



西安交通大学
XI'AN JIAOTONG UNIVERSITY

西安交通大学人工智能先导计划之**师生素养提升计划**

AI赋能交晓智平台 助力**校园智慧管理** (黄金篇)



网络信息中心



2025年5月



“交晓智”黄金篇知识点清单

● 上节知识点

● 本次案例涉及的知识点

● 黄金篇总体知识点

大模型&提示词	知识库&数据库	插件	工作流	测试验证
大模型选型	文档解析与分段技巧	交晓智公共插件	工作流创建	工作流调试
大模型参数设置	知识库召回方式	绘图插件使用	选择器组件	智能体调试
提示词框架	数据库使用 (NL2SQL)	触发器使用	循环与终止循环组件	火焰图分析
提示词调优	问答库	自定义插件	代码组件	调试树
提示词模板	术语库	外接插件	多工作流协同	插件调试
...	...	...	...	...

“交晓智”平台助力校园管理业务

培训 要点

- 智能体 workflow 设计
- 知识库在工作流的应用
- 工作流节点设置
- 工作流调试

培训 收获

- 掌握智能体 workflow 的概念、架构及高效设计方法
- 熟练在工作流中集成知识库，提升系统功能
- 熟练使用调试方法和工具，快速解决工作流中的问题

智能体建设

阶段一
需求分析

阶段二
实现思路设计

阶段三
智能体开发

阶段四
测试验证

阶段五
提交发布

阶段一：需求分析-需求提出

➤ 背景

在青铜篇我们已经通过构建知识库、编写提示词、调用插件创建了一个入门级网信中心智能客服。但是在实际使用过程中遇到了以下2个问题：

1. **回答不准确问题：**目前的智能客服在回答某些问题时，回复的内容不完整，信息准确性不足。
2. **多语言翻译问题：**目前的智能客服无法满足留学生多种语言的沟通需求。



- **需求：实现多语言无障碍交流、增强回答的准确性和完整性。**

阶段一：需求分析

需求分析

1.明确赛道：智慧管理

2.明确目标：实现多语言的咨询服务，快速准确地解答师生的各种咨询问题。

3.明确优化方向：

(1) 增强回答的准确性和完整性

-解决回答某些问题时，回复的内容不完整，信息准确性不高的问题。

(2) 增加多语言咨询服务

-满足留学生的日常咨询服务。

阶段二：实现思路设计

(1) 增强回答的准确性和完整性

通过调整知识库相关参数，增强回答的准确性和完整性

(2) 增加多语言咨询服务

留学生用户提问



回复给用户

流程化的实现过程

阶段二：实现思路设计

通过纯提示词实现	通过工作流实现
难以处理复杂问题： 随着应用场景变得越来越复杂，仅靠调整提示词可能难以满足需求。	支持处理复杂问题： 适合于包含较多步骤、流程化、具备一定逻辑的复杂工作 将复杂任务分解为多个简单步骤，逐步处理，支持根据不同情况选择不同的处理路径。
输出稳定性差： 不同情况下使用的提示词可能会导致输出结果的一致性和稳定性较差。	输出稳定性高： 通过预设的步骤和规则，确保每次执行的结果一致，保证稳定性。
维护成本高： 随着项目规模的扩大，维护和管理大量的提示词会变得困难。	易于后期维护： 良好的模块化设计更便于后期维护。
灵活性高： 可以即时调整输入给AI模型的提示词，以适应不同的任务需求。	可扩展性强： 每个步骤独立开发和测试，易于扩展。
快速迭代： 无需编写复杂的代码逻辑，只需修改文本提示即可快速测试新的想法或功能。	迭代效率低： 每次更改都需要重新设计和测试整个工作流。

