√	√	-	√
标注管理员	√	√	-	√
标注作业员	-	√	-	-
标注审核员	-	-	√	-

当前支持标注的数据集类型为：文本类、视频类、图片类、音频类。

标注文本类数据集

1. 在“标注任务”页面，单击当前标注任务的“作业”，可执行标注作业任务。
   其中，对于“标注作业员”角色，可单击“标注”执行标注作业任务。

2. 进入标注页面后，逐一对数据进行标注。

   以标注单轮问答数据为例，需要逐条确认文本类数据的问题（Q）与答案（A）是否正确。如果Q或A不正确，可对其进行编辑。

3. 一条数据标注完成后，单击“提交”继续标注剩余数据。
   - 页面左侧可查看已标注的数量与进度。
   - 如果需将标注任务移交给其他人员，可返回至“标注任务”页面，单击操作列“移交”，选择移交人员以及移交数量。
4. 所有数据标注完成后，页面会出现标注任务成功的提示。

标注视频/音频类数据集

1. 在“标注任务”页面，单击当前标注任务的“作业”，可执行标注作业任务。

   其中，对于“标注作业员”角色，可单击“标注”执行标注作业任务。

2. 进入标注页面后，逐一对数据进行标注。
3. 一条数据标注完成后，单击“提交”继续标注剩余数据。
   - 页面左侧可查看已标注的数量与进度。
   - 如果需将标注任务移交给其他人员，可返回至“标注任务”页面，单击操作列“移交”，选择移交人员以及移交数量。
4. 所有数据标注完成后，页面会出现标注任务成功的提示。

标注图片类数据集

1. 在“标注任务”页面，单击当前标注任务的“作业”，可执行标注作业任务。

   其中，对于“标注作业员”角色，可单击“标注”执行标注作业任务。

2. 进入标注页面后，逐一对数据进行标注。
3. 一条数据标注完成后，单击“提交”继续标注剩余数据。
   - 页面左侧可查看已标注的数量与进度。
   - 如果需将标注任务移交给其他人员，可返回至“标注任务”页面，单击操作列“移交”，选择移交人员以及移交数量。
   - 如果在创建标注任务时开启了“AI预标注”，且将标注要求设置为“可部分标注”，则可在标注部分数据后，单击右上角“提交全部标注数据”，可自动标注剩余数据。
4. 所有数据标注完成后，页面会出现标注任务成功的提示。

8 审核数据集

解释说明

创建数据集标注任务时，如果启用了标注审核，在完成标注后可以审核标注结果。

对于审核不合格的数据可以填写不合格原因并驳回给标注员重新标注。

该操作需要具备**标注审核员**角色权限。

操作步骤

1. 在审核页面，可通过单击“通过”或“不通过”逐一对数据进行审核，直至所有数据审核完成。
2. 可开启“标注前后对比”功能(仅用于文本类数据)，查看当前数据标注前后的内容。
3. 如果在创建标注任务时开启了“可部分审核”功能，则在审核过程中可单击右上角“提交全部审核数据”，系统将自动审核剩余的待审核数据，默认为审核通过。
4. 如果需将审核任务移交给其他人员，可返回至“标注任务”页面，单击操作列“移交”设置移交人员以及移交的数量。

9 创建评估标准

解释说明

ModelArts Studio大模型开发平台预设了一套基础评估标准，涵盖了不同数据多个评估维度，用户可以直接使用该标准或在该标准的基础上创建评估标准。

当前支持评估的数据集类型为：文本类、视频类、图片类。

操作步骤

1. 在“创建评估标准”页面。选择需要参考的预置标准或自定义标准，并填写“评估标准名称”和“描述”，也可选择不参考。
2. 编辑评估项。

   用户可以基于实际需求删减评估项，或创建自定义评估项。创建自定义评估项时，需要将评估类别、评估项、评估项说明填写清晰，填写时确保描述无歧义。
3. 单击“完成创建”创建评估标准。评估标准创建完成后可以在“评估标准”页面查看创建的评估标准，并支持编辑、删除操作。

10 创建评估任务

解释说明

数据评估通过对数据集进行系统的质量检查，依据评估标准评估数据的多个维度，旨在发现潜在问题并加以解决。

当前支持评估的数据集类型为：文本类、视频类、图片类。

操作步骤

1. 在“创建评估任务”页面。选择需要进行评估的数据集，并设置抽样规格，即从数据集中抽取一定比例数据进行评估。
2. 单击“下一步”选择数据集评估标准。平台支持选择用户自定义的评估标准。
3. 评估标准选择完成后，单击“下一步”，设置评估人员。
4. 单击“下一步”填写任务名称。单击“完成创建”。

11 评估数据集

解释说明

数据评估通过对数据集进行系统的质量检查，依据评估标准评估数据的多个维度，旨在发现潜在问题并加以解决。

当前支持评估的数据集类型为：文本类、视频类、图片类。

操作步骤

1. 在评估页面，单击右上角“评估标准”查看当前数据的评估标准。
2. 单击“通过”或“不通过”逐一对数据进行评估。
   对于文本类数据集而言，可选中文本并右键标注数据问题。
3. 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。
4. 评估完成后可以回退至“数据评估 > 人工评估”页面，单击操作列“报告”，查看数据集质量评估报告。

12 创建配比任务

解释说明

数据配比是将多个数据集按照特定比例关系组合的过程，确保数据的多样性、平衡性和代表性。

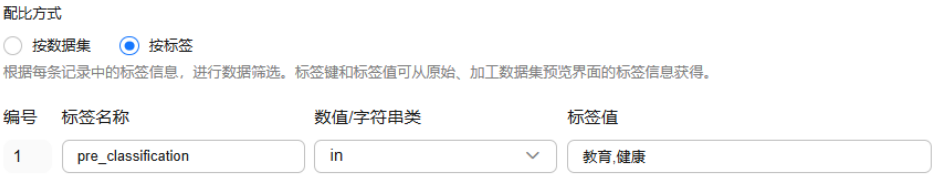
当前支持配比的数据集类型为：文本类、图片类，视频类，多模态类，预测类。

操作步骤

1. 在“创建配比任务”页面。选择数据集模态，并选择多个需要配比的数据集，单击“下一步”。
2. 在“数据配比”页面，对于文本类数据集支持两种配比方式，“按数据集”和“按标签”。
   - 按数据集：可以设置不同数据集的配比数量，单击“确定”。
   - 按标签：该场景适用于通过数据打标类加工算子进行加工的文本类数据集，具体标签名称与标签值可在完成数据集加工操作后，进入数据集详情页面获取。

     填写示例如图12-1所示。

     图 12-1 “按标签”配比方式填写示例

3. 单击“确定”创建配比任务。

13 创建发布任务

解释说明

数据发布是将数据集发布为特定格式的“发布数据集”的过程，用于后续模型训练等操作。

当前支持发布的数据集类型为：文本类、图片类、视频类、音频类、气象类、预测类、多模态类和其他类。

针对文本类、图片类数据集，平台支持将数据集发布为标准格式或盘古格式。如果该数据集用于训练盘古大模型，请确保已发布为“盘古格式”。

操作步骤

1. 在“创建发布任务”页面，选择数据集模态与数据集，单击“下一步”。
2. 在“基本配置”中选择数据用途、数据集可见性、适用场景。

   如果当前数据集拟用于训练盘古大模型，请选择适用场景为“STUDIO模型训练 > 盘古格式”。
3. 输入数据集名称，可填写数据集描述与扩展信息，单击“下一步”，进入任务配置。
4. 在任务配置页面可以进行资源配置。

   下拉资源配置，可以设置任务资源 。也支持自定义参数配置，单击添加参数，输入参数名称和参数值。
5. 任务配置完毕后，单击“确定”，发布数据集。

14 创建训练任务

解释说明

平台当前支持NLP大模型、多模态大模型、CV大模型、预测大模型、科学计算大模型的训练任务。

创建 NLP 大模型全量微调任务

1. 在“创建训练任务”页面，参考表14-1完成训练参数设置，参数默认值在创建训练任务的时候会带出。

表 14-1 NLP 大模型全量微调参数说明

参数分类	训练参数	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“大语言模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“微调”。
	训练目标	选择“全量微调”。 ● 全量微调：在模型进行有监督微调时，对大模型的所有参数进行更新。这种方法通常能够实现最佳的模型性能，但需要消耗大量计算资源和时间，计算开销较大。
	高级设置	checkpoints：在模型训练过程中，用于保存模型权重和状态的机制。 ● 关闭：关闭后不保存checkpoints，无法基于checkpoints执行续训操作。 ● 自动：自动保存训练过程中的所有checkpoints。 ● 自定义：根据设置保存指定数量的checkpoints。

参数分类	训练参数	参数说明
训练参数	热身比例	热身比例是指在模型训练初期逐渐增加学习率的过程。 由于训练初期模型的权重通常是随机初始化的，预测能力较弱，若直接使用较大的学习率，可能导致更新过快，进而影响收敛。为解决这一问题，通常在训练初期使用较小的学习率，并逐步增加，直到达到预设的最大学习率。通过这种方式，热身比例能够避免初期更新过快，从而帮助模型更好地收敛。
	序列长度	sequence_length，训练单条数据的最大长度，超过该长度的数据在训练时将被截断。
	数据批量大小	数据批量是指训练过程中将数据集分成小批次进行读取，并设定每个批次的数据大小。 通常，较大的批量能够使梯度更加稳定，有助于模型的收敛。然而，较大的批量也会占用更多显存，可能导致显存不足，并延长每次训练时间。
	单步迭代时处理的数据批量大小	指定每次迭代时处理的数据批量大小。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。
	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。
	学习率衰减比率	用于控制训练过程中学习率下降的幅度。 计算公式为：最低学习率 = 初始学习率 × 学习率衰减比率。
	Agent微调	在训练Agent所需的NLP大模型时，可以开启此参数。通过调整训练数据中的Prompt，引导模型在特定领域或任务上生成更符合预期的回答。 在使用此参数前，请先联系盘古客服，调整Prompt和训练数据。

参数分类	训练参数	参数说明
	模型保存步数	每训练一定数量的步骤（或批次），模型的状态将会被保存。可以通过以下公式预估已训练的数据量： token_num = step * batch_size * sequence ● token_num：已训练的数据量（以Token为单位）。 ● step：已完成的训练步数。 ● batch_size：每个训练步骤中使用的样本数量。 ● sequence：每个数据样本中的Token数量。
	模型保存策略	save_checkpoint_steps/save_checkpoint_epoch，训练过程中是按迭代步数，还是训练轮数保存Checkpoint文件。
	模型保存间隔	save_checkpoint_steps，训练过程中每隔多少个训练步长保存一次模型Checkpoint文件。 save_checkpoint_epoch，训练过程中每个多少训练轮数保存一次模型Checkpoint文件。
	模型保存个数	checkpoints：在模型训练过程中，用于保存模型权重和状态的机制。 ● 关闭：关闭后不保存checkpoints，无法基于checkpoints执行续训操作。 ● 自动：自动保存训练过程中的所有checkpoints。 ● 自定义：根据设置保存指定数量的checkpoints。
	旋转位置编码	rotary_base，位置编码的基底值，增强模型对序列中位置信息的捕捉能力，数值越大，模型能够处理的序列长度更长，泛化能力更好，建议使用默认值。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂。
	优化器	优化器参数用于更新模型的权重，常见包括adamw。 ● adamw是一种改进的Adam优化器，增加了权重衰减机制，有效防止过拟合。
训练数据配置	训练集	选择训练模型所需的数据集。
	验证集	● 若选择“分割训练集”，则需进一步配置数据拆分比例。 ● 若选择“选择数据集”，则需选择导入的数据集。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。

