解决方案实践

数字人交互智能问答解决方案

文档版本 2.0

发布日期 2025-08-25

华为技术有限公司

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。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：
https://www.huawei.com/cn/psirt/vul-response-process
如企业客户须获取漏洞信息，请参见如下网址：
https://securitybulletin.huawei.com/enterprise/cn/security-advisory

目 录

1 方案概述

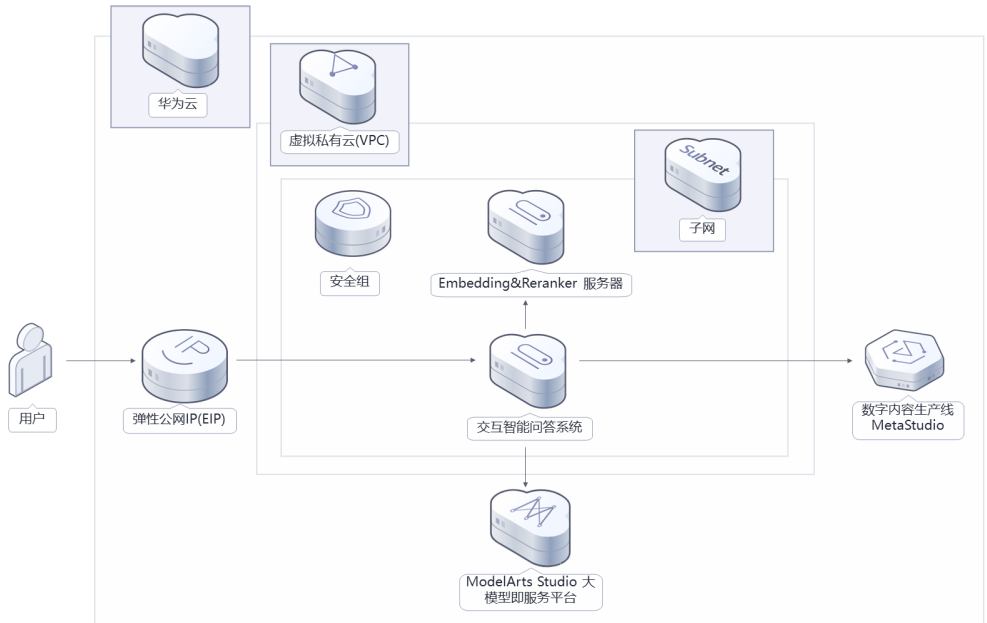
应用场景

该解决方案基于华为云**数字内容生产线** MetaStudio，ModelArts Studio**大模型即服务平台**和**Dify**快速部署数字人交互服务，部署后用户只需简单配置几项关键参数即可直接使用数字人交互服务。数字人交互服务是通过人工智能技术驱动的虚拟形象，提供拟人化、多模态的实时交互体验，已在多个领域实现商业化应用。

方案架构

该解决方案基于MetaStudio，ModelArts Studio以及Flexus云服务器X实例帮助您快速部署数字人交互服务。

图 1-1 方案架构图



该解决方案将会部署如下资源：

- 创建一个**弹性公网IP EIP**并关联部署Dify的Flexus云服务器X实例，提供访问公网和被公网访问能力。

- 创建两台**Flexus云服务器X实例**，分别用于搭建Dify-LLM应用开发平台和用于知识库优化的Embedding，Reranker模型。
- 创建一个安全组，通过配置安全组规则，为云服务器提供安全防护。
- 创建一个**数字内容生产线 MetaStudio 智能交互**。
- 开通**ModelArts Studio大模型即服务平台**用于大模型在线推理服务

方案优势

- 开箱即用

  快速部署，用户只需填写必要参数几个步骤即可使用智能数字人交互解决方案。
- 低成本

  提供高性价比的云服务器，用户可以根据实际需求自定义不同规格的云服务器。
- 一键部署

  一键轻松部署，即可完成智能数字人交互解决方案搭建。

约束与限制

- 该解决方案部署前，需注册华为账号并开通华为云，完成实名认证，且账号不能处于欠费或冻结状态。如果计费模式选择“包年包月”，请确保账户余额充足以便一键部署资源的时候可以自动支付；或者在一键部署的过程进入费用中心，找到“待支付订单”并手动完成支付。

2 资源和成本规划

该解决方案主要部署如下资源，以下费用仅供参考，具体请参考华为云官网**价格详情**，实际收费以账单为准。

表 2-1 资源和成本规划（按需计费）

华为云服务	资源名称	配置示例	数量	每月预估花费
虚拟私有云 VPC	digital-human-interaction	• VPC网段：172.16.0.0/16 • 区域：华北-北京四	1	0.00元
子网 Subnet	digital-human-interaction-subnet	• 子网网段：172.16.1.0/24 • 区域：华北-北京四	1	0.00元
安全组 SecurityGroup	digital-human-interaction	• 允许ping：0.0.0.0/0 • 开放端口22允许Cloud Shell 登录：121.36.59.153/32 • 开放端口8000，用于与数字人交互服务通信 • 区域：华北-北京四	1	0.00元

华为云服务	资源名称	配置示例	数量	每月预估花费
Flexus云服务器X实例	digital-human-interaction-dify	• 按需计费：1.00元/小时 • 区域：华北-北京四 • 规格：通用计算型 | x1.8u.16g | 8vCPUs | 16GiB • 镜像：dify0.15.2_workflow_rag_v2.1 • 系统盘：高IO | 100GB	1	720.00元
Flexus云服务器X实例	digital-human-interaction-model	• 按需计费：2.28元/小时 • 区域：华北-北京四 • 规格：通用计算增强型 | x1e.16u.16g | 16vCPUs | 16GiB • 镜像：embedding-reranker-models_v2.1 • 系统盘：高IO | 40GB	1	1645.06元
弹性公网IP EIP	digital-human-interaction-eip	• 按需计费：0.80元/GB • 区域：华北-北京四 • 线路：动态BGP • 公网带宽：按流量计费 • 带宽大小：300Mbit/s	1	0.80元/GB

华为云服务	资源名称	配置示例	数量	每月预估花费
数字内容生产线 MetaStudio	digital-human-interaction	• 包年包月 • 区域：华北-北京四 • 资源：分身数字人智能交互 并发 x 1 • 数量：1	1	1800元
MaaS tokens 计费	-	• 计费模式：按Token计费 • 模型：DeepSeek-V3-32K • 输入价格：0.002元 / 千tokens • 输出价格：0.008元 / 千tokens	-	输入0.002元 / 千tokens 输出0.008元 / 千tokens
合计		-		**4165.06元 + 弹性公网IP EIP费用 + MaaS tokens费用**

表 2-2 资源和成本规划（包年包月）

华为云服务	资源名称	配置示例	数量	每月预估花费
虚拟私有云 VPC	digital-human-interaction	• VPC网段：172.16.0.0/16 • 区域：华北-北京四	1	0.00元
子网 Subnet	digital-human-interaction-subnet	• 子网网段：172.16.1.0/24 • 区域：华北-北京四	1	0.00元

华为云服务	资源名称	配置示例	数量	每月预估花费
安全组 SecurityGroup	digital-human-interaction	• 允许ping：0.0.0.0/0 • 开放端口22允许Cloud Shell 登录：121.36.59.153/32 • 开放端口8000，用于与数字人交互服务通信 • 区域：华北-北京四	1	0.00元
Flexus云服务器X实例	digital-human-interaction-dify	• 包年包月 • 区域：华北-北京四 • 规格：通用计算型 | x1.8u.16g | 8vCPUs | 16GiB • 镜像：dify0.15.2_workflow_rag_v2.1 • 系统盘：高IO | 100GB	1	502.00元
Flexus云服务器X实例	digital-human-interaction-model	• 包年包月 • 区域：华北-北京四 • 规格：通用计算增强型 | x1e.16u.16g | 16vCPUs | 16GiB • 镜像：embedding-reranker-models_v2.1 • 系统盘：高IO | 40GB	1	1108.00元