在步骤较多、流程化、具备一定逻辑的场景下，选择工作流方式实现

阶段二：实现思路设计

“交晓智”平台 实现思路

1.选择**大模型**——选择一个擅长多语言翻译、智能客服的大模型

2.调整**知识库参数**——增强回答的准确性和完整性

3.创建 **workflow**——增加多语言咨询服务

阶段三：智能体开发

步骤一：LLM选型：本案例为智能客服场景，选择模型deepseek-v3/moonshot-v1-8k

大模型名称	交晓智中模型版本	特点	适用场景
智谱清言 (GLM)	GLM-4-Air	轻量级模型，推理速度快，资源占用低	移动端应用、实时对话（如客服机器人）、简单问答、对响应速度要求高的场景。
	GLM-4-Long	擅长处理长文本，上下文窗口较大（可能达数万token）	长文档摘要（如论文、报告）、多轮对话记忆、连续剧情分析、 法律合同审查。
	GLM-4-Plus	功能增强版，支持复杂任务处理	代码生成、数学推理、跨模态任务（图文生成）、深度数据分析。
通义千问 (Qwen)	qwen-turbo	极致推理速度，低成本，适合高频调用	高频聊天场景（社交APP）、简单信息检索、轻量级对话（如智能硬件）。
	qwen-long	长上下文支持，平衡性能与效率	中等长度文本生成（如新闻稿）、长对话管理（教育辅导）、 基础代码补全
	qwen-plus	综合性能强，准确率与泛化能力提升	商业文案创作、复杂问题解答（如医疗咨询）、多语言翻译
	qwen-max	最大参数量版本，支持高复杂度任务，可能具备高级推理能力	科研分析、金融预测、多轮逻辑推理（如辩论）、创意内容生成（小说、剧本）
豆包 (Doubao)	Doubao-Pro-32k	专注长文本处理、响应快、综合能力较强	中长文本分析（企业报告）、知识库问答、多文档信息整合
DeepSeek	deepseek-v3	通用性强、性能均衡	企业级智能助手、实时翻译、搜索引擎增强、个性化推荐系统
	deepseek-r1	擅长复杂推理与逻辑能力，擅长数学、编程和自然语言推理任务，响应较慢	学术研究辅助、行业数据分析（金融/医疗）、精准推荐（电商、内容平台）
Kimi	moonshot-v1-8k	基础版，适合短文本，响应快	实时翻译、短对话交互、社交媒体评论生成
	moonshot-v1-32k	扩展上下文，支持长文本理解	技术文档解析、会议纪要整理、中等长度代码审查
	moonshot-v1-128k	超长上下文（约10万字），行业领先的文本容量	整本书籍摘要、长篇小说续写、复杂代码库维护（如GitHub项目）、学术论文复现

阶段三：智能体开发

步骤二：调整知识库参数：解决回复不准确、不完整的问题

知识库文件提供2种分段方式。

- (1) **自动分割**：平台自动设置分段规则与预处理规则，可能会出现回复的内容不完整、不准确。
- (2) **自定义分割**：自定义知识库文件的分段规则、分段长度以及预处理规则等，提高检索结果的完整性和准确性。

类型	说明	默认
分段标识	按照自定义分段标识将完整文本分割成多个分段。	\n
按版面自动切分	自动识别出标题、正文段落、图片和表格元素。使用此方法后将忽略分段标识。	默认勾选
分段长度	每个分段的文字长度上限，单位为字符数。	1000
内容处理规则	移除空白字符、启用OCR	默认不勾选
内容过滤规则	过滤链接、邮箱地址、参考资料	默认不勾选

知识库文件的2种分段方式

点击或拖拽文件到此处上传

支持 doc, docx, html, htm, txt, pdf, png, jpg, jpeg, ppt, pptx 格式文件
pdf 文件不超过 100 MiB, 图片文件不超过 10 MiB, html/htm/txt 文件不超过 30 MiB
其他单文件不超过 50 MiB
单次最多上传 300 个文件
文本类型编码格式要求 utf-8 否则可能解析成乱码

文件数 (1/1) 100%

2025春季学期选修课-金融大数据应用-董志龙-第1章-大数据技术及其在经济金融中的应用... 4.02 MiB ✓

分段方式 ①

自动分割 自动设置分段规则与预处理规则

自定义分割 自定义分段规则、分段长度以及预处理规则等

分段标识 * ①

#####... 32/32

定义自己手动插入的分隔符

按版面自动切分

分段长度 * ①

1000

支持长度范围: 200 - 64000 整数

内容处理规则 ①

☐ 移除空白字符

☒ 启用 OCR

识别图片中的内容

内容过滤规则 ①

☐ 过滤链接和邮箱地址

☐ 过滤参考资料

下一步

阶段三：智能体开发

扩展知识1：对知识库文件进行预处理，提升检索准确性

预处理类型	对检索准确性的影响	具体操作
重复数据处理	重复数据会导致检索结果冗余，干扰召回排序优先级。	1. 精确去重： 删除完全重复的条目 2. 模糊去重： 删除高度相似内容 3. 分组去重： 按知识类别分组后去重，避免跨类别误删
去除噪声与无效信息	噪声数据会引入无关关键词，导致检索时匹配错误； 逻辑矛盾可能使智能体返回矛盾答案，损害可信度。	1. 文本清洗： 删除乱码（如\x00、控制字符）、冗余空格、换行符；统一全角 / 半角标点（如将 “！” 转为 “！”） 2. 语义过滤： 移除广告、版权声明、重复模板 3. 逻辑修正： 修正明显矛盾（如 “登录密码长度 6-8 位” 与 “密码需至少 10 位” 冲突），标注待确认的模糊表述（如 “可能” “大概” 需补充具体条件）
术语统一化	术语不统一会导致漏检（如用户搜 “CPU” 但知识中只有 “中央处理器”）； 歧义未处理时，可能匹配到不相关领域的知识（如 “苹果” 匹配到水果而非公司）。	1. 同义词统一： 建立术语库（如 “客服” → “客户服务”、“CPU” → “中央处理器”），检索时自动扩展匹配 2. 歧义消除： 对多义词标注上下文（如 “端口” 标注为 “网络端口” 或 “硬件端口”） 3. 行业术语标准化： 按目标领域（如金融、医疗）统一专业术语（如 “贷记卡” 统一为 “信用卡”，避免同一概念多种表述）