参数分类	训练参数	参数说明
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建 NLP 大模型 LoRA 微调任务

1. 在“创建训练任务”页面，参考**表14-2**完成训练参数设置。

表 14-2 NLP 大模型 LoRA 微调参数说明

参数分类	训练参数	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“大语言模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“微调”。
	训练目标	选择“LoRA微调”。 ● LoRA微调：在模型微调过程中，只对特定的层或模块的参数进行更新，而其余参数保持冻结状态。这种方法可以显著减少计算资源和时间消耗，同时在很多情况下，依然能够保持较好的模型性能。

参数分类	训练参数	参数说明
训练参数	热身比例	热身比例是指在模型训练初期逐渐增加学习率的过程。 由于训练初期模型的权重通常是随机初始化的，预测能力较弱，若直接使用较大的学习率，可能导致更新过快，进而影响收敛。为解决这一问题，通常在训练初期使用较小的学习率，并逐步增加，直到达到预设的最大学习率。通过这种方式，热身比例能够避免初期更新过快，从而帮助模型更好地收敛。
	数据批量大小	数据批量是指训练过程中将数据集分成小批次进行读取，并设定每个批次的数据大小。 通常，较大的批量能够使梯度更加稳定，有助于模型的收敛。然而，较大的批量也会占用更多显存，可能导致显存不足，并延长每次训练时间。
	学习率衰减比率	用于控制训练过程中学习率下降的幅度。 计算公式为：最低学习率 = 初始学习率 × 学习率衰减比率。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。
	序列长度	sequence_length，训练单条数据的最大长度，超过该长度的数据在训练时将被截断。
	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。
	LoRA矩阵中的秩	lora_rank，在Lora矩阵中，Rank的值用于衡量矩阵的复杂度和信息量。数值较大，增强模型的表示能力，但会增加训练时长；数值越小可以减少参数数量，降低过拟合风险。
	Agent微调	在训练Agent所需的NLP大模型时，可以开启此参数。通过调整训练数据中的Prompt，引导模型在特定领域或任务上生成更符合预期的回答。 在使用此参数前，请先联系盘古客服，调整Prompt和训练数据。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂。
	优化器	优化器参数用于更新模型的权重，常见包括adamw。 ● adamw是一种改进的Adam优化器，增加了权重衰减机制，有效防止过拟合。

参数分类	训练参数	参数说明
	旋转位置编码	位置编码的基底值， 建议使用默认值。
训练数据配置	训练集	选择训练模型所需的数据集。
	验证集	• 若选择“分割训练集”，则需进一步配置数据拆分比例。 • 若选择“选择数据集”，则需选择导入的数据集。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建 NLP 大模型预训练任务

1. 在“创建训练任务”页面，参考**表14-3**完成训练参数设置。

表 14-3 NLP 大模型预训练参数说明

参数分类	训练参数	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“大语言模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“预训练”。
	高级设置checkpoints	checkpoints：在模型训练过程中，用于保存模型权重和状态的机制。 ● 关闭：关闭后不保存checkpoints，无法基于checkpoints执行续训操作。 ● 自动：自动保存训练过程中的所有checkpoints。 ● 自定义：根据设置保存指定数量的checkpoints。
训练参数	热身比例	热身比例是指在模型训练初期逐渐增加学习率的过程。 由于训练初期模型的权重通常是随机初始化的，预测能力较弱，若直接使用较大的学习率，可能导致更新过快，进而影响收敛。为解决这一问题，通常在训练初期使用较小的学习率，并逐步增加，直到达到预设的最大学习率。通过这种方式，热身比例能够避免初期更新过快，从而帮助模型更好地收敛。
	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。
	数据批量大小	数据集进行分批读取训练，设定每个批次数据的大小。 通常情况下，较大的数据批量可以使梯度更加稳定，从而有利于模型的收敛。然而，较大的数据批量也会占用更多的显存资源，这可能导致显存不足，并且会延长每次训练的时长。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。
	序列长度	sequence_length，训练单条数据的最大长度，超过该长度的数据在训练时将被截断。
	热身比例	热身比例是指在模型训练初期逐渐增加学习率的过程。 由于训练初期模型的权重通常是随机初始化的，预测能力较弱，若直接使用较大的学习率，可能导致更新过快，进而影响收敛。为解决这一问题，通常在训练初期使用较小的学习率，并逐步增加，直到达到预设的最大学习率。通过这种方式，热身比例能够避免初期更新过快，从而帮助模型更好地收敛。

参数分类	训练参数	参数说明
	学习率衰减比率	用于控制训练过程中学习率下降的幅度。 计算公式为：最低学习率 = 初始学习率 × 学习率衰减比率。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂。
	优化器	优化器参数用于更新模型的权重，常见包括adamw。 ● adamw是一种改进的Adam优化器，增加了权重衰减机制，有效防止过拟合。
	模型保存步数	每训练一定数量的步骤（或批次），模型的状态将会被保存。可以通过以下公式预估已训练的数据量： token_num = step * batch_size * sequence ● token_num：已训练的数据量（以Token为单位）。 ● step：已完成的训练步数。 ● batch_size：每个训练步骤中使用的样本数量。 ● sequence：每个数据样本中的Token数量。
	数据预处理并发个数	定义了在预处理数据时，能够同时处理文件的并行进程数量。设定这个参数的主要目的是通过并发处理来加速数据预处理，从而提升训练效率。
	旋转位置编码	rotary_base，位置编码的基底值，增强模型对序列中位置信息的捕捉能力，数值越大，模型能够处理的序列长度更长，泛化能力更好，建议使用默认值。
训练数据配置	训练集	选择训练模型所需的数据集。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。

参数分类	训练参数	参数说明
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建 NLP 大模型 DPO 强化学习任务

1. 在“创建训练任务”页面，参考**表14-4**完成训练参数设置。

表 14-4 NLP 大模型 DPO 强化学习参数说明

参数分类	训练参数	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“我的资产”。 ● 类型：选择“大语言模型”，并选择训练所用的模型。
	训练类型	选择“强化学习”。
	训练目标	选择“DPO”。
	高级设置checkpoints	checkpoints：在模型训练过程中，用于保存模型权重和状态的机制。 ● 关闭：关闭后不保存checkpoints，无法基于checkpoints执行续训操作。 ● 自动：自动保存训练过程中的所有checkpoints。 ● 自定义：根据设置保存指定数量的checkpoints。
训练参数	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。
	数据批量大小	数据集进行分批读取训练，设定每个批次数据的大小。 通常情况下，较大的数据批量可以使梯度更加稳定，从而有利于模型的收敛。然而，较大的数据批量也会占用更多的显存资源，这可能导致显存不足，并且会延长每次训练的时长。

参数分类	训练参数	参数说明
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。
	序列长度	sequence_length，训练单条数据的最大长度，超过该长度的数据在训练时将被截断。
	热身比例	热身比例是指在模型训练初期逐渐增加学习率的过程。 由于训练初期模型的权重通常是随机初始化的，预测能力较弱，若直接使用较大的学习率，可能导致更新过快，进而影响收敛。为解决这一问题，通常在训练初期使用较小的学习率，并逐步增加，直到达到预设的最大学习率。通过这种方式，热身比例能够避免初期更新过快，从而帮助模型更好地收敛。
	权重衰减系数	weight_decay，是一种对模型参数值大小进行衰减的正则化方法，防止模型过拟合，提高模型泛化能力。
	模型保存策略	save_checkpoint_steps，训练过程中是按迭代步数，还是训练轮数保存Checkpoint文件。
	Checkpoint保存间隔	save_checkpoint_steps，训练过程中每隔多少个训练步长保存一次模型Checkpoint文件。
	Checkpoint保存轮数	save_checkpoint_epoch，训练过程中每隔多少训练轮数保存一次模型Checkpoint文件，当值为0以save_checkpoint_steps为准，当值大于0以save_checkpoint_epoch为准。
	验证步数	eval_steps，模型每隔多少步跑一次验证集。
	旋转位置编码	rotary_base，位置编码的基底值，增强模型对序列中位置信息的捕捉能力，数值越大，模型能够处理的序列长度更长，泛化能力更好，建议使用默认值。
	DPO loss温度超参	Beta，用于控制模型输出分布的集中程度。较高的beta值会使输出更具确定性，而较低的beta值则使输出更具多样性。
	学习率衰减比率	用于控制训练过程中学习率下降的幅度。 计算公式为：最低学习率 = 初始学习率 × 学习率衰减比率。

参数分类	训练参数	参数说明
	模型保存步数	每训练一定数量的步骤（或批次），模型的状态将会被保存。可以通过以下公式预估已训练的数据量： token_num = step * batch_size * sequence ● token_num：已训练的数据量（以Token为单位）。 ● step：已完成的训练步数。 ● batch_size：每个训练步骤中使用的样本数量。 ● sequence：每个数据样本中的Token数量。
训练数据配置	训练集	选择训练模型所需的数据集。
	验证集	● 若选择“分割训练集”，则需进一步配置数据拆分比例。 ● 若选择“选择数据集”，则需选择导入的数据集。
资源配置	计费模式	选择训练RFT强化任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建 NLP 大模型 RFT 强化学习任务

1. 在“创建训练任务”页面，参考**表14-5**完成训练参数设置。

表 14-5 NLP 大模型 RFT 强化学习参数说明

参数分类	训练参数	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“大语言模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“强化学习”。
	训练目标	选择“RFT”。
训练参数	热身比例	热身比例是指在模型训练初期逐渐增加学习率的过程。 由于训练初期模型的权重通常是随机初始化的，预测能力较弱，若直接使用较大的学习率，可能导致更新过快，进而影响收敛。为解决这一问题，通常在训练初期使用较小的学习率，并逐步增加，直到达到预设的最大学习率。通过这种方式，热身比例能够避免初期更新过快，从而帮助模型更好地收敛。
	数据批量大小	指定每个数据并行下处理的数据批量大小。在数据并行和流水线并行开启情况下，全局batch_size等于per_batch_size乘micro_size乘data_parallelism。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。
	训练轮数	完成全部训练数据集训练的次数。
	优化器	训练中对模型参数进行更新的算法，默认为adamw。
	权重衰减系数	权重衰减系数，通过在损失函数中增加一个与模型权重大小相关的惩罚项，来鼓励模型保持权重较小，从而防止模型过于复杂或过拟合训练数据。
	DPO loss温度超参	用于控制模型输出分布的集中程度。较高的beta值会使输出更具确定性，而较低的beta值则使输出更具多样性。
	打分器类型	rft训练中的样本打分器。
训练数据配置	训练集	选择训练模型所需的数据集。
资源配置	计费模式	选择训练RFT强化任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。