华为云服务	资源名称	配置示例	数量	每月预估花费
弹性公网IP EIP	digital-human-interaction-eip	• 按需计费：0.80元/GB • 区域：华北-北京四 • 线路：动态BGP • 公网带宽：按流量计费 • 带宽大小：300Mbit/s	1	0.80元/GB
数字内容生产线 MetaStudio	digital-human-interaction	• 包年包月 • 区域：华北-北京四 • 资源：分身数字人智能交互 并发 x 1 • 数量：1	1	1800元
MaaS tokens 计费	-	• 计费模式：按Token计费 • 模型：DeepSeek-V3-32K • 输入价格：0.002元 / 千tokens • 输出价格：0.008元 / 千tokens	-	输入0.002元 / 千tokens 输出0.008元 / 千tokens
合计		-		**3410.00元 + 弹性公网IP EIP费用 + MaaS tokens费用**

3 实施步骤

3.1 准备工作

创建 rf_admin_trust 委托（可选）

当您首次使用华为云时注册的账号，则无需执行该准备工作，如果您使用的是IAM用户账户，请确认您是否在admin用户组中，如果您不在admin组中，则需要为您的账号**授予相关权限**，并完成以下准备工作。

步骤1 进入华为云官网，打开**控制台管理**界面，鼠标移动至个人账号处，打开“统一身份认证”菜单。

图 3-1 控制台管理界面

图 3-2 统一身份认证菜单

步骤2 进入“委托”菜单，搜索“rf_admin_trust”委托。

图 3-3 委托列表

- 如果委托存在，则不用执行接下来的创建委托的步骤
- 如果委托不存在时执行接下来的步骤创建委托

步骤3 单击步骤2界面中的“创建委托”按钮，在委托名称中输入“rf_admin_trust”，委托类型选择“云服务”，选择“RFS”，单击“下一步”。

图 3-4 创建委托

步骤4 在搜索框中输入“Tenant Administrator”权限，并勾选搜索结果，单击“下一步”。

图 3-5 选择策略

步骤5 选择“所有资源”，并单击“下一步”完成配置。

图 3-6 设置授权范围

步骤6 “委托”列表中出现“rf_admin_trust”委托则创建成功。

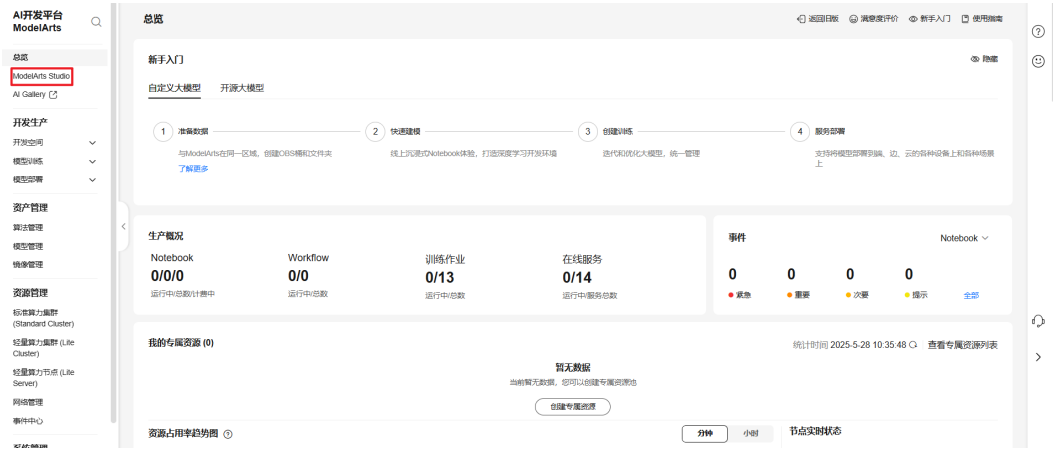
图 3-7 委托列表

----结束

与 MaaS 服务对接

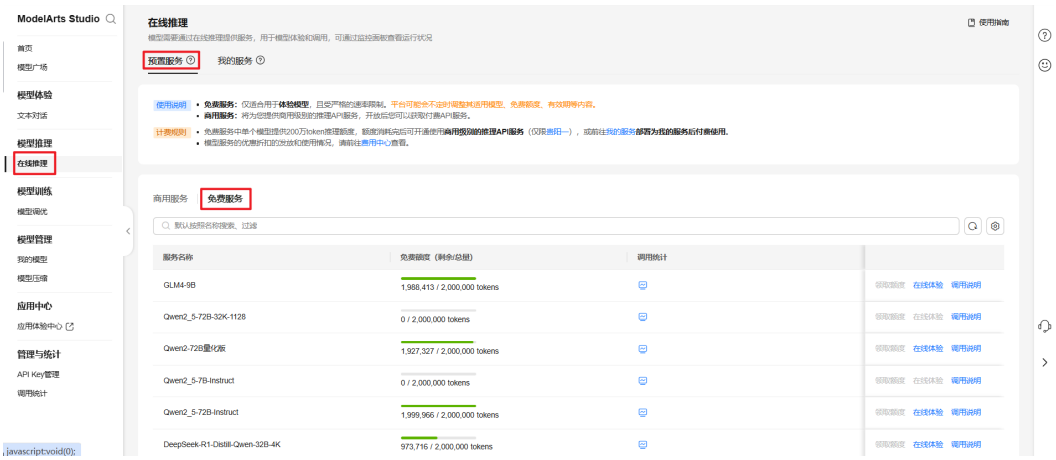
步骤1 登录**ModelArts Studio** 平台，本文以部署华东二的DeepSeek-R1-Distill-Qwen-32B-4K为例。

图 3-8 ModelArts Studio



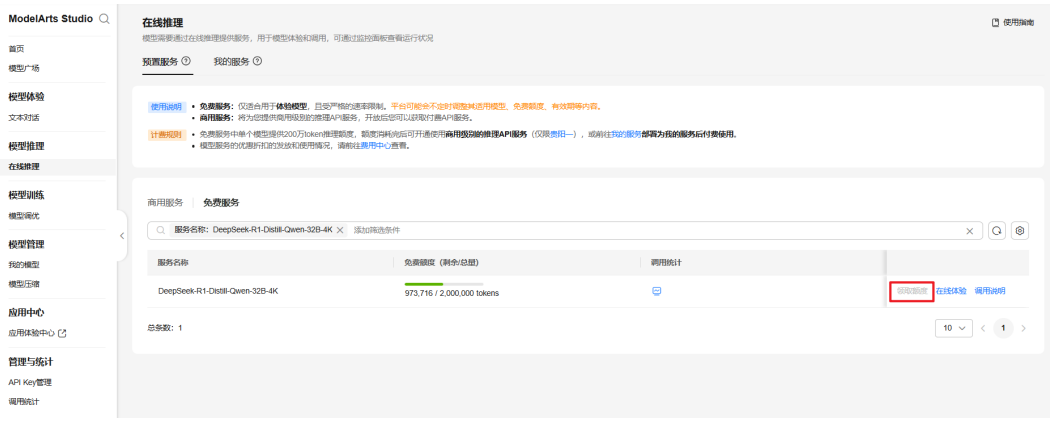
步骤2 在ModelArts Studio左侧导航栏中，选择“在线推理”进入“预置服务”服务列表，选择“免费服务”。

图 3-9 免费服务



步骤3 领取免费调用额度。在免费服务列表，选择所需的服务，单击右侧操作列的“领取额度”。当领取置灰时，表示该服务的免费额度已领取。

图 3-10 领取额度



步骤4 成功领取后，在免费服务列表，单击所需模型服务右侧的“调用说明”，在弹出的调用说明中接口类型选择“OpenAI SDK”获取API地址和模型名称。

注意：要选择OpenAI SDK后再复制API地址，末尾不要带chat/completion

图 3-11 选择模型服务

图 3-12 调用说明

步骤5 免费服务中单个模型提供200万token推理额度，额度消耗完后可开通使用商用级别的推理API服务（仅限**贵阳一**），或前往**我的服务**部署为我的服务后付费使用。