阶段三：智能体开发

步骤三：创建 workflows-工作流概述

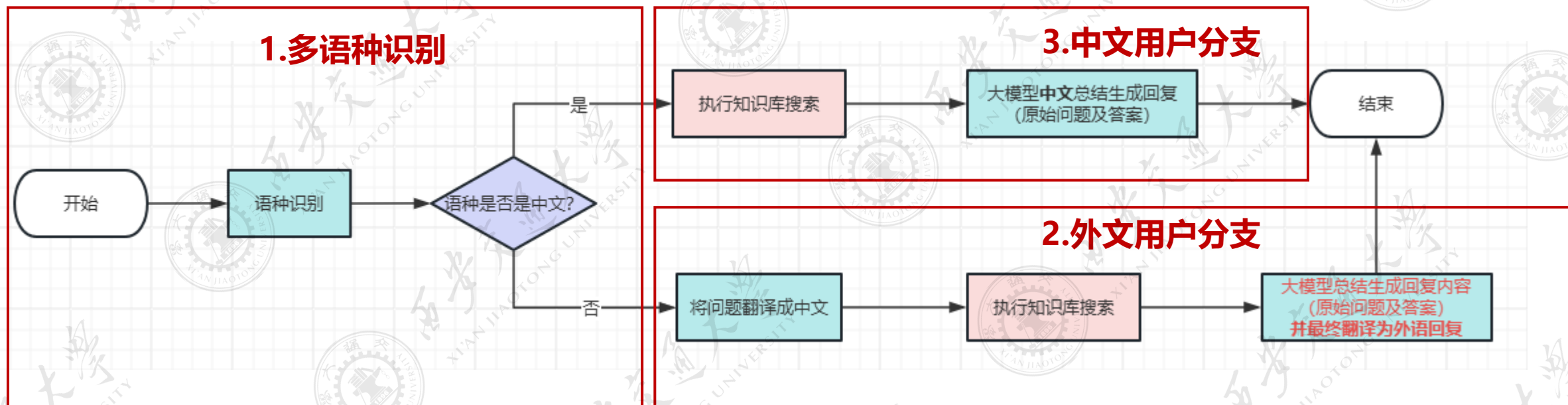
工作流由多个节点构成，节点是组成工作流的基本单元，包含大模型、代码、知识库、选择器、插件、工作流等**多种节点类型**。

节点类型	功能描述	应用场景	备注
Start节点	工作流起始节点，设定启动所需信息	接收用户输入或触发条件	默认存在于所有工作流中
End节点	工作流终止节点，返回最终结果	输出最终处理结果	默认存在于所有工作流中
大模型节点	调用大语言模型生成回复	- 文档生成/翻译/摘要 - 数据分析与预测 - 智能客服意图识别/自动回复 - 流程优化	支持单节点调试
知识库节点	根据相似度召回知识库分段，以List形式返回。	匹配用户问题并返回相关知识内容	支持单节点调试
选择器节点	按条件顺序匹配分支，执行符合条件的下游节点。类似if-else	问题分类分发（如故障排查转特定流程）	无
代码节点	编写代码处理输入变量，支持Python和JavaScript	- 数据清洗/分析 - 文件管理（如备份） - 系统集成	提供IDE编辑器，支持测试代码，支持单节点调试
插件节点	扩展智能体能力，连接外部系统或服务	调用外部API或工具（如绘图、联网搜索）	支持单节点调试
工作流节点	嵌套其他工作流实现模块化流程	拆分复杂任务为子流程（如订单处理拆分为支付、物流等子流程）	仅支持两级嵌套，被嵌套的工作流不可再嵌套，支持单节点调试

循环、终止循环、数据库、意图识别、智能体.....

阶段三：智能体开发

步骤三：创建工作流-总体流程设计



节点类型的选择:

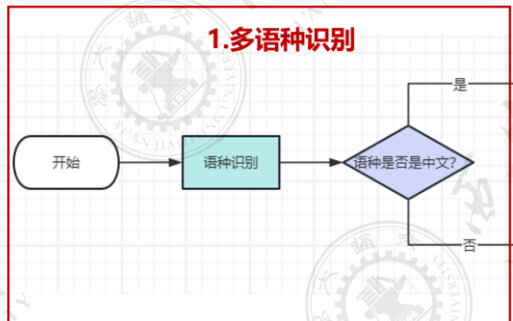
利用**大模型节点** (蓝绿色) 实现语种识别、翻译和生成回复

利用**知识库节点** (粉色) 实现知识库检索

利用**选择器节点** (紫色) 来实现中文用户分支和外文用户分支的分流

阶段三：智能体开发

步骤三：创建工作流：1.多语种识别



Start

工作流的起始节点，用于设定启动工作流需要的信息

输入 ①

变量名 * 变量类型 * 描述 *

变量名	变量类型	描述	必填
input	String	用户输入的	☒

语种识别

调用大语言模型，使用变量和提示词生成回复

模型

deepseek-v3

输入参数 ①

参数名	参数值
input	引用 Start/input

提示词 ① 从模板导入

输出值 ①

变量名	变量类型
raw_output	String
language	String

输出格式 ① JSON

IF 选择器01

连接多个下游分支，根据设定的条件按照顺序查找的方式来匹配运行的分支，如果匹配到某条件则只运行该...