参数分类	训练参数	参数说明
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建 CV 大模型微调任务

1. 在“创建训练任务”页面，参考**表14-6**完成训练参数设置。

   其中，“训练参数”展示了各场景涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

表 14-6 CV 大模型微调参数说明

参数分类	训练参数	说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“CV大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“微调”。
	训练目标	选择“全量微调”。 ● 全量微调：在模型进行有监督微调时，对大模型的所有参数进行更新。这种方法通常能够实现最佳的模型性能，但需要消耗大量计算资源和时间，计算开销较大。
训练参数	训练参数	模型训练参数，参考**表14-7**。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。

参数分类	训练参数	说明
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

表 14-7 CV 大模型训练参数说明

模型类型	训练参数	说明	典配值
Pangu-CV-物体检测-S-2.1.0	数据集	训练数据集。	-
	训练轮次	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。	100
	批次大小	表示每个训练步骤中使用的样本数量。较大的批量可以提供更稳定的梯度，但可能会增加计算资源的使用和训练时间。	8
	验证集图片保存最大数量	用于限制验证过程中模型预测错误（与标注不符）的图片保存数量，便于后续分析错误类型，同时避免存储空间浪费。	10000
	学习率	学习率决定了每次训练时模型参数更新的幅度。	0.01
	权重衰减	通过在每次参数更新时对模型权重进行小幅度缩减，避免模型过度依赖特定数据。	0.0005
Pangu-CV-物体检测-N-2.1.0	数据集	训练数据集。	-
	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。	3

模型类型	训练参数	说明	典配值
	热身轮次	表示在模型训练初期，逐步增加学习率到预设值的训练轮次，用于帮助模型在训练初期稳定收敛，避免大幅度的参数更新导致不稳定的学习过程。	3
	热身阶段学习率	热身轮次中使用的初始学习率。	0.0001
	权重衰减	用于防止模型过拟合。在更新模型权重时，它会对模型参数施加惩罚，使得权重值趋于较小，从而提高模型的泛化性能。	0.0005
	优化器	选择用于训练模型的优化算法。这里选择的“sgd”是随机梯度下降法，它是深度学习中常用的优化算法之一，适合大规模数据集训练。	sgd
	锚框的长边和短边的比例	定义检测物体锚框的长宽比。通过设置不同的长短比例，模型可以更好地适应多种尺寸和形状的物体。	[0.5, 1.0, 2.0]
	锚框大小	指锚框的初始尺寸。锚框是物体检测中的一个关键概念，通过合理设置，可以帮助模型检测出多种尺寸的目标。	8
	框重叠比例阈值	用于判定模型预测的边界框与真实边界框之间是否为同一物体。该阈值用于计算IoU（交并比），影响模型的精确度。	0.5
	滑动平滑训练	一种训练策略，通过在模型预测的标签上添加少量噪声来避免过拟合，常用于提升模型在测试数据集上的泛化能力。	True
	极大值抑制阈值	在预测多个边界框时，用于去除高度重叠的边界框。此阈值控制相似的边界框保留的条件。	0.4
	类别无关极大值抑制开关	决定是否在不同类别中应用极大值抑制阈值。	False
Pangu-CV-图像分类-2.1.0	数据集	训练数据集。	-
	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。	20

模型类型	训练参数	说明	典配值
	每卡批次大小	表示每个训练步骤中使用的样本数量。较大的批量可以提供更稳定的梯度，但可能会增加计算资源的使用和训练时间。	64
	分类模式	图像分类模式，支持单标签分类和多标签分类。	single
	超参搜索	是否启用超参搜索获取最优的超参用于训练。	False
	初始学习率	学习率决定了每次训练时模型参数更新的幅度。初始学习率是模型训练最开始阶段所设定的学习率，它是学习率的初始值。	0.0005
	超参搜索训练轮数	在超参搜索阶段，设置模型会经过几轮训练以找到最优的超参数组合。	5

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建 CV 大模型预训练任务

1. 在“创建训练任务”页面，参考**表14-8**完成训练参数设置。

   其中，“训练参数”展示了各场景涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

表 14-8 CV 大模型预训练参数说明

参数分类	训练参数	说明
训练配置	选择模型	可以修改如下信息： • 来源：选择“模型广场”。 • 类型：选择“CV大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“预训练”。
训练参数	训练参数	模型训练参数，参考**表14-9**。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。

参数分类	训练参数	说明
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

表 14-9 CV 大模型训练参数说明

模型类型	训练参数	说明	典配值
Pangu-CV-物体检测-S-2.1.0	数据集	训练数据集。	-
	训练轮次	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。	100
	批次大小	表示每个训练步骤中使用的样本数量。较大的批量可以提供更稳定的梯度，但可能会增加计算资源的使用和训练时间。	8
	验证集图片保存最大数量	用于限制验证过程中模型预测错误（与标注不符）的图片保存数量，便于后续分析错误类型，同时避免存储空间浪费。	10000
	学习率	学习率决定了每次训练时模型参数更新的幅度。	0.01
	权重衰减	通过在每次参数更新时对模型权重进行小幅度缩减，避免模型过度依赖特定数据。	0.0005
Pangu-CV-物体检测-N-2.1.0	数据集	训练数据集。	-
	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。	3
	热身轮次	表示在模型训练初期，逐步增加学习率到预设值的训练轮次，用于帮助模型在训练初期稳定收敛，避免大幅度的参数更新导致不稳定的学习过程。	3
	热身阶段学习率	热身轮次中使用的初始学习率。	0.0001
	权重衰减	用于防止模型过拟合。在更新模型权重时，它会对模型参数施加惩罚，使得权重值趋于较小，从而提高模型的泛化性能。	0.0005

模型类型	训练参数	说明	典配值
	优化器	选择用于训练模型的优化算法。这里选择的“sgd”是随机梯度下降法，它是深度学习中常用的优化算法之一，适合大规模数据集训练。	sgd
	锚框的长边和短边的比例	定义检测物体锚框的长宽比。通过设置不同的长短比例，模型可以更好地适应多种尺寸和形状的物体。	[0.5, 1.0, 2.0]
	锚框大小	指锚框的初始尺寸。锚框是物体检测中的一个关键概念，通过合理设置，可以帮助模型检测出多种尺寸的目标。	8
	框重叠比例阈值	用于判定模型预测的边界框与真实边界框之间是否为同一物体。该阈值用于计算IoU（交并比），影响模型的精确度。	0.5
	滑动平滑训练	一种训练策略，通过在模型预测的标签上添加少量噪声来避免过拟合，常用于提升模型在测试数据集上的泛化能力。	True
	极大值抑制阈值	在预测多个边界框时，用于去除高度重叠的边界框。此阈值控制相似的边界框保留的条件。	0.4
	类别无关极大值抑制开关	决定是否在不同类别中应用极大值抑制阈值。	False
Pangu-CV-图像分类-2.1.0	数据集	训练数据集。	-
	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。	20
	每卡批次大小	表示每个训练步骤中使用的样本数量。较大的批量可以提供更稳定的梯度，但可能会增加计算资源的使用和训练时间。	64
	分类模式	图像分类模式，支持单标签分类和多标签分类。	single
	超参搜索	是否启用超参搜索获取最优的超参用于训练。	False
	初始学习率	学习率决定了每次训练时模型参数更新的幅度。初始学习率是模型训练最开始阶段所设定的学习率，它是学习率的初始值。	0.0005
	超参搜索训练轮数	在超参搜索阶段，设置模型会经过几轮训练以找到最优的超参数组合。	5

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建预测大模型微调任务

1. 在“创建训练任务”页面，参考**表14-10**完成训练参数设置。

   其中，“训练参数”展示了各场景涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

   表 14-10 预测大模型微调参数说明

参数分类	训练参数	说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“预测大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“微调”。
	训练目标	选择“全量微调”。 ● 全量微调：在模型进行有监督微调时，对大模型的所有参数进行更新。这种方法通常能够实现最佳的模型性能，但需要消耗大量计算资源和时间，计算开销较大。
训练参数	训练参数	模型训练参数，参考**表14-11**。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

表 14-11 预测大模型训练参数说明

模型类型	训练参数	说明
盘古时序预测分类大模型（Pangu-Predict-Cla-TS-2.1.0）	数据集	选择训练所需的数据集。
	非特征列	不作为输入特征的列。此处填写的特征列名将不作为模型训练的输入特征，不要将预测目标列的列名填入。填写格式为中括号，其中列举带引号的特征列名，默认设置为中括号。
	预测目标列	预测目标的列名，指定预测目标变量列名，填写格式为中括号，时序分类任务为样本标签列。默认设置为中括号，表示选择最后一列作为预测目标变量。 注意：二次微调不支持分类任务的分类数发生改变，需与原资产保持一致，推理服务接口输入字段名称以二次微调为准。
	标识列	用于对连续时间段样本点的区分标识。时序分类任务会将相同ID的行数据视作同一个序列样本，因此时序分类任务必须填写。默认设置为中括号。
	类别特征列	使用LabelEncoder处理的特征列的列表，用于处理字符串类型的类别特征。填写格式为中括号，其中列举带引号的特征列名，默认设置为中括号，表示没有需要使用LabelEncoder处理的特征。
	协变量列	用于显式指定协变量利用时序任务建模。如果不填写，则自动选择协变量列。
	训练集切分比例	训练集切分比例。注意：训练集切分比例、验证集切分比例、测试集切分比例三者之和为1。
	验证集切分比例	验证集切分比例。
	测试集切分比例	测试集切分比例。
	历史窗口大小	时序预测输入的窗口长度。数值越大建模包含的历史信息越多，但相应的显存占用会增加以及模型拟合难度也会提升。请根据实际任务特点选择合适的输入窗口，建议取值范围128至512。
	训练轮数	迭代训练的epoch数量，请根据实际情况选择，不低于1，建议范围5至10。
	数据批量大小	单卡的批处理大小，通常来说数据批量越大，梯度会越稳定。但是同时也会使用更大的显存，受硬件限制可能会OOM，并延长单步训练时长。