图 3-13 商用服务

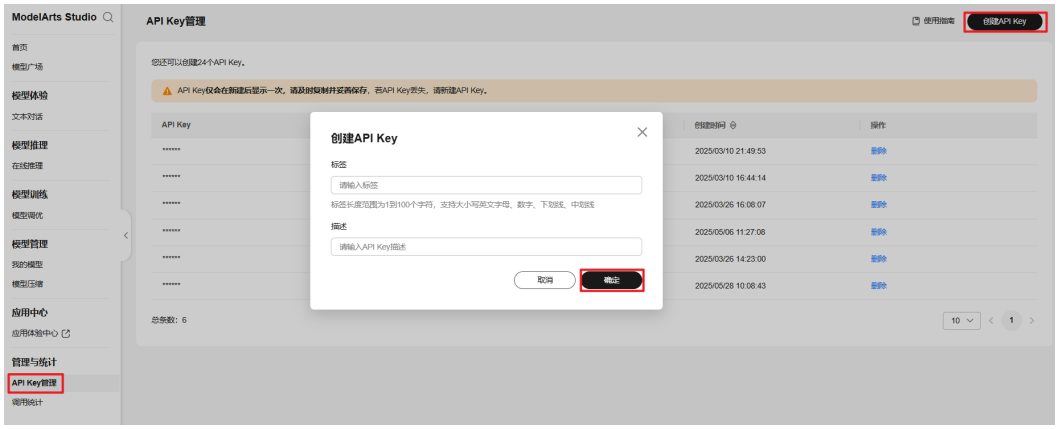
图 3-14 调用说明

步骤6 在调用MaaS部署的模型服务时，需要填写API Key用于接口的鉴权认证。在左侧导航栏，单击“API Key管理”（最多可创建30个密钥。**每个密钥仅在创建时显示一次，请确保妥善保存。如果密钥丢失，无法找回，需要重新创建API Key以获取新的访问密钥**）。

图 3-15 API Key 管理

步骤7 在“API Key管理”页面，单击右上角“创建API Key”，填写标签（自定义API Key的标签，标签具有唯一性，不可重复。仅支持大小写英文字母、数字、下划线、中划线，长度范围为1~100个字符）和描述（自定义API Key的描述，长度范围为1~100个字符）信息后，单击“确定”。标签和描述信息在创建完成后，不支持修改。注意复制并保存，以便后续步骤使用。

图 3-16 创建 API Key



----结束

3.2 一键部署（参数配置）

操作场景

本章节帮助用户高效地部署“数字人交互智能问答解决方案”解决方案。一键部署该解决方案时，参照本章节中的步骤和说明进行操作，即可完成快速部署。

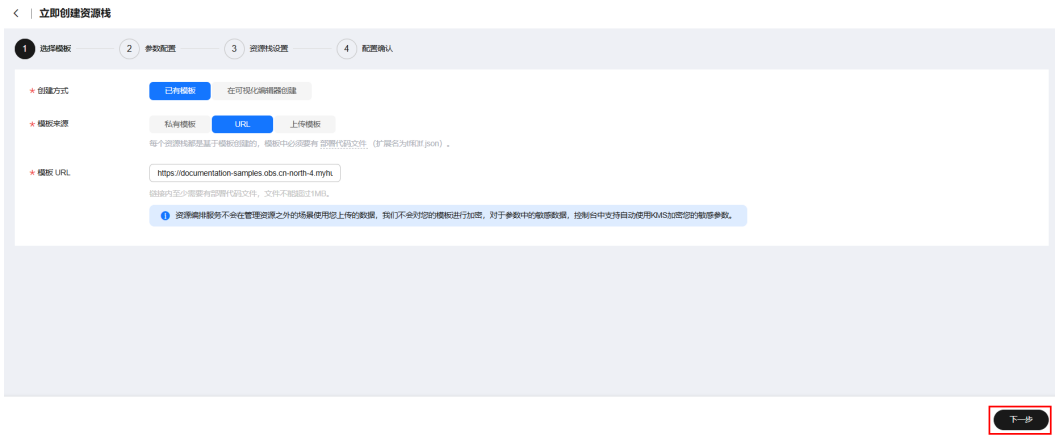
操作步骤

步骤1 登录华为云解决方案实践，选择“数字人交互智能问答解决方案”，单击“一键部署”，跳转至解决方案创建资源栈界面。

图 3-17 解决方案主页

步骤2 在选择模板界面中，单击“下一步”。

图 3-18 选择模板



步骤3 在配置参数界面中，参考“**表1 参数填写说明**”完成自定义参数填写，部分参数会自动默认填充参数值。如需修改请在参数配置页面删除文本框内的默认值后填写新的参数值，所有参数填写完成后方可单击“下一步”。

图 3-19 配置参数

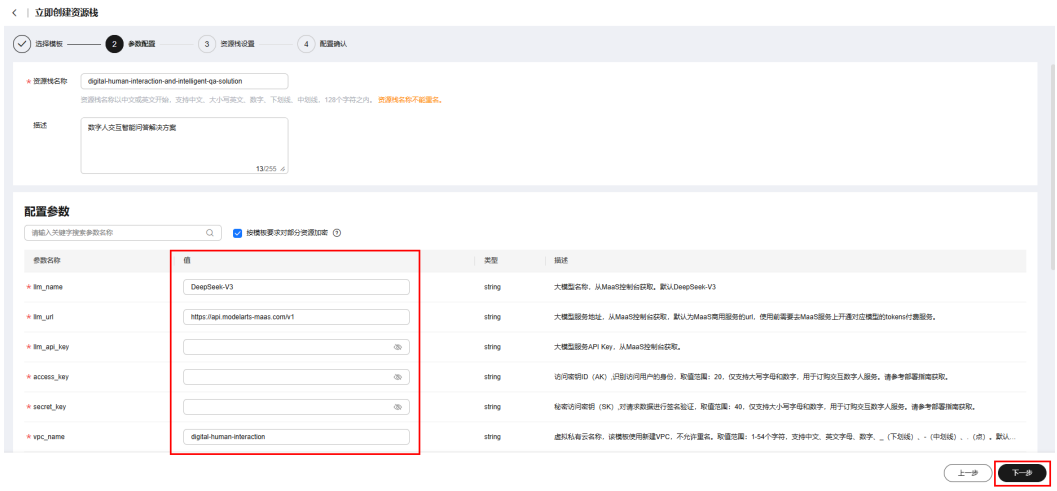


表 3-1 参数填写说明

参数名称	类型	是否可选	参数解释	默认值
llm_name	string	必填	大模型名称，从MaaS控制台获取，具体参考**与MaaS服务对接**。	DeepSeek-V3
llm_url	string	必填	大模型服务地址，从MaaS控制台获取，默认为MaaS商用服务的url，使用前需要去MaaS服务上开通对应模型的tokens付费服务，具体参考**与MaaS服务对接**。	https://api.modelarts-maas.com/v1