所有条件

If ^

条件 AND

引用变量	选择条件	比较值
语种识别/language	等于	输入 中文

Else

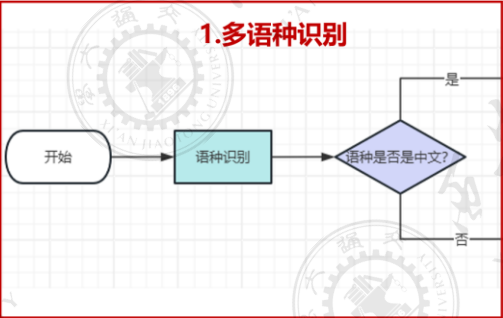
(1) Start节点
输入：用户输入的问题

(2) 大模型节点（语种识别）
输入：用户输入的问题
输出：识别的语种

(3) 选择器节点
判断条件：语种是否为中文

阶段三：智能体开发

步骤三：创建工作流：1.多语种识别-Start节点



Start

工作流的起始节点，用于设定启动工作流需要的信息

输入 ⓘ

变量名 *

变量类型 *

描述 *

必填

input

String

用户输入的问题

☒

⊖

步骤：

1.添加输入：添加input变量，接收用户输入的问题，变量类型String。

变量类型	含义
String	字符串
Integer	整数值，无小数点
Number	数值，可以是整数或浮点数
Object	键值对集合，用于存储结构化数据
Boolean	逻辑值，仅有两个可能值：true 或 false
Array<String>	字符串列表
.....	

阶段三：智能体开发

步骤三：创建工作流：1.多语种识别-大模型节点（语种识别）

步骤：

- 1.选择大模型：选择合适的大模型deepseek-v3
- 2.添加输入：添加用户输入的问题 input变量（引用Start节点的input）
- 3.添加输出：识别的语种language变量
- 4.提示词：编写提示词

第一步：分析以下文本是什么语言：{{input}}。

第二步：输出结果到{{language}} 变量中。

例如识别到语言是中文，则直接输出：中文

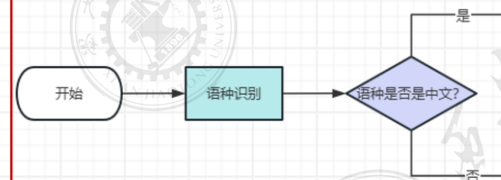
限制：

- 只专注于识别用户问题所使用的语种并输出到指定变量，拒绝回答与语种识别无关的话题。

注意事项：

- ✓ 通过连接线连接2个节点，才能传递参数
- ✓ 引用变量时，只能引用上游的变量
- ✓ 输出的变量需要向下游传递时，输出格式选择JSON，方便提取变量
- ✓ 提示词中引用变量时，使用 {{变量名}}

1.多语种识别



语种识别

调用大语言模型，使用变量和提示词生成回复

模型

deepseek-v3

输入参数 ① ^

参数名

参数值

input

引用

Start/input

提示词 ① ^

从模板导入

第一步：分析以下文本是什么语言：{{input}}。

第二步：输出结果到{{language}} 变量中。

例如识别到语言是中文，则直接输出：中文

限制：

- 只专注于识别用户问题所使用的语种并输出到指定变量，拒绝回答与语种识别无关的话题。

输出值 ① ^

输出格式 ① JSON ^

变量名

变量类型

raw_output

String

language

String

阶段三：智能体开发

步骤三：创建工作流：1.多语种识别-选择器节点

- 选择器可以连接多个下游分支，根据设定的条件按照**顺序查找**的方式来匹配运行的分支。
- 如果匹配到某条件则只运行该条件对应的分支，否则继续匹配下一条件直至结束

判断条件：语种识别是否中文

IF 选择器01

连接多个下游分支，根据设定的条件按照顺序查找的方式来匹配运行的分支，如果匹配到某条件则只运行该...

所有条件

IF ^

条件

AND

引用变量 选择条件 比较值

语种识别/language 等于 输入 中文

Else

IF 选择器01

连接多个下游分支，根据设定的条件按照顺序查找的方式来匹配运行的分支，如果匹配到某条件则只运行该...