模型类型	训练参数	说明
	学习率	学习率用于控制每个训练step参数更新的幅度，一般来说需要选择一个合适的学习率，否则当学习率过大的时候会导致模型难以收敛，过小的时候会收敛速度过慢。
	热身比率	热身阶段占整体训练的比例。刚开始训练时若选择一个较大的学习率，可能带来模型的不稳定，选择使用warmup热身的方式，可以使得开始训练的热身阶段内学习率较小，在热身阶段的小学习率下，模型可以慢慢趋于稳定，等模型相对稳定后再逐渐提升至预设的最大学习率进行训练。使用热身可以使得模型收敛速度更快，效果更佳。
	权重衰减系数	权重衰减系数，权重衰减是一种根据参数值大小对参数进行衰减的正则化方法，用于减少过拟合，权重衰减系数用于控制正则化的力度，权重衰减系数越大，正则化力度越强。
	模型保存步数	定义模型每隔多少步保存一次，注意训练集总步数必须大于该值。
盘古时序预测回归大模型（Pangu-Predict-Reg-TS-2.1.0）	数据集	选择训练所需的数据集。
	非特征列	不作为输入特征的列。此处填写的特征列名将不作为模型训练的输入特征，不要将预测目标列的列名填入。填写格式为中括号，其中列举带引号的特征列名，默认设置为中括号。
	预测目标列	预测目标的列名，指定预测目标变量列名，填写格式为中括号，时序分类任务为样本标签列。默认设置为中括号，表示选择最后一列作为预测目标变量。 注意：二次微调支持列名/变量数发生变化，但回归任务的目标变量数需要与之前保持一致，推理服务接口输入字段名称以二次微调为准。
	标识列	用于对连续时间段样本点的区分标识。时序分类任务会将相同ID的行数据视作同一个序列样本，因此时序分类任务必须填写。默认设置为中括号。
	类别特征列	使用LabelEncoder处理的特征列的列表，用于处理字符串类型的类别特征。填写格式为中括号，其中列举带引号的特征列名，默认设置为中括号，表示没有需要使用LabelEncoder处理的特征。
	协变量列	用于显式指定协变量例用时序任务建模。如果不填写，则自动选择协变量列。
	训练集切分比例	训练集切分比例。注意：训练集切分比例、验证集切分比例、测试集切分比例三者之和为1。

模型类型	训练参数	说明
	验证集切分比例	验证集切分比例。
	测试集切分比例	测试集切分比例。
	特征是否独立建模	时序变量建模模式设置。若选择独立建模，则模型只关注预测目标列的选择的数据，并不做列数据区分，统一视作出自一个序列分布。若选择非独立建模，建模会考虑列变量之间的相关性，建模的变量范围为输入的数据列除去非特征列以及标识列的剩余数据内容。
	历史窗口大小	时序预测输入的窗口长度。数值越大建模包含的历史信息越多，但相应的显存占用会增加以及模型拟合难度也会提升。请根据实际任务特点选择合适的输入窗口，建议取值范围128至512。
	预测目标窗口大小	时序预测输出的窗口长度。数值越大输出预测的时间范围越大，但相应的显存占用会增加以及模型预测精度可能会下降。请根据实际任务特点选择合适的输出窗口，建议取值范围48至96，且尽量选择预测目标周期的整数倍。注若基于已发布模型的进行增量微调，输出窗口需要与其保持一致。 注意：二次微调支持输入窗口可变，但输出窗口的长度需要与之前保持一致。
	训练轮数	迭代训练的epoch数量，请根据实际情况选择，不低于1，建议范围5至10。
	数据批量大小	单卡的批处理大小，通常来说数据批量越大，梯度会越稳定。但是同时也会使用更大的显存，受硬件限制可能会OOM，并延长单步训练时长。
	学习率	学习率用于控制每个训练step参数更新的幅度，一般来说需要选择一个合适的学习率，否则当学习率过大的时候会导致模型难以收敛，过小的时候会收敛速度过慢。
	热身比率	热身阶段占整体训练的比例。刚开始训练时若选择一个较大的学习率，可能带来模型的不稳定，选择使用warmup热身的方式，可以使得开始训练的热身阶段内学习率较小，在热身阶段的小学习率下，模型可以慢慢趋于稳定，等模型相对稳定后再逐渐提升至预设的最大学习率进行训练，使用热身可以使得模型收敛速度更快，效果更佳。

模型类型	训练参数	说明
	权重衰减系数	权重衰减系数。权重衰减是一种根据参数值大小对参数进行衰减的正则化方法，用于减少过拟合，权重衰减系数用于控制正则化的力度，权重衰减系数越大，正则化力度越强。
	模型保存步数	定义模型每隔多少步保存一次，注意训练集总步数必须大于该值。
盘古融合推荐回归大模型（Pangu-Predict-Reg-Table-2.0.0）	数据集选择	选择训练所需的数据集。
	类别特征列	指定使用LabelEncoder处理的字符串类型类别特征的列表。格式为["列名1","列名2"]，默认设置为[]，表示没有需要处理的类别特征。 LabelEncoder的作用是将类别特征转换为数值型特征，使模型能够处理这些特征。
	非特征列	列出不需要输入到模型中的特征列，用于排除冗余或无意义的特征。格式为["列名1","列名2"]，默认设置为[]，表示所有特征都用于训练。
	标准化列	指定需要进行最大最小值标准化处理的数值特征的列表。格式为["列名1","列名2"]，默认设置为[]，表示没有特征需要标准化。标准化将特征值缩放到0到1的范围，处理分布差异较大的数值特征。
	预测目标列	指定预测目标变量的列名，gatednet算法支持多目标变量预测，其余算法仅支持单目标变量预测。格式为["列名"]，默认设置为[]，表示选择最后一列作为预测目标变量。
	训练集&验证集比例	将数据集划分为训练集和验证集，填入验证集比例即可，默认设置为0.2，即训练集占0.8，验证集占0.2，可选范围为0.1、0.2、0.3、0.4。

模型类型	训练参数	说明
	基模型算法池	从预定义的算法池中选择用于训练模型的算法，算法包括：["svm", "ada", "lgb", "xgb", "rf", "et", "gb", "gauss","mlp","gatednet"]，其中： ● svm表示支持向量机。 ● ada表示adaboost。 ● lgb表示lightgbm。 ● xgb表示xgboost。 ● rf表示随机森林。 ● et表示extraTree。 ● gb表示梯度提升树。 ● gauss表示高斯过程，gauss适合维度小于10且数据量小于500的样本数据。 ● mlp表示多层感知机，默认设置为5lgb，多种类算法示例: 3lgb,2rf,1xgb（表示使用3个LightGBM算法、2个随机森林算法和1个XGBoost算法）。 ● gatednet表示门控自适应网络，可用于多目标预测。gatednet不支持跟其他算法同时选择，如果同时指定gatednet和其他模型，将只会执行gatednet算法。gatednet仅支持单个生效，例如，5gatednet也只会使用一个gatednet训练。
	推荐的模型个数	从推荐模型中选择的模型个数，指定推荐模型的个数，使得模型的多样性更丰富，有助于提高最终模型的性能。 推荐模型的数量参数的范围是0到20。设置为0表示不使用推荐模型。 假设基模型算法池中有5个LightGBM（lgb）模型，且推荐的模型数量设置为5。这意味着除了基模型池中的5个LightGBM模型外，系统还会再推荐5个不同的模型。因此，总共有10个模型用于训练，其中5个是LightGBM模型，另外5个是系统根据数据情况推荐的不同模型。 当基模型算法池包含gatednet时，该参数不生效。
	计算模型权重特征重要性	是否在训练完成之后，计算模型的权重特征重要性，并在界面展示各特征的重要性分值及排序。支持权重特征重要性的模型有ada, lgb, xgb, rf, et, gb。当“基模型算法池”中至少配置以上模型中的一个，或“推荐的模型个数”至少为1时，用户可以打开此选项。否则此项无法打开，界面不展示相关信息。

模型类型	训练参数	说明
盘古融合推荐异常检测大模型（Pangu-Predict-Anom-Table-2.0.0）	数据集选择	选择训练所需的数据集。
	类别特征列	指定使用LabelEncoder处理的字符串类型类别特征的列表。格式为["列名1","列名2"]，默认设置为[]，表示没有需要处理的类别特征。 LabelEncoder的作用是将类别特征转换为数值型特征，使模型能够处理这些特征。
	非特征列	列出不需要输入到模型中的特征列，用于排除冗余或无意义的特征。格式为["列名1","列名2"]，默认设置为[]，表示所有特征都用于训练。
	标准化列	指定需要进行最大最小值标准化处理的数值特征的列表。格式为["列名1","列名2"]，默认设置为[]，表示没有特征需要标准化。标准化将特征值缩放到0到1的范围，处理分布差异较大的数值特征。
	预测目标列	指定预测目标变量的列名，仅支持单目标变量预测。格式为["列名"]，默认设置为[]，表示选择最后一列作为预测目标变量。
	训练集&验证集比例	将数据集划分为训练集和验证集，填入验证集比例即可，默认设置为0.2，即训练集占0.8，验证集占0.2，可选范围为0.1、0.2、0.3、0.4。
	基模型算法池	从预定义的算法池中选择用于训练模型的算法，算法包括：["knn", "iforest", "loda", "ocsvm"]，其中： ● knn表示k最近邻算法。 ● iforest表示孤立森林算法。 ● loda表示Loda算法。 ● ocsvm表示单类支持向量机算法。
	推荐的模型个数	从推荐模型中选择的模型个数，指定推荐模型的个数，使得模型的多样性更丰富，有助于提高最终模型的性能。 推荐模型的数量参数的范围是0到20。设置为0表示不使用推荐模型。 假设基模型算法池中有5个LightGBM（lgb）模型，且推荐的模型数量设置为5。这意味着除了基模型池中的5个LightGBM模型外，系统还会再推荐5个不同的模型。因此，总共有10个模型用于训练，其中5个是LightGBM模型，另外5个是系统根据数据情况推荐的不同模型。

模型类型	训练参数	说明
盘古融合推荐分类大模型（Pangu-Predict-Cla-Table-2.0.0）	数据集选择	选择训练所需的数据集。
	类别特征列	使用LabelEncoder处理的特征列的列表，用于处理字符串类型的类别特征。填写格式为中括号，其中列举带引号的特征列名，默认设置为中括号，表示没有需要使用LabelEncoder处理的特征。
	非特征列	不作为输入特征的列，此处填写的特征列名将不作为模型训练的输入特征，不要将预测目标列的列名填入。填写格式为中括号，其中列举带引号的特征列名，默认设置为中括号，表示选取全部特征用于训练。
	标准化列	使用最大最小值标准化处理的特征列的列表，用于处理分布差异较大的数值特征，填写格式为中括号，其中列举带引号的特征列名，默认设置为中括号，表示全部特征都不需要进行标准化。
	预测目标列	预测目标的列名，指定预测目标变量列名，仅支持单目标变量预测，填写格式为中括号，其中列举带引号的特征列名。默认设置为中括号，表示选择最后一列作为预测目标变量。
	训练集&验证集比例	数据集划分为训练集和验证集，填入验证集比例即可，默认设置为0.2，即训练集占0.8，验证集占0.2，可选范围为0.1、0.2、0.3、0.4。
	基模型算法池	从预定义的算法池中选择用于训练模型的算法，算法包括：["ada", "lgb", "xgb", "rf", "et", "gb", "gauss","mlp"]，其中： ● ada表示adaboost。 ● lgb表示lightgbm。 ● xgb表示xgboost。 ● rf表示随机森林。 ● et表示extraTree。 ● gb表示梯度提升树。 ● gauss表示高斯过程，gauss适合维度小于10且数据量小于500的样本数据。 ● mlp表示多层感知机，默认设置为5lgb，多种类算法示例: 3lgb,2rf,1xgb（表示使用3个LightGBM算法、2个随机森林算法和1个XGBoost算法）。