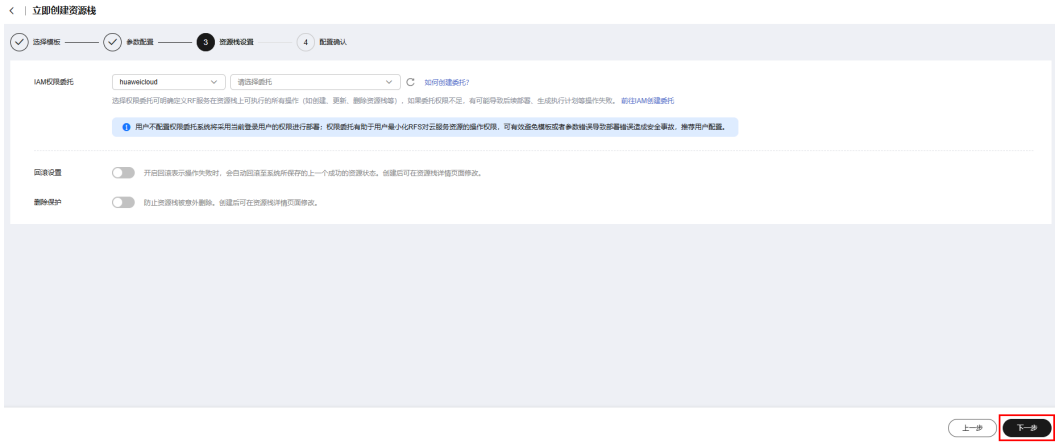
参数名称	类型	是否可选	参数解释	默认值
llm_api_key	string	必填	大模型服务API Key，从MaaS控制台获取，具体参考**与MaaS服务对接**。	空
access_key	string	必填	访问密钥ID（AK）,识别访问用户的身份，取值范围：20，仅支持大写字母和数字，用于订购数字内容生产线 MetaStudio 智能交互服务。详见**如何获取访问密钥AK/SK**。	空
secret_key	string	必填	秘密访问密钥（SK）,对请求数据进行签名验证，取值范围：40，仅支持大小写字母和数字，用于订购数字内容生产线 MetaStudio 智能交互服务。详见**如何获取访问密钥AK/SK**。	空
vpc_name	string	必填	虚拟私有云名称，该模板使用新建VPC，不允许重名。取值范围：1-54个字符，支持中文、英文字母、数字、_（下划线）、-（中划线）、.（点）。。	digital-human-interaction
security_group_name	string	必填	安全组名称，该模板新建安全组，安全组规则请参考**安全组规则修改（可选）**进行配置。取值范围：1-64个字符，支持数字、字母、中文、_(下划线)、-（中划线）、.（点）。	digital-human-interaction
ecs_name	string	必填	云服务器实例名称前缀，不支持重名。取值范围：1-58个字符，支持中文、英文字母、数字、_（下划线）、-（中划线）、.（点）。	digital-human-interaction
dify_ecs_flavor	string	必填	云服务器实例规格用于部署Dify，支持弹性云服务器 ECS及华为云Flexus 云服务器X实例。Flexus 云服务器X实例规格ID命名规则为x1.?u.?g，例如2vCPUs4GiB规格ID为x1.2u.4g，具体华为云Flexus云服务器X实例规格请参考控制台。弹性云服务器 ECS规格信息具体请参考官网**弹性云服务器规格清单**。	x1.8u.16g

参数名称	类型	是否可选	参数解释	默认值
model_ecs_flavor	string	必填	云服务器实例规格用于部署Embedding、Rerank，支持弹性云服务器 ECS及华为云Flexus云服务器X实例。Flexus 云服务器X实例规格ID命名规则为x1.?u.?g，例如2vCPUs4GiB规格ID为x1.2u.4g，具体华为云Flexus云服务器X实例规格请参考控制台。弹性云服务器 ECS规格信息具体请参考官网**弹性云服务器规格清单**。	x1e.16u.16g
ecs_password	string	必填	云服务器密码，长度为8-26位，密码至少必须包含大写字母、小写字母、数字和特殊字符（!@$%^-_=+[{}]:,./?）中的三种。修改密码，请参考**重置云服务器密码**登录ECS控制台修改密码。管理员账户默认root。	空
system_disk_size	number	必填	用于部署Dify的云服务器实例的系统盘大小，磁盘类型默认为高IO，单位：GB，取值范围为40-1,024，不支持缩盘。	100
bandwidth_size	number	必填	弹性公网带宽大小，该模板计费方式为按流量计费。单位：Mbit/s，取值范围：1-300Mbit/s。	300
ecs_charging_mode	string	必填	云服务器计费模式，默认自动扣费，可选值为：postPaid（按需计费）、prePaid（包年包月）。	postPaid
ecs_charging_unit	string	必填	云服务器订购周期类型，仅当charging_mode为prePaid（包年/包月）生效，此时该参数为必填参数。取值范围：month（月），year（年）。	month
ecs_charging_period	number	必填	云服务器订购周期，仅当charging_mode为prePaid（包年/包月）生效，此时该参数为必填参数。取值范围：charging_unit=month（周期类型为月）时，取值为1-9；charging_unit=year（周期类型为年）时，取值为1-3。	1

参数名称	类型	是否可选	参数解释	默认值
metastudio_charging_unit	string	必填	数字内容生产线 MetaStudio智能交互服务订购周期类型。取值范围：month（月），year（年）。	month
metastudio_charging_period	number	必填	数字内容生产线 MetaStudio智能交互服务订购周期。取值范围： metastudio_charging_unit=month（周期类型为月）时，取值为1-9； metastudio_charging_unit=year（周期类型为年）时，取值为1-3	1

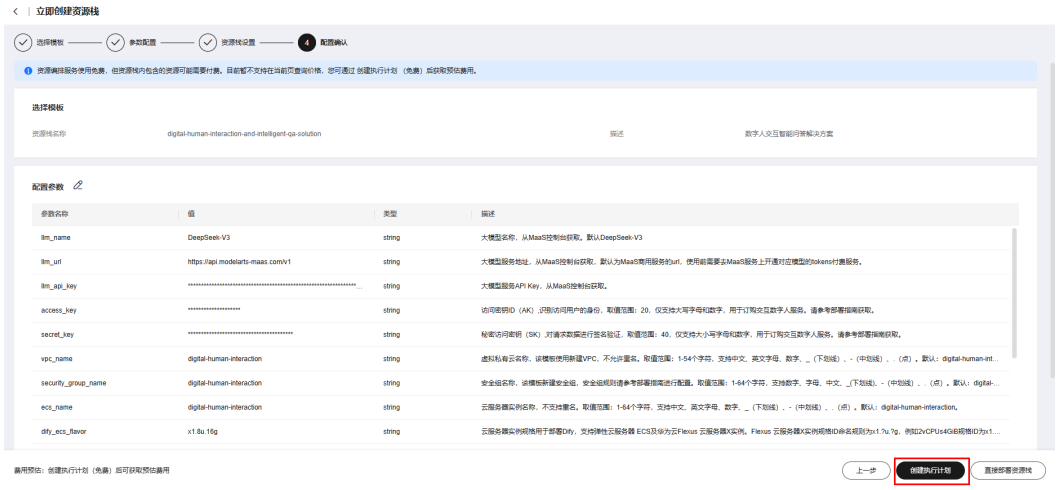
步骤4 （可选，如果使用华为主账号或admin用户组下的IAM子账户可不选委托）在资源设置界面中，在权限委托下拉框中选择“rf_admin_trust”委托，单击“下一步”。

图 3-20 资源栈设置



步骤5 在配置确认界面中，确认填写参数并单击“创建执行计划”。

图 3-21 配置确认



步骤6 在弹出的创建执行计划框中，自定义填写执行计划名称，单击“确定”。

图 3-22 创建执行计划

步骤7 单击“部署”，并且在弹出的执行计划确认框中单击“执行”。

图 3-23 执行计划

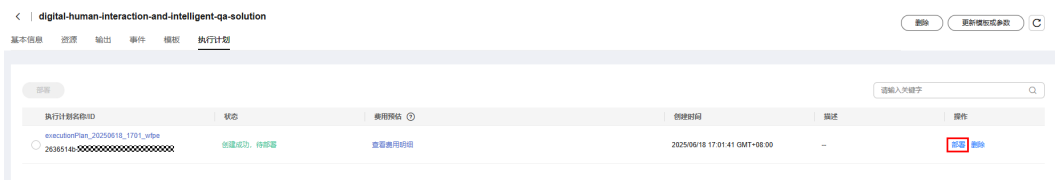
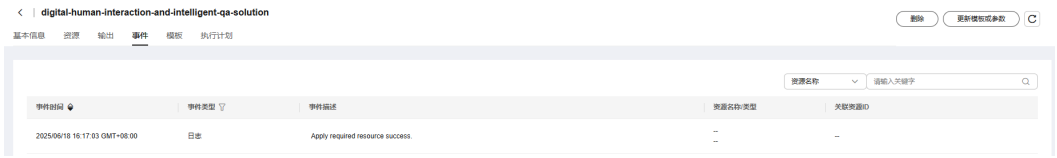