所有条件

IF ^

条件

AND

引用变量 选择条件 比较值

语种识别/language 等于 输入 中文

Else If ^

条件

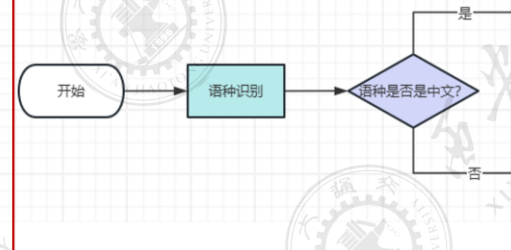
AND

引用变量 选择条件 比较值

请选择 等于 引用 请选择

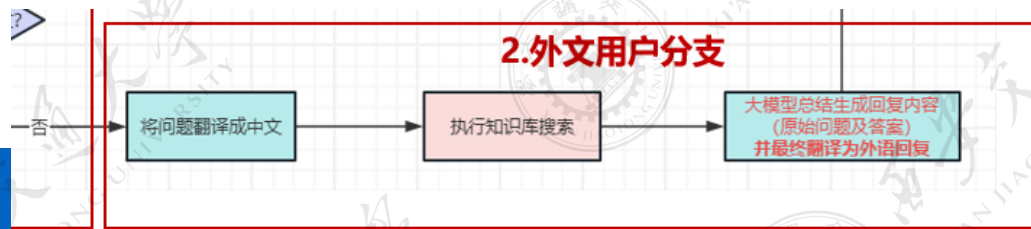
Else

1.多语种识别



阶段三：智能体开发

步骤三：创建工作流：2.外文用户分支



外文翻译

调用大语言模型，使用变量和提示词生成回复

模型：deepseek-v3

输入参数：input

提示词：第一步：接收用户输入((input))的多语言内容，运用专业的翻译能力，将其精准翻译为中文。第二步：仅输出翻译结果，输出到((chineseQuestion))变量中，不包含多余解释。## 限制：- 仅专注于将用户输入的多语言内容翻译为中文，不提供其他语言之间的翻译或与翻译无关的服务。- 翻译结果必须符合中文的表达习惯和语法规则。- 输出内容简洁明了，避免出现模糊或歧义的表达。

输出值：chineseQuestion (String)

知识库02

在选定的知识库中，根据设置的召回分段数量和要求相似度设置...

输入参数：query

知识库：xjtunic系统安全审查

检索方式：向量检索

召回分段数：5

相似度设置：0.44

输出值：knowledgeList (List)

结果总结并翻译为外文

调用大语言模型，使用变量和提示词生成回复

模型：deepseek-v3

输入参数：language, answer, question

提示词：第一步：拿到用户问题((question))和答案((answer))后，进行有条理的总结。第二步：将总结内容准确翻译为目标语言((language))，仅输出翻译后的结果。对其中的时间、地点、联系方式等关键信息，要进行加粗显示。## 限制：- 仅围绕结合原始问题总结回答相关内容和翻译结果进行操作，不涉及其他无关话题。- 总结回答需逻辑清晰、表达准确，符合正常语言表达习惯，不要起主标题。

输出值：response (String)

(1) 大模型节点 (外文翻译)

输入：用户输入的问题
输出：用户输入的问题 (中文版)

(2) 知识库节点

输入：用户输入的问题 (中文版)
输出：检索到的知识列表

(3) 大模型节点 (结果总结并翻译为外文)

输入：用户输入的问题 (中文版)、检索到的知识列表、识别的语种
输出：回复给用户的内容 (外文版)

阶段三：智能体开发

步骤三：创建工作流：2.外文用户分支-大模型节点（外文翻译）

步骤：

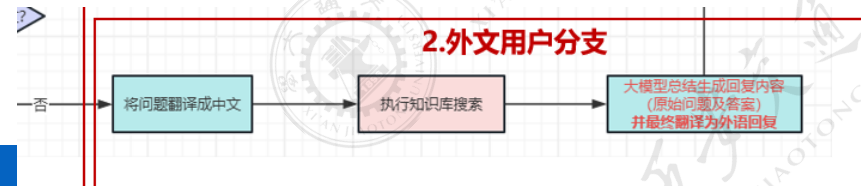
- 1.选择大模型：选择合适的大模型deepseek-v3
- 2.输入：用户输入的问题 **input**
- 3.输出：用户输入的问题（中文版） **chineseQuestion**
- 4.提示词：编写提示词

第一步：接收用户输入{{input}}的多语言内容，运用专业的翻译能力，将其精准翻译为中文。

第二步：仅输出翻译结果，输出到{{chineseQuestion}}变量中，不包含多余解释。

限制：

- 仅专注于将用户输入的多语言内容翻译为中文，不提供其他语言之间的翻译或与翻译无关的服务。
- 翻译结果必须符合中文的表达习惯和语法规则。
- 输出内容简洁明了，避免出现模糊或歧义的表述。



外文翻译

调用大语言模型，使用变量和提示词生成回复

模型

deepseek-v3

输入参数 ① ^

参数名

input

参数值

引用

Start/input

提示词 ① ^

从模板导入

第一步：接收用户输入{{input}}的多语言内容，运用专业的翻译能力，将其精准翻译为中文。

第二步：仅输出翻译结果，输出到{{chineseQuestion}}变量中，不包含多余解释。

限制：

- 仅专注于将用户输入的多语言内容翻译为中文，不提供其他语言之间的翻译或与翻译无关的服务。
- 翻译结果必须符合中文的表达习惯和语法规则。
- 输出内容简洁明了，避免出现模糊或歧义的表述。

输出值 ① ^

输出格式① JSON ^

变量名

raw_output

变量类型

String

chineseQuesti...

String

阶段三：智能体开发

步骤三：创建工作流：2.外文用户分支-知识库节点

步骤：

- 1.输入：引用上个节点输出的chineseQuestion变量
- 2.输出：无需单独设置，知识库节点会将检索到的内容，封装成List输出
- 3.添加知识库：添加已经创建好的知识库，设置知识库召回参数

知识库 ① ^

- xjtunic系统安全审查 标准类型
- 网信服务知识库分类 标准类型
- xjtunic网站内容 标准类型
- xjtunic服务手册 标准类型
- xjtunic问答规章制度 标准类型
- xjtunic问答知识库 标准类型

检索方式 ① 向量检索

召回分段数 ① 5

相似度设置 ① 0.44

知识增强 ①

知识库02

在选定的知识中，根据设置的召回分段数数量和要求的相似度设置值...