模型类型	训练参数	说明
	推荐的模型个数	由推荐模型推荐的模型数，参数的范围为0至20。0代表不使用推荐模型。假设基模型算法池为5lgb，推荐的模型个数为5，表示基模型有5个为lgb，另外5个是由推荐模型推荐的。 假设基模型算法池中有5个LightGBM（lgb）模型，且推荐的模型数量设置为5。这意味着除了基模型池中的5个LightGBM模型外，系统还会再推荐5个不同的模型。因此，总共有10个模型用于训练，其中5个是LightGBM模型，另外5个是系统根据数据情况推荐的不同模型。
	计算模型权重特征重要性	是否在训练完成之后，计算模型的权重特征重要性，并在界面展示各特征的重要性分值及排序。支持权重特征重要性的模型有ada, lgb, xgb, rf, et, gb。当“基模型算法池”中至少配置以上模型中的一个，或“推荐的模型个数”至少为1时，用户可以打开此选项。否则此项无法打开，界面不展示相关信息。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建科学计算大模型中期天气要素预测微调任务

1. 在“创建训练任务”页面，参考**表14-12**完成训练参数设置。

   其中，“数据配置”展示了各训练数据涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

表 14-12 科学计算大模型中期天气要素预测微调训练参数说明

参数分类	参数名称	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“科学计算大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“微调”。
	训练目标	选择“全量微调”。 ● 全量微调：在模型进行有监督微调时，对大模型的所有参数进行更新。这种方法通常能够实现最佳的模型性能，但需要消耗大量计算资源和时间，计算开销较大。
数据配置	训练数据	选择数据集中已发布的数据集，这里数据集需为再分析类型数据，同时需要完成加工作业，加工时需选择气象预处理算子。

参数分类	参数名称	参数说明
训练参数	训练集	选择训练数据中的部分时间数据，训练数据集尽可能多一些。
	验证集	选择验证集中的部分时间数据，验证集数据不能跟训练集数据重合。
	层次	设置训练数据的层次信息。在“预训练”场景中，可以添加或去除高空层次，训练任务将根据配置的层次信息重新训练模型。
	高空变量	设置训练数据的高空变量信息。在“预训练”场景中，可以添加或去除新的高空变量，选中后会在变量权重中增加或移除该变量，训练任务将根据配置的高空变量重新训练模型。
	表面变量	设置训练数据的表面变量信息。在“预训练”场景中，可以添加或去除新的表面变量，选中后会在变量权重中增加或移除该变量，训练任务将根据配置的表面变量重新训练模型。
	表面静态量	表面静态量默认包括地形高度、LAND_MASK和SOIL_TYPE，用于初始化模型状态并提供地表特性信息。当前不支持添加或去除这些静态量。 ● LAND_MASK：一个二维数组，表示模型网格中每个单元格是否是陆地。 ● SOIL_TYPE：表示地表土壤分类，影响土壤的物理和化学特性，如水分保持能力、热容量和导热性。
模型输出控制参数	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。取值范围：[1-1000]。
	损失类型	用来衡量模型预测结果与真实结果之间的差距的函数，提供MAE（平均绝对误差）、MSE（均方误差）两种损失函数。 ● MSE对于异常值非常敏感，因为它会放大较大的误差。因此，如果您数据中没有异常值，或者希望模型对大的误差给予更大的惩罚，可选择MSE。 ● 如果数据中存在异常值，或者希望模型对所有的误差都一视同仁，可选择MAE。
	表面变量相对高空变量的权重	指在模型训练过程中对表面变量相对于深海层变量赋予的权重，总Loss=高空Loss+surface_loss_weight*表面Loss。取值范围：(0.05, 10)。
正则化参数	路径删除概率	用于定义路径删除机制中的删除概率。路径删除是一种正则化技术，它在训练过程中随机删除一部分的网络连接，以防止模型过拟合。这个值越大，删除的路径越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0，1）。

参数分类	参数名称	参数说明
	特征删除概率	用于定义特征删除机制中的删除概率。特征删除（也称为特征丢弃）是另一种正则化技术，它在训练过程中随机删除一部分的输入特征，以防止模型过拟合。这个值越大，删除的特征越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1)。
	给输入数据加噪音的概率	定义了给输入数据加噪音的概率。加噪音是一种正则化技术，它通过在输入数据中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输入数据加噪音的尺度	定义了给输入数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	给输出数据加噪音的概率	定义了给输出数据加噪音的概率。加噪音是一种正则化技术，它通过在模型的输出中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输出数据加噪音的尺度	定义了给输出数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
优化器种类	优化器种类	优化器是用于更新模型参数的算法，目前支持ADAM优化器。
	第一个动量矩阵的指数衰减率(beta1)	用于定义ADAM优化器中的一阶矩估计的指数衰减率。一阶矩估计相当于动量，可以加速梯度在相关方向的下降并抑制震荡。取值范围：(0,1)。
	第二个动量矩阵的指数衰减率(beta_2)	用于定义ADAM优化器中的二阶矩估计的指数衰减率。二阶矩估计相当于RMSProp，可以调整学习率。取值范围：(0,1)。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂，取值需≥0。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。
	学习率调整策略	用于选择学习率调度器的类型。学习率调度器可以在训练过程中动态地调整学习率，以改善模型的训练效果。目前支持CosineDecayLR调度器。
变量权重	变量权重	训练数据设置完成后，会显示出各变量以及默认的权重。您可以基于变量的重要情况调整权重。

参数分类	参数名称	参数说明
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建科学计算大模型中期天气要素预测预训练任务

1. 在“创建训练任务”页面，参考**表14-13**完成训练参数设置。

   其中，“数据配置”展示了各训练数据涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

表 14-13 科学计算大模型中期天气要素预测预训练参数说明

参数分类	参数名称	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“科学计算大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“预训练”。
数据配置	训练数据	选择数据集中已发布的数据集，这里数据集需为再分析类型数据，同时需要完成加工作业，加工时需选择气象预处理算子。
	训练集	选择训练数据中的部分时间数据，训练数据集尽可能多一些。
	验证集	选择验证集中的部分时间数据，验证集数据不能跟训练集数据重合。

参数分类	参数名称	参数说明
	层次	设置训练数据的层次信息。在“预训练”场景中，可以添加或去除高空层次，训练任务将根据配置的层次信息重新训练模型。
	高空变量	设置训练数据的高空变量信息。在“预训练”场景中，可以添加或去除新的高空变量，选中后会在变量权重中增加或移除该变量，训练任务将根据配置的高空变量重新训练模型。
	表面变量	设置训练数据的表面变量信息。在“预训练”场景中，可以添加或去除新的表面变量，选中后会在变量权重中增加或移除该变量，训练任务将根据配置的表面变量重新训练模型。
	表面静态量	表面静态量默认包括地形高度、LAND_MASK和SOIL_TYPE，用于初始化模型状态并提供地表特性信息。当前不支持添加或去除这些静态量。 ● LAND_MASK：一个二维数组，表示模型网格中每个单元格是否是陆地。 ● SOIL_TYPE：表示地表土壤分类，影响土壤的物理和化学特性，如水分保持能力、热容量和导热性。
模型输出控制参数	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。取值范围：[1-1000]。
	损失类型	用来衡量模型预测结果与真实结果之间的差距的函数，提供MAE（平均绝对误差）、MSE（均方误差）两种损失函数。 ● MSE对于异常值非常敏感，因为它会放大较大的误差。因此，如果您数据中没有异常值，或者希望模型对大的误差给予更大的惩罚，可选择MSE。 ● 如果数据中存在异常值，或者希望模型对所有的误差都一视同仁，可选择MAE。
	表面变量相对高空变量的权重	指在模型训练过程中对表面变量相对于深海层变量赋予的权重，总Loss=高空Loss+surface_loss_weight*表面Loss。取值范围：(0.05, 10)。
正则化参数	路径删除概率	用于定义路径删除机制中的删除概率。路径删除是一种正则化技术，它在训练过程中随机删除一部分的网络连接，以防止模型过拟合。这个值越大，删除的路径越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0，1）。
	特征删除概率	用于定义特征删除机制中的删除概率。特征删除（也称为特征丢弃）是另一种正则化技术，它在训练过程中随机删除一部分的输入特征，以防止模型过拟合。这个值越大，删除的特征越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1)。

参数分类	参数名称	参数说明
	给输入数据加噪音的概率	定义了给输入数据加噪音的概率。加噪音是一种正则化技术，它通过在输入数据中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输入数据加噪音的尺度	定义了给输入数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	给输出数据加噪音的概率	定义了给输出数据加噪音的概率。加噪音是一种正则化技术，它通过在模型的输出中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输出数据加噪音的尺度	定义了给输出数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
优化器种类	优化器种类	优化器是用于更新模型参数的算法，目前支持ADAM优化器。
	第一个动量矩阵的指数衰减率(beta1)	用于定义ADAM优化器中的一阶矩估计的指数衰减率。一阶矩估计相当于动量，可以加速梯度在相关方向的下降并抑制震荡。取值范围：(0,1)。
	第二个动量矩阵的指数衰减率(beta_2)	用于定义ADAM优化器中的二阶矩估计的指数衰减率。二阶矩估计相当于RMSProp，可以调整学习率。取值范围：(0,1)。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂，取值需≥0。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。 预训练时，默认值为：0.00001，范围为[0, 0.001]
	学习率调整策略	用于选择学习率调度器的类型。学习率调度器可以在训练过程中动态地调整学习率，以改善模型的训练效果。目前支持CosineDecayLR调度器。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。

参数分类	参数名称	参数说明
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建科学计算大模型区域中期海洋智能预测微调任务