图 3-24 执行计划确认

步骤8 （可选）如果计费模式选择“包年包月”，在余额不充足的情况下（所需总费用请参考**表2-2**）请及时登录费用中心，手动完成待支付订单的费用支付。

步骤9 待“事件”中出现“Apply required resource success”，堆栈部署成功，表示顺利完成资源的下发和部署。堆栈部署成功后，脚本开始执行，耐心等待10分钟左右（受网络波动影响）。

图 3-25 部署完成



----结束

3.3 一键部署（快速选购）

本章节帮助用户高效地部署“数字人交互智能问答解决方案”解决方案。一键部署该解决方案时，参照本章节中的步骤和说明进行操作，即可完成快速部署。

步骤1 进入**DeepSeek应用专场**，选择“智能交互数字人”，单击“购买解决方案（单机版）”跳转至解决方案购买页面。

图 3-26 购买解决方案

步骤2 进入解决方案购买页面，设置基础配置。

图 3-27 基础配置

表 3-2 基础配置说明

参数	说明
计费模式	根据业务特点选择适用的计费模式。 ● 包年/包月是预付费模式，按购买周期计费，适用于可预估资源使用周期的场景，价格比按需计费模式更优惠。 ● 按需计费是后付费模式，按资源的实际使用时长计费，可以随时开通、删除。 **说明** 仅对解决方案中支持该计费模式的服务或资源生效，例如选择了包年/包月计费模式，弹性公网IP仍会按实际使用的流量计费。详情请查看具体服务的配置项。
区域	请就近选择靠近您业务的区域，可以降低网络时延，提高访问速度。创建后无法更换区域，请谨慎选择。
解决方案名称	系统自动生成，建议自定义为方便您识别的解决方案名称。 支持添加描述，可填写解决方案的更多相关信息。
虚拟私有云名称	系统自动新建VPC和子网。名称支持自定义、不支持重名。
安全组名称	系统自动新建安全组。名称支持自定义。 更多信息，请参见**安全组规则**。
购买时长	单次购买最短为1个月，最长为3年。

步骤3 设置1台Flexus云服务器X实例配置（应用节点），用于部署Dify-LLM应用开发平台。

图 3-28 Flexus 云服务器 X 实例配置

表 3-3 Flexus 云服务器 X 实例配置说明

参数	说明
规格	请根据业务需要选择合适的规格。单击 可调整CPU/内存配比、系统盘容量。
云服务器名称	系统自动创建所选规格的云服务器。名称支持自定义、不支持重名。
密码	设置云服务器密码。长度为8~26位，密码至少必须包含大写字母、小写字母、数字和特殊字符（!@$%^-_=+[{}]:,./?）中的三种。用户名默认为root。
购买数量	不支持修改，默认购买数量为1。

步骤4 设置1台Flexus云服务器X实例配置（模型节点），用于部署Embedding、Reranker模型。

图 3-29 Flexus 云服务器 X 实例配置

表 3-4 Flexus 云服务器 X 实例配置说明

参数	说明
规格	请根据业务需要选择合适的规格。单击 可调整CPU/内存配比。
购买数量	不支持修改，默认购买数量为1。

步骤5 设置MetaStudio服务配置。

图 3-30 MetaStudio 服务配置

表 3-5 MetaStudio 服务配置说明

参数	说明
产品规格	默认分身数字人智能交互：基于已有数字分身形象和声音，结合知识库和观众进行适配交互对话，清晰度1080p。
访问密钥ID（AK）	访问密钥ID（AK），识别访问用户的身份，长度为20位，仅支持大写字母和数字，用于订购数字内容生产线 MetaStudio 智能交互服务。请参见**获取访问密钥AK/SK**。
秘密访问密钥（SK）	秘密访问密钥（SK），对请求数据进行签名验证，长度为40位，仅支持大小写字母和数字，用于订购数字内容生产线 MetaStudio 智能交互服务。请参见**获取访问密钥AK/SK**。
模型服务	若未开通模型服务默认置灰，请先开通模型服务。请参见**在MaaS预置服务中开通商用服务**。 **注意** ModelArts Studio（MaaS）模型服务“在线推理-预置服务-商用服务/自定义接入点”，服务调用将产生费用。请参见**MaaS模型推理计费项**。
模型版本	选择模型服务的模型版本。

参数	说明
API地址	选择模型服务和模型版本后，自动获取所选模型的API地址。
API Key	调用ModelArts Studio（MaaS）模型服务时，需填写API Key用于接口的鉴权认证，保障服务访问的安全性和合法性。请参见**创建API Key**。
购买数量	不支持修改，默认购买数量为1。

步骤6 设置弹性公网 IP配置。

图 3-31 弹性公网 IP 配置

表 3-6 弹性公网 IP 配置说明

参数	说明
线路	默认全动态BGP：可根据设定的寻路协议实时自动优化调整网络结构，以保证客户网络的持续稳定和高效运行。
公网带宽	默认按流量计费：按照实际使用的流量来计费。
带宽大小	根据业务需要，选择所需的带宽大小。单位：Mbit/s，取值范围：1~300Mbit/s。
购买数量	不支持修改，默认购买数量为1。

步骤7 查看页面右侧的“配置概要”，确认解决方案配置详情。

图 3-32 配置概要

步骤8 单击“一键部署”，系统将通过自动支付完成扣款，请确保账户余额充足。

⚠ 注意

- 一键部署自动支付的金额不包含ModelArts Studio服务数字人交互的费用，这部分费用请您前往**费用中心**单独进行支付。具体请参见**开始使用步骤一**。
- 创建新账户或账户余额不足时，请进入“费用中心>**充值**”页面进行充值。具体请参见**账户充值**。
- 若选择了包年/包月计费模式，因账户余额不足导致自动支付失败，请进入“费用中心>**待支付订单**”页面，手动完成费用支付。

步骤9 一键部署后自动跳转至资源栈详情页面。

步骤10 待“事件”页面出现“Apply required resource success”，表示该解决方案已经部署完成。

图 3-33 部署完成

说明

请耐心等待5~10分钟左右（受网络波动影响），待应用下载成功后，即可输入网址访问Dify开发平台，具体请参见**开始使用**。

----结束

3.4 开始使用

该解决方案使用22端口CloudShell方式远程登录云服务器，默认已配置IP地址白名单，若需远程登录云服务器，可直接使用CLoudShell远程登录。

该解决方案部署成功后，环境初始化及应用安装预计10~20分钟不等，受网络、带宽影响，部署时间会有波动，部署完成方可正常使用。

安全组规则修改（可选）

安全组实际是网络流量访问策略，包括网络流量入方向规则和出方向规则，通过这些规则为安全组内具有相同保护需求并且相互信任的云服务器、云容器、云数据库等实例提供安全保护。