输入参数 ① ^

参数名	参数值
query * ①	引用 外文翻译/chin...

知识库 ① ^

- xjtunic系统安全审查 标准类型
- 网信服务知识库分类 标准类型
- xjtunic网站内容 标准类型
- xjtunic服务手册 标准类型
- xjtunic问答规章制度 标准类型
- xjtunic问答知识库 标准类型

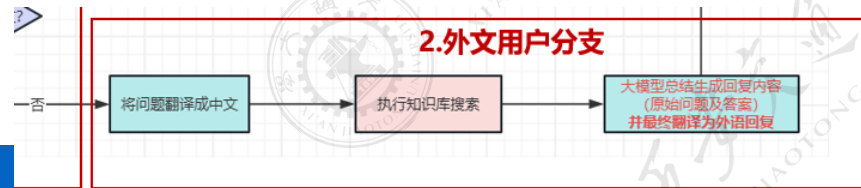
检索方式 ① 向量检索

召回分段数 ① 5

相似度设置 ① 0.44

知识增强 ①

输出值 ① v



阶段三：智能体开发

扩展知识2：知识库召回参数设置

插件名称	功能	本案例设置值
检索方式	1.向量检索：把用户问题向量化后，和知识库语义匹配，计算向量之间的相似度（余弦距离、欧氏距离等） 2.全文检索：把用户问题分词后，和知识库做关键词匹配（BM25算法） 3.混合检索：针对用户问题分别做向量检索和全文检索，将结果做一次排序后作为最终结果（选择混合检索时候需要用到Rerank模型，减轻向量检索和全文检索因相似度计算算法不同导致的打分不均衡的情况）	向量检索
召回分段数	从检索结果中选取相似度较高的分段数，值越小越精确，但可能会遗漏内容， 建议在3-5之间	5
相似度设置	相似度设置用于过滤不相干的分段内容，值越大越精确，但可能导致检索不到目标知识， 建议设置在0.4-0.6之间	0.5
知识增强	召回分段时，会把这些分段的上一个和下一个分段都带上，增加了内容完整性，但比较消耗大模型输入token。	不开启

阶段三：智能体开发



步骤三：创建工作流：2.外文用户分支的编排-大模型节点（结果总结并翻译为外文）

步骤：

- 1.选择大模型：选择合适的大模型deepseek-v3
- 2.输入：识别的语种language、用户输入的问题（中文版）question、检索到的知识列表answer
- 3.输出：回复给用户的内容（外文版）
- 4.提示词：编写提示词

第一步：拿到用户问题{{question}}和答案{{answer}}后，进行有条理的总结。

第二步：将总结内容准确翻译为目标语言{{language}}，仅输出翻译后的结果。对其中的时间、地点、联系方式等关键信息，要进行加粗显示。

限制：

- 仅围绕结合原始问题总结回答相关内容和翻译结果进行操作，不涉及其他无关话题。
- 总结回答需逻辑清晰、表达准确，符合正常语言表达习惯，不要起主标题。

结果总结并翻译为外文

调用大语言模型，使用变量和提示词生成回复。

模型

deepseek-v3

输入参数

参数名	参数值
language	引用 语种识别/lan...
answer	引用 知识库02/ou...
question	引用 外文翻译/chi...

提示词

从模板导入

第一步：拿到用户问题{{question}}和答案{{answer}}后，进行有条理的总结。

第二步：将总结内容准确翻译为目标语言{{language}}，仅输出翻译后的结果。对其中的时间、地点、联系方式等关键信息，要进行加粗显示。

限制：

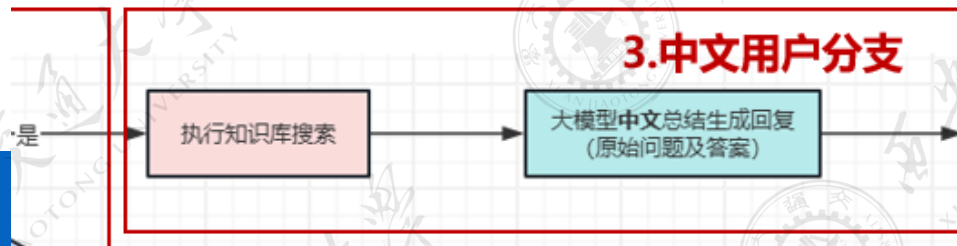
- 仅围绕结合原始问题总结回答相关内容和翻译结果进行操作，不涉及其他无关话题。
- 总结回答需逻辑清晰、表达准确，符合正常语言表达习惯，不要起主标题。

输出值

输出格式 文本

阶段三：智能体开发

步骤三：创建工作流：3.中文用户分支



知识库节点

- 1.输入：用户输入的问题
- 2.输出：检索到的知识列表
- 3.添加知识库：添加已经创建好的知识库，设置知识库召回参数

知识库01

在选定的知识库中，根据设置参数来召回最匹配的分段，并返回List。

输入参数

参数名	参数值
query *	引用 Start/input

知识库

- xjtunic系统安全审查 标准类型
- 网信服务知识库分类 标准类型
- xjtunic网站内容 标准类型
- xjtunic服务手册 标准类型
- xjtunic问答规章制度 标准类型
- xjtunic问答知识库 标准类型