1. 在“创建训练任务”页面，参考表14-14完成训练参数设置。

   其中，“数据配置”展示了各训练数据涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

   表 14-14 科学计算大模型区域中期海洋智能预测微调参数说明

参数分类	参数名称	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“科学计算大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“微调”。
	训练目标	选择“全量微调”。 ● 全量微调：在模型进行有监督微调时，对大模型的所有参数进行更新。这种方法通常能够实现最佳的模型性能，但需要消耗大量计算资源和时间，计算开销较大。
	模型水平分辨率	模型网格在水平方向上的精细程度，通常用来表示模拟或预测中空间网格的大小。根据训练数据和业务需求，自行定义模型水平分辨率，取值>0。
数据配置	训练数据	选择数据集中已发布的数据集，这里数据集需为再分析类型数据，同时需要完成加工作业。
	训练集	选择训练数据中的部分时间数据，训练数据集尽可能多一些。

参数分类	参数名称	参数说明
	验证集	选择验证集中的部分时间数据，验证集数据不能跟训练集数据重合。
	深海层深	海深层深是指海洋模型将整个水柱（从海面到海底）按一定深度间隔划分成多个层次，每个深度值代表模型在这个深度层进行计算和模拟。例如，"0m"代表海平面，"6m"代表在海平面以下6米处的一层，以此类推。范围包括：0m、6m、10m、20m、30m、50m、70m、100m、125m、150m、200m、250m、300m、400m、500m。
	深海变量	深海变量是用于模拟和描述海洋状态的关键物理量。 T：15层：海温(℃) S：15层：海盐(PSU) U：15层：海流经向速率 (ms-1) V：15层：海流纬向速率 (ms-1)
	海表变量	海表变量用于描述海洋表层和其上方大气的状态的关键物理量。它们主要用于模拟和分析海洋表面的风速、温度、和气压等特征。 U10：1层：海表面10m经向风速(ms-1) V10：1层：海表面10m纬向风速(ms-1) T2m：1层：海表面2m温度 (℃) MSL：1层：平均海平面气压 (Pa) SP：1层：海表面气压 (Pa)
	表面静态量	表面静态量默认支持地形高度、LAND_MASK、SOIL_TYPE，用于初始化模型状态和在模型运行过程中提供必要的地表特性信息，暂时不支持添加和去除。 其中，LAND_MASK是一个二维数组，通常用于表示模型网格中每个单元格是否是陆地。SOIL_TYPE是指地表土壤的分类，它影响土壤的物理和化学特性，如土壤的水分保持能力、热容量和导热性。
模型输出控制参数	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。取值范围：[1-1000]。
	损失类型	用来衡量模型预测结果与真实结果之间的差距的函数，提供MAE（平均绝对误差）、MSE（均方误差）两种损失函数。 ● MSE对于异常值非常敏感，因为它会放大较大的误差。因此，如果您数据中没有异常值，或者希望模型对大的误差给予更大的惩罚，可选择MSE。 ● 如果数据中存在异常值，或者希望模型对所有的误差都一视同仁，可选择MAE。

参数分类	参数名称	参数说明
	海表变量相对深海变量的权重	指在模型训练过程中对海表变量相对于深海层变量赋予的权重，总Loss=深海层Loss+surface_loss_weight*海表Loss。取值范围：(0.05, 10)。
正则化参数	路径删除概率	用于定义路径删除机制中的删除概率。路径删除是一种正则化技术，它在训练过程中随机删除一部分的网络连接，以防止模型过拟合。这个值越大，删除的路径越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0，1）。
	特征删除概率	用于定义特征删除机制中的删除概率。特征删除（也称为特征丢弃）是另一种正则化技术，它在训练过程中随机删除一部分的输入特征，以防止模型过拟合。这个值越大，删除的特征越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1)。
	给输入数据加噪音的概率	定义了给输入数据加噪音的概率，定义了给输入数据加噪音的概率。加噪音是一种正则化技术，它通过在输入数据中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输入数据加噪音的尺度	给输入数据加噪音的尺度，定义了给输入数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	给输出数据加噪音的概率	给输出数据加噪音的概率，定义了给输出数据加噪音的概率。加噪音是一种正则化技术，它通过在模型的输出中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输出数据加噪音的尺度	给输出数据加噪音的尺度，定义了给输出数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
优化器参数	优化器种类	优化器种类。优化器是用于更新模型参数的算法，目前支持ADAM优化器。
	第一个动量矩阵的指数衰减率(beta1)	数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	第二个动量矩阵的指数衰减率(beta_2)	用于定义ADAM优化器中的二阶矩估计的指数衰减率。二阶矩估计相当于RMSProp，可以调整学习率。取值范围：(0,1)。

参数分类	参数名称	参数说明
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂，取值需≥0。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。
	学习率调整策略	用于选择学习率调度器的类型。学习率调度器可以在训练过程中动态地调整学习率，以改善模型的训练效果。目前支持CosineDecayLR调度器。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建科学计算大模型区域中期海洋智能预测预训练任务

1. 在“创建训练任务”页面，参考**表14-15**完成训练参数设置。

   其中，“数据配置”展示了各训练数据涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

表 14-15 科学计算大模型区域中期海洋智能预测预训练参数说明

参数分类	参数名称	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“科学计算大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“预训练”。
	模型水平分辨率	模型网格在水平方向上的精细程度，通常用来表示模拟或预测中空间网格的大小。根据训练数据和业务需求，自行定义模型水平分辨率，取值>0。
数据配置	训练数据	选择数据集中已发布的数据集，这里数据集需为再分析类型数据，同时需要完成加工作业。
	训练集	选择训练数据中的部分时间数据，训练数据集尽可能多一些。
	验证集	选择验证集中的部分时间数据，验证集数据不能跟训练集数据重合。
模型数据配置	深海层深	海深层深是指海洋模型将整个水柱（从海面到海底）按一定深度间隔划分成多个层次，每个深度值代表模型在这个深度层进行计算和模拟。例如，"0m"代表海平面，"6m"代表在海平面以下6米处的一层，以此类推。范围包括：0m、6m、10m、20m、30m、50m、70m、100m、125m、150m、200m、250m、300m、400m、500m。
	深海变量	深海变量是用于模拟和描述海洋状态的关键物理量。 T：15层：海温(℃) S：15层：海盐(PSU) U：15层：海流经向速率 (ms-1) V：15层：海流纬向速率 (ms-1)
	表面变量	海表变量用于描述海洋表层和其上方大气的状态的关键物理量。它们主要用于模拟和分析海洋表面的风速、温度、和气压等特征。 U10：1层：海表面10m经向风速(ms-1) V10：1层：海表面10m纬向风速(ms-1) T2m：1层：海表面2m温度 (℃) MSL：1层：平均海平面气压 (Pa) SP：1层：海表面气压 (Pa)

参数分类	参数名称	参数说明
	表面静态量	表面静态量默认支持地形高度、LAND_MASK、SOIL_TYPE，用于初始化模型状态和在模型运行过程中提供必要的地表特性信息，暂时不支持添加和去除。 其中，LAND_MASK是一个二维数组，通常用于表示模型网格中每个单元格是否是陆地。SOIL_TYPE是指地表土壤的分类，它影响土壤的物理和化学特性，如土壤的水分保持能力、热容量和导热性。
模型输出控制参数	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。取值范围：[1-1000]。
	损失类型	用来衡量模型预测结果与真实结果之间的差距的函数，提供MAE（平均绝对误差）、MSE（均方误差）两种损失函数。 ● MSE对于异常值非常敏感，因为它会放大较大的误差。因此，如果您数据中没有异常值，或者希望模型对大的误差给予更大的惩罚，可选择MSE。 ● 如果数据中存在异常值，或者希望模型对所有的误差都一视同仁，可选择MAE。
	海表变量相对深海变量的权重	指在模型训练过程中对海表变量相对于深海层变量赋予的权重，总Loss=深海层Loss+surface_loss_weight*海表Loss。取值范围：(0.05, 10)。
正则化参数	路径删除概率	用于定义路径删除机制中的删除概率。路径删除是一种正则化技术，它在训练过程中随机删除一部分的网络连接，以防止模型过拟合。这个值越大，删除的路径越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0，1）。
	特征删除概率	用于定义特征删除机制中的删除概率。特征删除（也称为特征丢弃）是另一种正则化技术，它在训练过程中随机删除一部分的输入特征，以防止模型过拟合。这个值越大，删除的特征越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1)。
	给输入数据加噪音的概率	定义了给输入数据加噪音的概率，定义了给输入数据加噪音的概率。加噪音是一种正则化技术，它通过在输入数据中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输入数据加噪音的尺度	给输入数据加噪音的尺度，定义了给输入数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	给输出数据加噪音的概率	给输出数据加噪音的概率，定义了给输出数据加噪音的概率。加噪音是一种正则化技术，它通过在模型的输出中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。

参数分类	参数名称	参数说明
	给输出数据加噪音的尺度	给输出数据加噪音的尺度，定义了给输出数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
优化器参数	优化器种类	优化器种类。优化器是用于更新模型参数的算法，目前支持ADAM优化器。
	第一个动量矩阵的指数衰减率(beta1)	数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	第二个动量矩阵的指数衰减率(beta_2)	用于定义ADAM优化器中的二阶矩估计的指数衰减率。二阶矩估计相当于RMSProp，可以调整学习率。取值范围：(0,1)。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂，取值需≥0。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。 预训练时，默认值为：0.00001，范围为[0, 0.001]。
	学习率调整策略	用于选择学习率调度器的类型。学习率调度器可以在训练过程中动态地调整学习率，以改善模型的训练效果。目前支持CosineDecayLR调度器。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。

参数分类	参数名称	参数说明
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建科学计算大模型区域大气污染物预测微调任务

1. 在“创建训练任务”页面，参考**表14-16**完成训练参数设置。

   其中，“数据配置”展示了各训练数据涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

   表 14-16 科学计算大模型区域大气污染物预测微调参数说明

参数分类	参数名称	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“科学计算大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“微调”。
	训练目标	选择“全量微调”。 ● 全量微调：在模型进行有监督微调时，对大模型的所有参数进行更新。这种方法通常能够实现最佳的模型性能，但需要消耗大量计算资源和时间，计算开销较大。
	模型水平分辨率	模型网格在水平方向上的精细程度，通常用来表示模拟或预测中空间网格的大小。根据训练数据和业务需求，自行定义模型水平分辨率，取值>0。
	模型时间分辨率	模型预报的时间间隔，类型为整数，可取值 1,3,6,24，单位为小时。
数据配置	训练数据	选择数据集中已发布的数据集，这里数据集需为再分析类型数据，同时需要完成加工作业。
	训练集	选择训练数据中的部分时间数据，训练数据集尽可能多一些。
	验证集	选择验证集中的部分时间数据，验证集数据不能跟训练集数据重合。