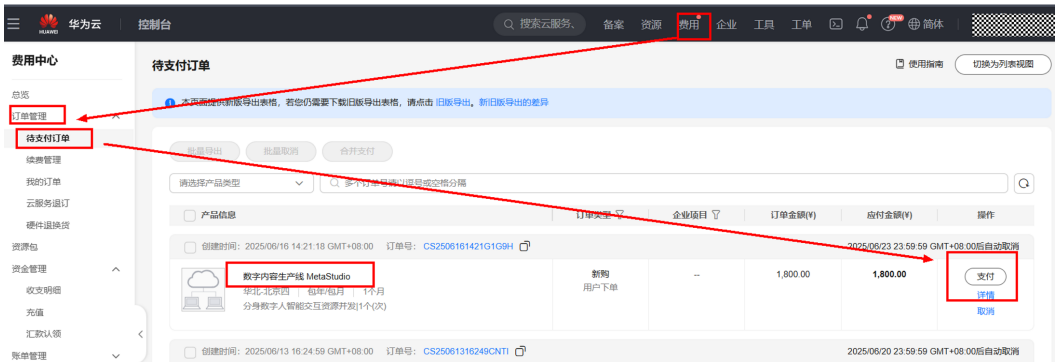
如果您的实例关联的安全组策略无法满足使用需求，比如需要添加、修改、删除某个TCP端口，请参考以下内容进行修改。

- 添加安全组规则：根据业务使用需求需要开放某个TCP端口，请参考**添加安全组规则**添加入方向规则，打开指定的TCP端口。
- 修改安全组规则：安全组规则设置不当会造成严重的安全隐患。您可以参考**修改安全组规则**，来修改安全组中不合理的规则，保证云服务器等实例的网络安全。
- 删除安全组规则：当安全组规则入方向、出方向源地址/目的地址有变化时，或者不需要开放某个端口时，您可以参考**删除安全组规则**进行安全组规则删除。

使用步骤

步骤1 **完成数字人智能交互服务订单：**访问**费用中心**支付自动创建好的数字人智能交互服务订单。

图 3-34 支付订单



步骤2 **登录Dify平台**：访问**云服务器控制台**，选择部署的后缀为“-dify”的云服务器，获取云服务器公网IP，浏览器访问地址http://{公网IP}登录Dify平台，邮箱默认：super@dify.com，密码默认：admin1234。

图 3-35 公网 IP

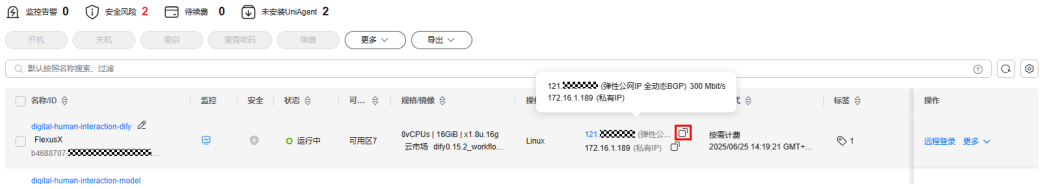


图 3-36 登录 Dify 平台

步骤3 **重置Dify用户密码**。

图 3-37 编辑账户

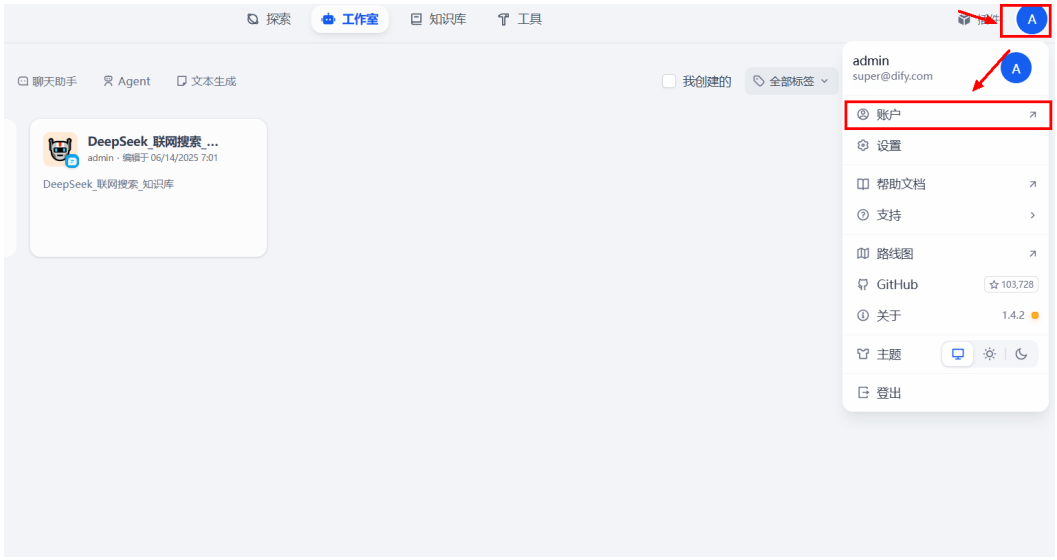


图 3-38 重置密码

步骤4 **上传知识库文档：**依次单击“知识库”，“默认知识库”；单击“添加文件”并上传文件，然后单击“下一步”进行创建。。

图 3-39 默认知识库

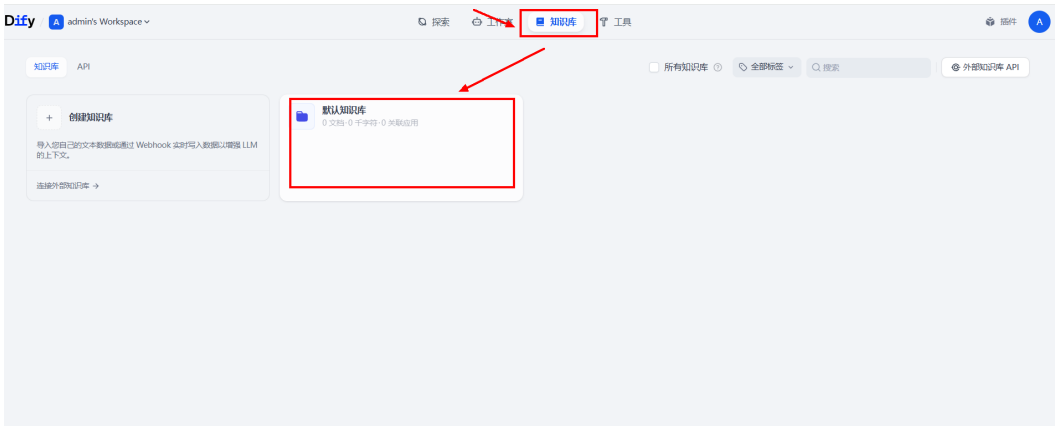


图 3-40 添加文件

图 3-41 上传文件

步骤5 **确认知识库参数并保存：**如图所示检查红框内的参数，完成后单击“保存并处理”，创建成功状态为“启用”。

图 3-42 确认参数

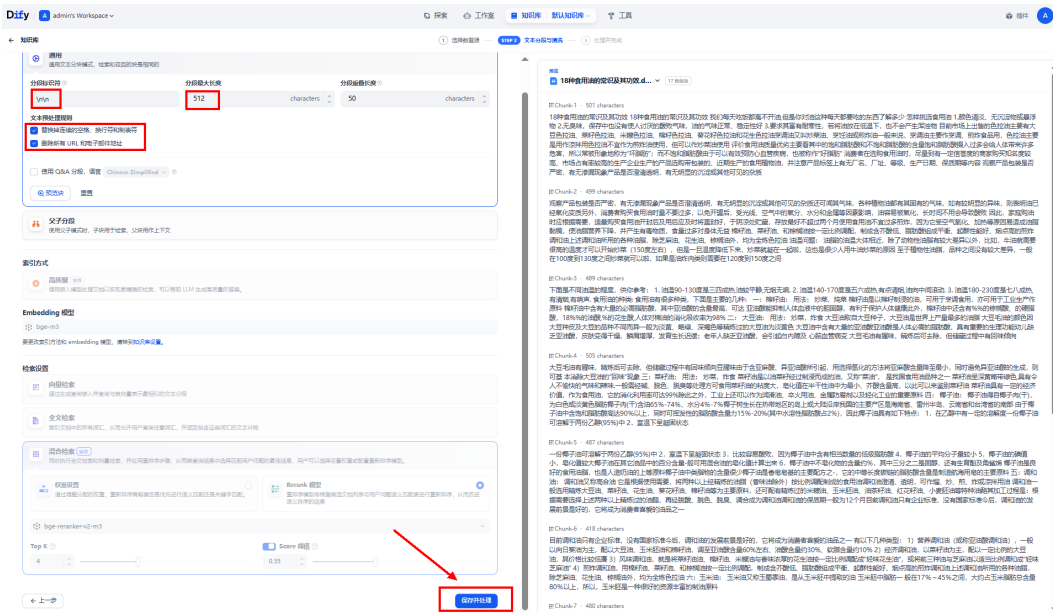


图 3-43 状态启用

步骤6 **测试并发布工作流：**在创建工作流并点单击“预览”按钮，在弹出的聊天框输入问题进行测试；之后在右上角单击“发布”按钮，再单击“发布”，即可完成工作流发布。