参数分类	参数名称	参数说明
	表面静态量	表面静态量默认支持地形高度、LAND_MASK、SOIL_TYPE，用于初始化模型状态和在模型运行过程中提供必要的地表特性信息，暂时不支持添加和去除。 其中，LAND_MASK是一个二维数组，通常用于表示模型网格中每个单元格是否是陆地。SOIL_TYPE是指地表土壤的分类，它影响土壤的物理和化学特性，如土壤的水分保持能力、热容量和导热性。
	污染物变量	污染物变量是用于模拟和描述大气污染物浓度的关键物理量。 PM10（$\mu g/m^3$） PM2.5（$\mu g/m^3$） NO_2：二氧化氮（$\mu g/m^3$） O_3：臭氧（$\mu g/m^3$） SO_2：二氧化硫（$\mu g/m^3$） CO：一氧化碳（mg/m^3）
模型输出控制参数	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。取值范围：[1-1000]。
	损失类型	用来衡量模型预测结果与真实结果之间的差距的函数，提供MAE（平均绝对误差）、MSE（均方误差）两种损失函数。 ● MSE对于异常值非常敏感，因为它会放大较大的误差。因此，如果您数据中没有异常值，或者希望模型对大的误差给予更大的惩罚，可选择MSE。 ● 如果数据中存在异常值，或者希望模型对所有的误差都一视同仁，可选择MAE。
正则化参数	路径删除概率	用于定义路径删除机制中的删除概率。路径删除是一种正则化技术，它在训练过程中随机删除一部分的网络连接，以防止模型过拟合。这个值越大，删除的路径越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0，1）。
	特征删除概率	用于定义特征删除机制中的删除概率。特征删除（也称为特征丢弃）是另一种正则化技术，它在训练过程中随机删除一部分的输入特征，以防止模型过拟合。这个值越大，删除的特征越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1)。
	给输入数据加噪音的概率	定义了给输入数据加噪音的概率，定义了给输入数据加噪音的概率。加噪音是一种正则化技术，它通过在输入数据中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输入数据加噪音的尺度	给输入数据加噪音的尺度，定义了给输入数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。

参数分类	参数名称	参数说明
	给输出数据加噪音的概率	给输出数据加噪音的概率，定义了给输出数据加噪音的概率。加噪音是一种正则化技术，它通过在模型的输出中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输出数据加噪音的尺度	给输出数据加噪音的尺度，定义了给输出数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
优化器参数	优化器种类	优化器种类。优化器是用于更新模型参数的算法，目前支持ADAM优化器。
	第一个动量矩阵的指数衰减率(beta1)	数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	第二个动量矩阵的指数衰减率(beta_2)	用于定义ADAM优化器中的二阶矩估计的指数衰减率。二阶矩估计相当于RMSProp，可以调整学习率。取值范围：(0,1)。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂，取值需≥0。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。 预训练时，默认值为：0.00001，范围为[0, 0.001]。
	学习率调整策略	用于选择学习率调度器的类型。学习率调度器可以在训练过程中动态地调整学习率，以改善模型的训练效果。目前支持CosineDecayLR调度器。
变量权重	变量权重	训练数据设置完成后，会显示出各变量以及默认的权重。您可以基于变量的重要情况调整权重。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。

参数分类	参数名称	参数说明
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

创建科学计算大模型区域大气污染物预测预训练任务

1. 在“创建训练任务”页面，参考**表14-17**完成训练参数设置。

   其中，“数据配置”展示了各训练数据涉及到的全部参数，请根据具体前端页面展示的参数进行设置。

   表 14-17 科学计算大模型区域大气污染物预测预训练参数说明

参数分类	参数名称	参数说明
训练配置	选择模型	可以修改如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“科学计算大模型”，并选择训练所用的基础模型和版本。
	训练类型	选择“预训练”。
	模型时间分辨率	模型预报的时间间隔，类型为整数，可取值 1,3,6,24，单位为小时。
	模型水平分辨率	模型网格在水平方向上的精细程度，通常用来表示模拟或预测中空间网格的大小。根据训练数据和业务需求，自行定义模型水平分辨率，取值>0。
数据配置	训练数据	选择数据集中已发布的数据集，这里数据集需为再分析类型数据，同时需要完成加工作业。
	训练集	选择训练数据中的部分时间数据，训练数据集尽可能多一些。
	验证集	选择验证集中的部分时间数据，验证集数据不能跟训练集数据重合。

参数分类	参数名称	参数说明
模型数据配置	表面静态量	表面静态量默认支持地形高度、LAND_MASK、SOIL_TYPE，用于初始化模型状态和在模型运行过程中提供必要的地表特性信息，暂时不支持添加和去除。 其中，LAND_MASK是一个二维数组，通常用于表示模型网格中每个单元格是否是陆地。SOIL_TYPE是指地表土壤的分类，它影响土壤的物理和化学特性，如土壤的水分保持能力、热容量和导热性。
	污染物变量	污染物变量是用于模拟和描述大气污染物浓度的关键物理量。 PM10（μg/m³） PM2.5（μg/m³） NO_2：二氧化氮（μg/m³） O_3：臭氧（μg/m³） SO_2：二氧化硫（μg/m³） CO：一氧化碳（mg/m³）
模型输出控制参数	训练轮数	表示完成全部训练数据集训练的次数。每个轮次都会遍历整个数据集一次。取值范围：[1-1000]。
	损失类型	用来衡量模型预测结果与真实结果之间的差距的函数，提供MAE（平均绝对误差）、MSE（均方误差）两种损失函数。 ● MSE对于异常值非常敏感，因为它会放大较大的误差。因此，如果您数据中没有异常值，或者希望模型对大的误差给予更大的惩罚，可选择MSE。 ● 如果数据中存在异常值，或者希望模型对所有的误差都一视同仁，可选择MAE。
正则化参数	路径删除概率	用于定义路径删除机制中的删除概率。路径删除是一种正则化技术，它在训练过程中随机删除一部分的网络连接，以防止模型过拟合。这个值越大，删除的路径越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0，1）。
	特征删除概率	用于定义特征删除机制中的删除概率。特征删除（也称为特征丢弃）是另一种正则化技术，它在训练过程中随机删除一部分的输入特征，以防止模型过拟合。这个值越大，删除的特征越多，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1)。
	给输入数据加噪音的概率	定义了给输入数据加噪音的概率，定义了给输入数据加噪音的概率。加噪音是一种正则化技术，它通过在输入数据中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输入数据加噪音的尺度	给输入数据加噪音的尺度，定义了给输入数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。

参数分类	参数名称	参数说明
	给输出数据加噪音的概率	给输出数据加噪音的概率，定义了给输出数据加噪音的概率。加噪音是一种正则化技术，它通过在模型的输出中添加随机噪音来增强模型的泛化能力。取值范围：[0,1]。
	给输出数据加噪音的尺度	给输出数据加噪音的尺度，定义了给输出数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
优化器参数	优化器种类	优化器种类。优化器是用于更新模型参数的算法，目前支持ADAM优化器。
	第一个动量矩阵的指数衰减率(beta1)	数据加噪音的尺度。这个值越大，添加的噪音越强烈，模型的正则化效果越强，但同时也可能会降低模型的拟合能力。取值范围：[0,1]。
	第二个动量矩阵的指数衰减率(beta_2)	用于定义ADAM优化器中的二阶矩估计的指数衰减率。二阶矩估计相当于RMSProp，可以调整学习率。取值范围：(0,1)。
	权重衰减系数	通过在损失函数中加入与模型权重大小相关的惩罚项，鼓励模型保持较小的权重，防止过拟合或模型过于复杂，取值需≥0。
	学习率	学习率决定每次训练中模型参数更新的幅度。 选择合适的学习率至关重要： ● 如果学习率过大，模型可能无法收敛。 ● 如果学习率过小，模型的收敛速度将变得非常慢。 预训练时，默认值为：0.00001，范围为[0, 0.001]。
	学习率调整策略	用于选择学习率调度器的类型。学习率调度器可以在训练过程中动态地调整学习率，以改善模型的训练效果。目前支持CosineDecayLR调度器。
变量权重	变量权重	训练数据设置完成后，会显示出各变量以及默认的权重。您可以基于变量的重要情况调整权重。
资源配置	计费模式	选择训练当前任务的计费模式。
	训练单元	选择训练模型所需的训练单元。 当前展示的完成本次训练所需要的最低训练单元要求。
	单实例训练单元数	选择单实例训练单元数。
	实例数	选择实例数。

参数分类	参数名称	参数说明
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
发布模型	开启自动发布	开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
基本信息	名称	训练任务名称。
	描述	训练任务描述。

2. 参数填写完成后，单击“立即创建”。
3. 创建好训练任务后，页面将返回“模型训练”页面，可随时查看当前任务的状态。

15 创建压缩任务

解释说明

模型压缩可以降低推理显存占用，节省推理资源提高推理性能。

当前支持压缩的模型类型为：NLP大模型。

操作步骤

1. 在“创建压缩任务”页面，模型来源选择“盘古大模型”。
2. 选择需要压缩的基础模型，支持选择已发布模型或未发布模型。
3. 选择压缩策略。除INT8压缩策略外，部分模型支持INT4压缩策略，可在选择模型后，根据页面展示的策略进行选择。
   - INT8：该压缩策略将模型参数压缩至8位字节，可以有效降低推理显存占用。
   - INT4：该压缩策略与INT8相比，可以进一步减少模型的存储空间和计算复杂度。
4. 选择计费模式并设置训练单元，可选择开启订阅提醒。
5. 选择开启自动发布后，模型训练完成的最终产物会自动发布为空间资产，以便对模型进行压缩、部署、评测等操作或共享给其他空间。
6. 填写基本信息，包括任务名称、压缩后模型名称与描述，单击“立即创建”。

16 创建部署任务

解释说明

模型训练完成后，可以执行模型的部署操作。

创建 NLP 大模型部署任务

1. 在“创建部署”页面，参考**表16-1**完成部署参数设置。

表 16-1 NlP 大模型部署参数说明

参数分类	部署参数	参数说明
部署配置	选择模型	选择模型需要选择如下信息： • 来源：选择“模型广场”。 • 类型：选择“大语言模型”，并选择需要进行部署的模型和版本。
	部署方式	选择“云上部署”。
	自定义名称	此名称是通过V2版本推理接口调用该推理服务时的唯一标识。创建后不支持修改。
安全护栏	开启并同意授权	安全护栏保障模型调用安全。
	版本选择	当前支持安全护栏基础版，内置了默认的内容审核规则。
资源配置	计费模式	限时免费。
	实例数	设置部署模型时所需的实例数。

参数分类	部署参数	参数说明
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
基本信息	服务名称	设置部署任务的名称。
	描述（选填）	设置部署任务的描述。

2. 参数填写完成后，单击“立即部署”。

创建向量&重排模型部署任务

1. 在“创建部署”页面，参考**表16-2**完成部署参数设置。

表 16-2 向量&重排模型部署参数说明

参数分类	部署参数	参数说明
部署配置	选择模型	选择模型需要选择如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“向量&重排模型”，并选择需要进行部署的模型和版本。
	部署方式	选择“云上部署”。
资源配置	计费模式	限时免费。
	实例数	设置部署模型时所需的实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
基本信息	服务名称	设置部署任务的名称。
	描述（选填）	设置部署任务的描述。

2. 参数填写完成后，单击“立即部署”。

创建 CV 大模型云上部署任务

1. 在“创建部署”页面，参考**表16-3**完成部署参数设置。

表 16-3 CV 大模型云上部署参数说明

参数分类	部署参数	参数说明
部署配置	选择模型	选择模型需要选择如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“CV大模型”，并选择需要进行部署的模型和版本。
	部署方式	● 选择“云上部署”。 ● 若选择“边缘部署”，详细步骤请参见《用户指南》。
资源配置	计费模式	限时免费。
	实例数	设置部署模型时所需的实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
基本信息	服务名称	设置部署任务的名称。
	描述（选填）	设置部署任务的描述。

2. 参数填写完成后，单击“立即部署”。

创建科学计算大模型部署任务

1. 在“创建部署”页面，参考**表16-4**完成部署参数设置。

表 16-4 科学计算大模型云上部署参数说明

参数分类	部署参数	参数说明
部署配置	选择模型	选择模型需要选择如下信息： ● 来源：选择“模型广场”。 ● 类型：选择“科学计算大模型”，并选择需要进行部署的模型和版本。
	部署方式	● 选择“云上部署”。 ● 若选择“边缘部署”，详细步骤请参见《用户指南》。
	作业输入方式	选择 “OBS”表示从OBS中读取数据。

参数分类	部署参数	参数说明
	作业输出方式	选择 “OBS”表示将输出结果存储在OBS中。
	作业配置参数	设置模型部署参数信息，平台已给出默认值。
资源配置	计费模式	限时免费。
	实例数	设置部署模型时所需的实例数。
订阅提醒	订阅提醒	该功能开启后，系统将在任务状态更新时，通过短信或邮件将提醒发送给用户。
基本信息	服务名称	设置部署任务的名称。
	描述（选填）	设置部署任务的描述。

2. 参数填写完成后，单击“立即部署”。

17 创建评测任务

解释说明

ModelArts Studio平台支持对NLP大模型、CV大模型、预测大模型进行评测。

在创建评测任务时，用户可以选择已部署的模型进行评测，或通过外部API接入模型进行评测。每次最多支持评测10个模型。

当前支持评测的模型类型为：NLP大模型、CV大模型、预测大模型。

创建 NLP 大模型自动评测任务

1. 在“创建自动评测任务”页面，参考**表17-1**或**表17-2**完成部署参数设置。

表 17-1 NLP 大模型自动评测任务参数说明（基于规则）

参数分类	参数名称	参数说明
选择服务	模型类型	选择“大语言模型”。
	服务来源	支持已部署服务、外部服务两种选项。单次最多可评测10个模型。 • 已部署服务：选择部署至ModelArts Studio平台的模型进行评测。 • 外部服务：通过API的方式接入外部模型进行评测。选择外部服务时，需要填写外部模型的接口名称、接口地址、请求体、响应体等信息。 – 请求体支持openai、tgi、自定义三种格式。openai格式即是由OpenAI公司开发并标准化的一种大模型请求格式；tgi格式即是Hugging Face团队推出的一种大模型请求格式。 – 接口的响应体需要按照jsonpath语法要求进行填写，jsonpath语法的作用是从响应体的json字段中提取出所需的数据。

参数分类	参数名称	参数说明
评测配置	评测规则	选择“基于规则”：基于规则自动打分，即基于相似度/准确率进行打分，对比模型预测结果与标注数据的差异，适合标准选择题或简单问答场景。
	评测数据集	● 预置评测集：使用预置的专业数据集进行评测。 ● 自定义评测集：由用户指定评测指标（F1分数、准去率、BLEU、Rouge）并上传评测数据集进行评测。 选择“自定义评测集”时需要上传待评测数据集。
	评测结果存储位置	模型评测结果的存储位置。
基本信息	评测任务名称	填写评测任务名称。
	描述	填写评测任务描述。

表 17-2 NLP 大模型自动评测任务参数说明（基于大模型）

参数分类	参数名称	参数说明
选择服务	模型类型	选择“大语言模型”。
	服务来源	支持已部署服务、外部服务两种选项。单次最多可评测10个模型。 ● 已部署服务：选择部署至ModelArts Studio平台的模型进行评测。 ● 外部服务：通过API的方式接入外部模型进行评测。选择外部服务时，需要填写外部模型的接口名称、接口地址、请求体、响应体等信息。 – 请求体支持openai、tgi、自定义三种格式。openai格式即是由OpenAI公司开发并标准化的一种大模型请求格式；tgi格式即是Hugging Face团队推出的一种大模型请求格式。 – 接口的响应体需要按照jsonpath语法要求进行填写，jsonpath语法的作用是从响应体的json字段中提取出所需的数据。
评测配置	评测规则	选择“基于大模型”：使用能力更强的大模型对被评估模型的生成结果进行自动化打分，适用于开放性或复杂问答场景。

参数分类	参数名称	参数说明
	选择模式	• 评分模式：裁判模型将根据设置的评分标准对模型推理结果自动进行打分 • 对比模式：裁判模型将对比两个模型在每个评测题目上的表现，选择win、lose、tie展示对比结果 ，对比模式下服务来源必须选择2个服务，默认所选择的第一个服务作为基准模型。
	打分Prompt	• 评分模式下默认为score_prompt，该Prompt包含当前场景的标准回复，将Prompt在评分环节输入至裁判模型中进行打分。 • 对比模式下默认为arena_prompt，该Prompt包含当前场景的标准回复，意图让裁判模型比较两个服务的优劣。 在此过程中，用户可在右侧“变量”中修改metric评价维度等指标。您可对评分指标和评分步骤等内容进行修改
	评测数据集	选择需要评测的数据集。NLP多轮问答场景仅支持基于大模型自动评测，可选择多轮问答评测数据集。
	评测结果存储位置	模型评测结果的存储位置。
裁判员配置	裁判模型	支持选择已部署的服务或外部服务。
	打分规则	打分规则支持自定义配置，裁判模型将根据设定的规则对模型结果进行打分或对比。
基本信息	评测任务名称	填写评测任务名称。
	描述	填写评测任务描述。

2. 参数填写完成后，单击“立即创建”。

创建 NLP 大模型人工评测任务

1. 在“创建人工评测任务”页面，参考**表17-3**完成部署参数设置。

表 17-3 NLP 大模型人工评测任务参数说明

参数分类	参数名称	参数说明
选择服务	模型类型	选择“大语言模型”。

参数分类	参数名称	参数说明
	服务来源	支持已部署服务、外部服务两种选项。单次最多可评测10个模型。 ● 已部署服务：选择部署至ModelArts Studio平台的模型进行评测。 ● 外部服务：通过API的方式接入外部模型进行评测。选择外部服务时，需要填写外部模型的接口名称、接口地址、请求体、响应体等信息。 – 请求体支持openai、tgi、自定义三种格式。openai格式即是由OpenAI公司开发并标准化的一种大模型请求格式；tgi格式即是Hugging Face团队推出的一种大模型请求格式。 – 接口的响应体需要按照jsonpath语法要求进行填写，jsonpath语法的作用是从响应体的json字段中提取出所需的数据。
评测配置	评测指标	由用户自定义评测指标并填写评测标准。
	评测数据集	待评测的数据集。
	评测结果存储位置	模型评测结果的存储位置。
基本信息	评测任务名称	填写评测任务名称。
	描述	填写评测任务描述。

2. 参数填写完成后，单击“立即创建”。

创建 CV 大模型自动评测任务

1. 在“创建自动评测任务”页面，参考**表17-4**完成部署参数设置。

表 17-4 创建 CV 大模型自动评测任务参数说明（基于规则）

参数分类	参数名称	参数说明
选择服务	模型类型	选择“CV大模型”。
	评测模型	选择“物体检测”

参数分类	参数名称	参数说明
	服务来源	支持已部署服务。单次最多可评测10个模型。 ● 已部署服务：选择部署至ModelArts Studio平台的模型进行评测。
评测配置	评测规则	选择“基于规则”。
	评测数据	待评测的数据集。
	数据集标注模式	选择“有标注模式”。
	评测结果存储位置	模型评测结果的存储位置。
基本信息	评测任务名称	填写评测任务名称。
	描述	填写评测任务描述。

2. 参数填写完成后，单击“立即创建”。

创建预测大模型自动评测任务

1. 在“创建自动评测任务”页面，参考**表17-5**完成评测任务参数设置。

表 17-5 创建预测大模型自动评测任务参数说明

参数分类	参数名称	参数说明
选择服务	模型类型	选择“预测大模型”。
	评测模型	当前支持预测大模型3种模型场景，分别是： ● 回归 ● 分类 ● 异常检测
	服务来源	当前仅支持已部署服务 ● 已部署服务：选择部署至ModelArts Studio平台的模型进行评测。
评测配置	评测规则	仅支持“基于规则”。

参数分类	参数名称	参数说明
	评测数据集	待评测的数据集。
	评测结果存储位置	模型评测结果的存储位置。
	数据集配置	预测目标列：指定预测目标变量列名（注：当前仅支持单变量预测）格式：[""]。如：["result"]，[]表示默认最后一列。
基本信息	评测任务名称	填写评测任务名称。
	描述	填写评测任务描述。

2. 参数填写完成后，单击“立即创建”。

18 模型资产

模型资产介绍

用户在平台中可试用已订购或训练后发布的模型，将被视为模型资产并存储在空间资产内，方便统一管理与操作。用户可以查看模型的所有历史版本及操作记录，从而追踪模型的演变过程。同时，平台支持一系列便捷操作，包括模型训练、部署、评测等，帮助用户简化模型开发及应用流程。此外，平台还具备导入、导出盘古大模型的功能，便于用户将其他局点的盘古大模型迁移至本局点，确保模型资产在多局点环境中的灵活使用与共享。这些功能有助于用户高效管理模型生命周期，提高资产管理效率。

模型资产包含以下两种形式：

- **预置模型。**
  用户在平台中可试用、已订购的预置模型。
- **用户自行发布的模型**。
  用户可以将训练完成的模型发布为模型资产。发布的模型支持查看详细信息、编辑属性、删除、导出、导入等操作。

管理模型资产

1. 登录ModelArts Studio大模型开发平台，在“我的空间”模块，单击进入所需空间。
2. 在左侧导航栏中选择“空间资产 > 模型”。
3. 单击“本空间”或“其他空间”页签，可对用户在当前空间或其他空间发布的模型执行以下操作：
   - 查看模型信息。单击模型名称，进入模型信息页面，可产模型的基本信息与操作记录。
   - 编辑属性。单击操作列的“编辑属性”，可修改模型资产名称、描述以及资产可见性。
   - 训练、压缩、部署。可在模型列表页面，对模型执行训练、压缩或部署操作。单击相应按钮，将跳转至相关操作页面。