图 3-44 离线内容测试

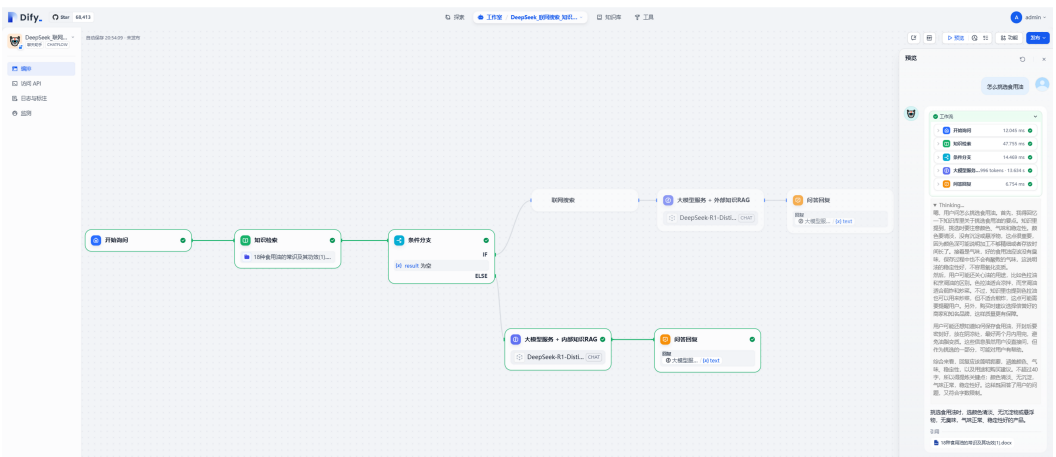


图 3-45 联网搜索测试

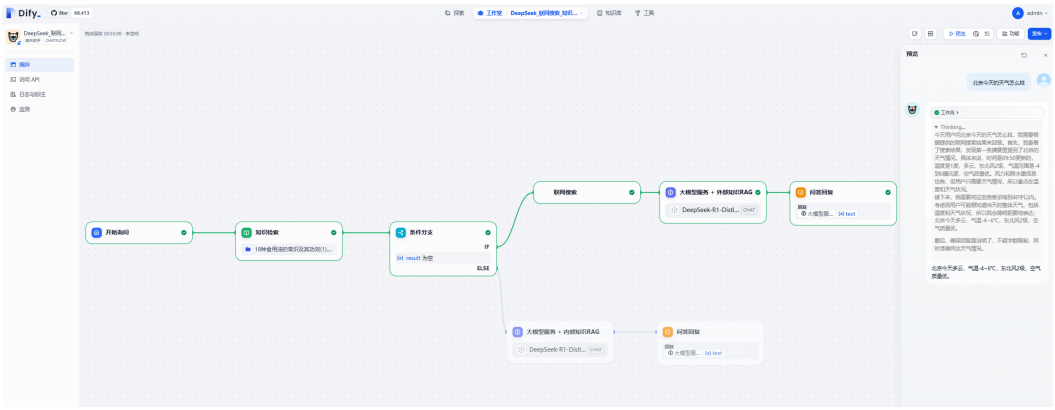
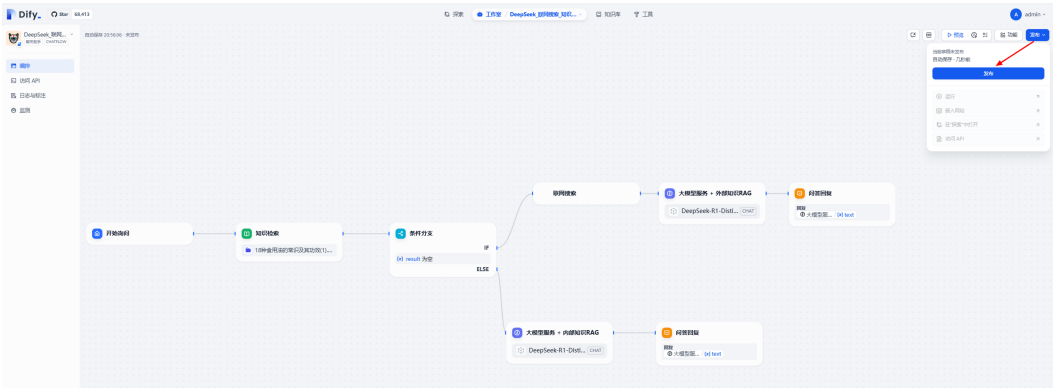


图 3-46 发布工作流



步骤7 **获取API密钥（API Key）：**访问工作流，单击左侧导航栏的“访问API”，单击右上角的“API密钥”，复制“API密钥”。

图 3-47 访问 API

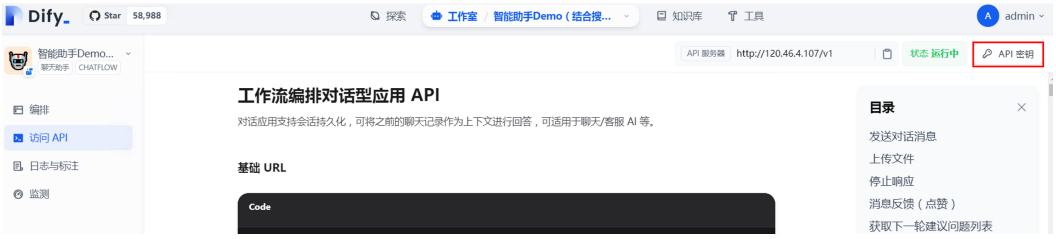
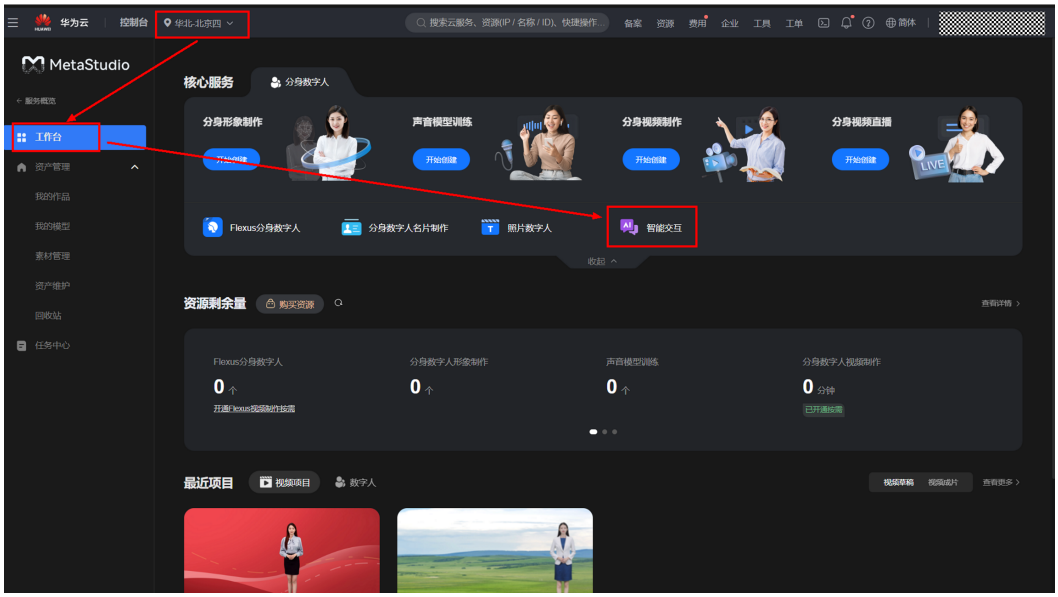


图 3-48 复制密钥

步骤8 **访问MetaStudio智能交互：**访问MetaStudio控制台，进入MetaStudio工作台，在MetaStudio工作台进入“智能交互”。

图 3-49 MetaStudio 控制台

图 3-50 MetaStudio 工作台



步骤9 **配置并发布智能交互数字人：**按图中所示依次填写配置，完成后单击右上角的“发布”。

第三方应用：第三方大脑（大模型）

应用名称：自定义

APPID：自定义

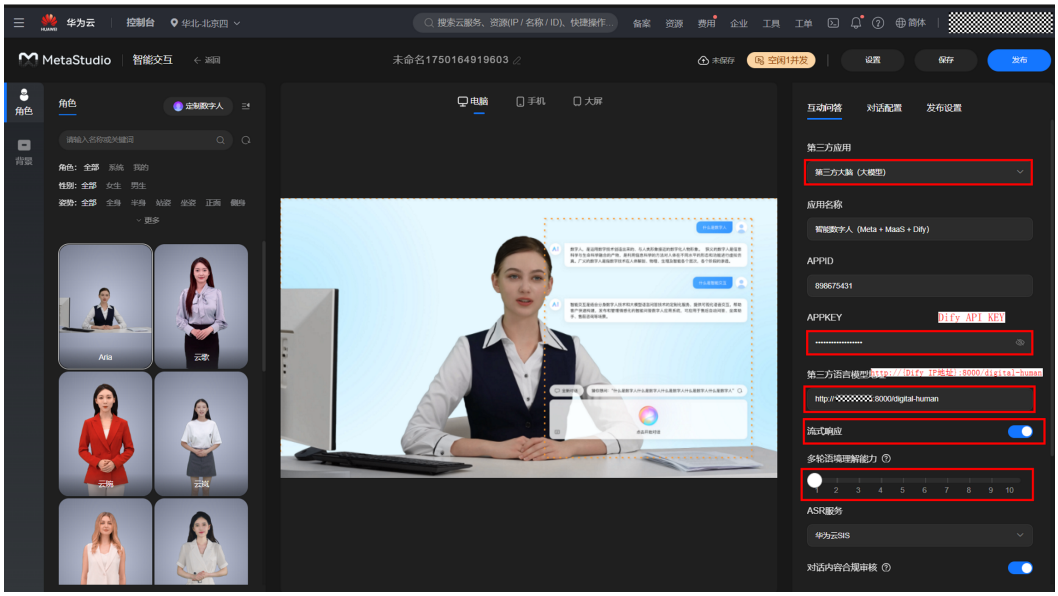
APPKEY：步骤10中获取的Dify API KEY

第三方语言模型地址（Dify IP地址从步骤2获取）：http://{Dify IP地址}:8000/digital-human

流式响应：开启（流式响应必须开启，否则影响交互响应速度）

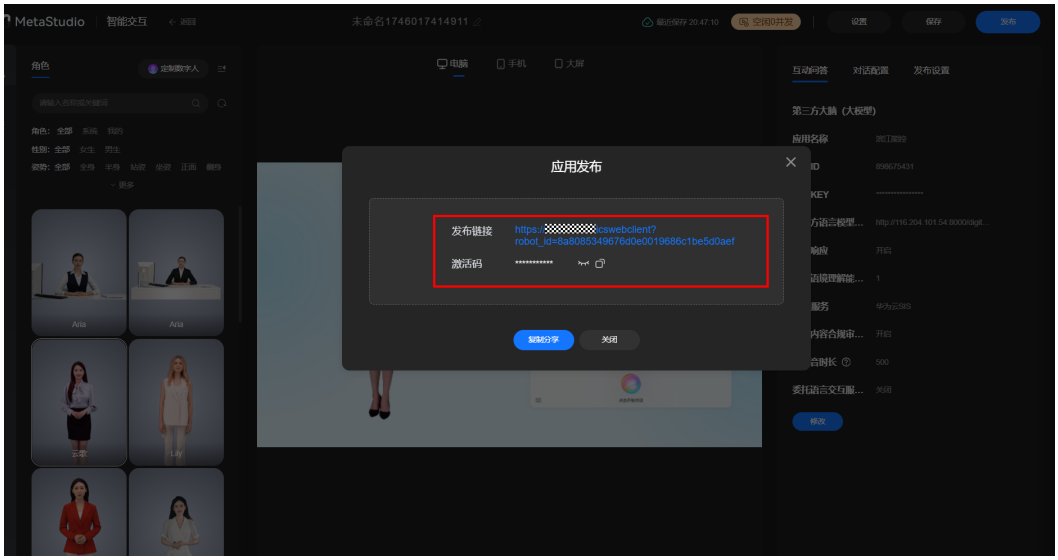
多轮对话语境理解能力：固定为1（方案已在后端自动实现多轮对话能力，此处必须固定为1即可，大于1会影响方案正常使用）

图 3-51 配置智能交互



步骤10 **使用智能交互数字人：**获取发布链接和激活码，访问链接填写激活码即可使用智能数字人交互服务。

图 3-52 链接和激活码



----结束

3.5 快速卸载

步骤1 登录**资源编排 RFS资源栈**，找到该解决方案创建的资源栈，单击资源栈名称右侧“删除”按钮。

图 3-53 一键卸载

步骤2 在弹出的删除资源栈确定框中，删除方式选择删除资源，输入Delete，单击右下角“确定”，即可卸载解决方案。

图 3-54 删除资源栈确认

步骤3 访问费用中心单击“云服务退订”，单击“分身数字人智能交互基础版”右侧的“退订资源”。

图 3-55 退订智能交互服务

----结束

4 附录

名词解释

- Flexus云服务器X实例：Flexus云服务器X实例是新一代面向中小企业和开发者打造的柔性算力云服务器。Flexus云服务器X实例功能接近ECS， 同时还具备独有特点，例如Flexus云服务器X实例具有更灵活的vCPU内存配比、支持热变配不中断业务变更规格、支持性能模式等。
- 弹性云服务器 ECS：是一种云上可随时自助获取、可弹性伸缩的计算服务，可帮助您打造安全、可靠、灵活、高效的应用环境。
- 虚拟私有云 VPC：是用户在华为云上申请的隔离的、私密的虚拟网络环境。用户可以基于VPC构建独立的云上网络空间，配合**弹性公网IP**、**云连接**、**云专线**等服务实现与Internet、云内私网、跨云私网互通，帮您打造可靠、稳定、高效的专属云上网络。
- 弹性公网IP EIP：提供独立的公网IP资源，包括公网IP地址与公网出口带宽服务。可以与弹性云服务器、裸金属服务器、虚拟IP、弹性负载均衡、NAT网关等资源灵活地绑定及解绑，提供访问公网和被公网访问能力。
- 数字内容生产线 MetaStudio：数字内容生产线，提供数字人视频制作、视频直播、智能交互、企业代言等多种服务能力，使能千行百业降本增效。

5 修订记录

表 5-1 修订记录

发布日期	修订记录
2025-06-18	第一次正式发布。