检索方式 向量检索

召回分段数 5

相似度设置 0.44

知识增强 关闭

输出值

总结中文

调用大语言模型，使用变量和提示词生成回复。

模型

deepseek-v3

输入参数

参数名	参数值
answer	引用 知识库01/ou...
question	引用 Start/input

提示词

从模板导入

第一步：接收用户问题({{question}})和答案({{answer}})。
第二步：对问题和答案进行分析、提炼，提取关键信息。
用流畅自然的语言进行总结回答，确保内容完整、准确，
对其中的时间、地点、联系方式等关键信息，要进行加粗显示。

限制:

- 所输出的内容必须符合正常的语言表达习惯，逻辑清晰。
- 总结回答应简洁明了，避免冗长复杂的表述。

输出值

输出格式 文本

变量名

变量类型

raw_output

String

大模型节点

- 1.输入：用户输入的问题、检索到的知识列表
- 2.输出：回复给用户的内容
- 3.提示词：编写提示词

阶段三：智能体开发

步骤三：创建工作流：将两个分支连接到End节点

步骤：

1.设置回复模式：

使用设定的内容直接回答：可以自己设定回复内容
返回变量由Bot生成回答：由外层调用该工作流的智能体进行最终总结回答。

2.输出：中文用户分支和外文用户分支中的任一回复。

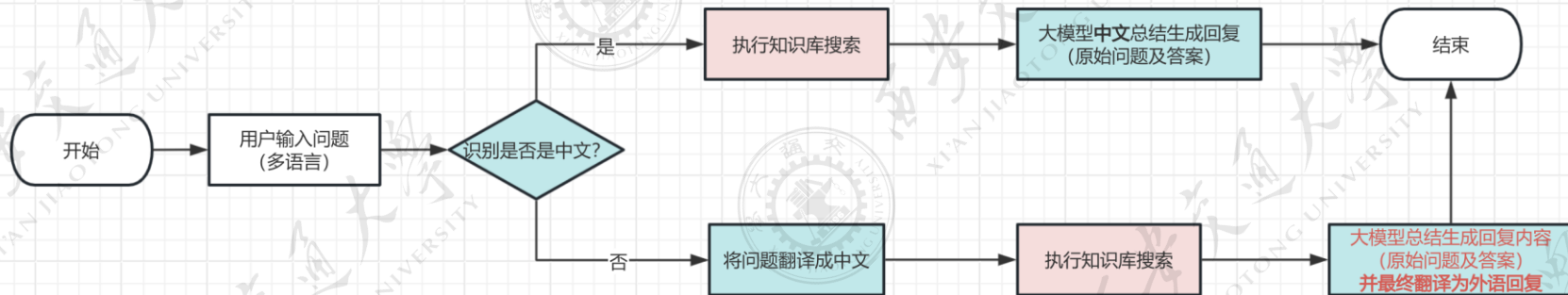
3.开启流式输出：

开启时，回复内容将会逐字流式输出；关闭时，回复内容将全部生成后一次性输出。

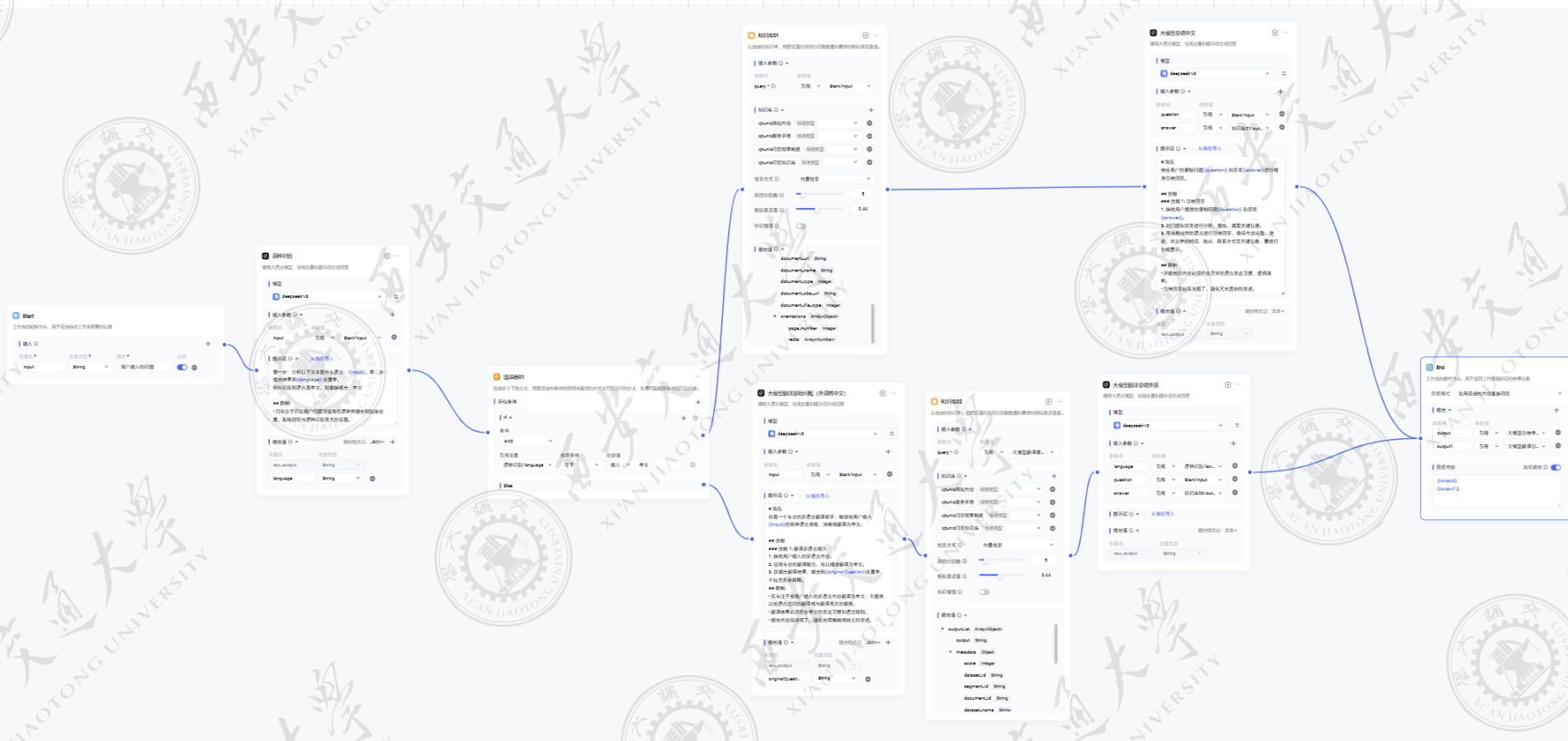


阶段三：智能体开发

流程设计图



工作流实现



阶段三：智能体开发

步骤四：调试 workflow——工作流调试

点击试运

网信中心，客服多语言工作流

15145 ms 1245 Tokens

隐藏上一次运行结果

复制

试运行

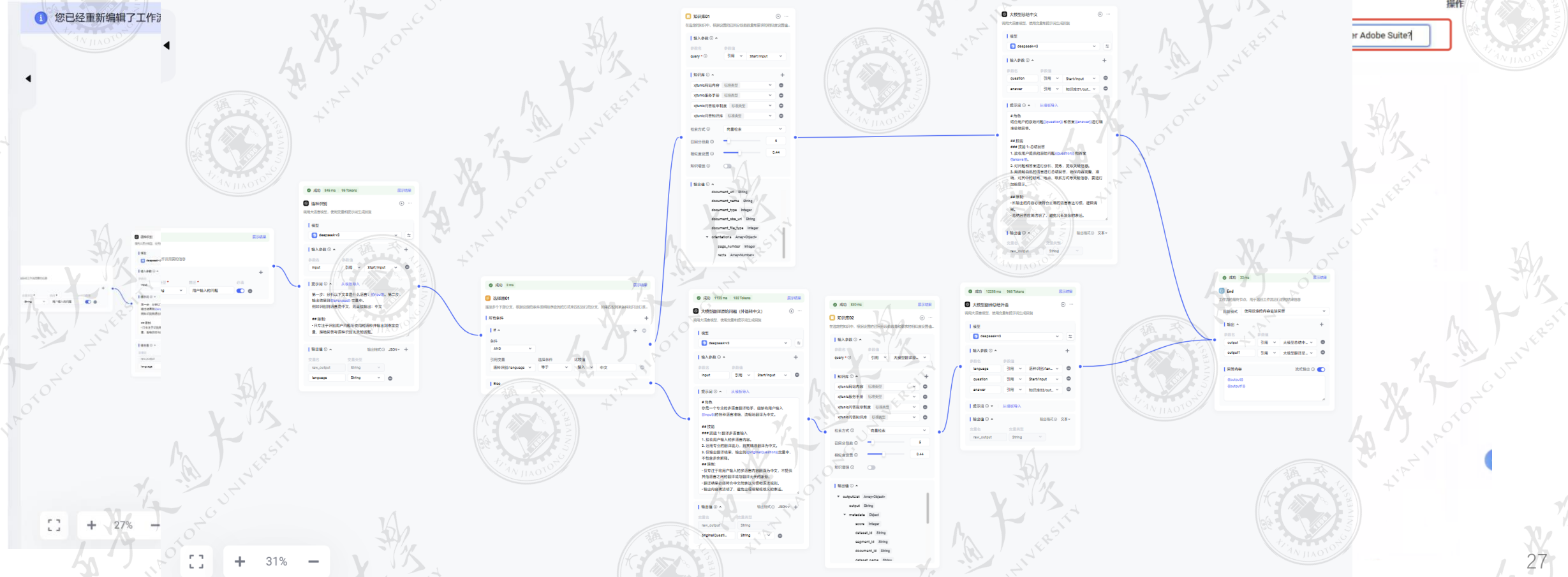
发布

流

网信中心，客服多语言工

您已经重新编辑了工作流，在再次发布之前，智能体中引用的工作流仍然是最后发布的版本。

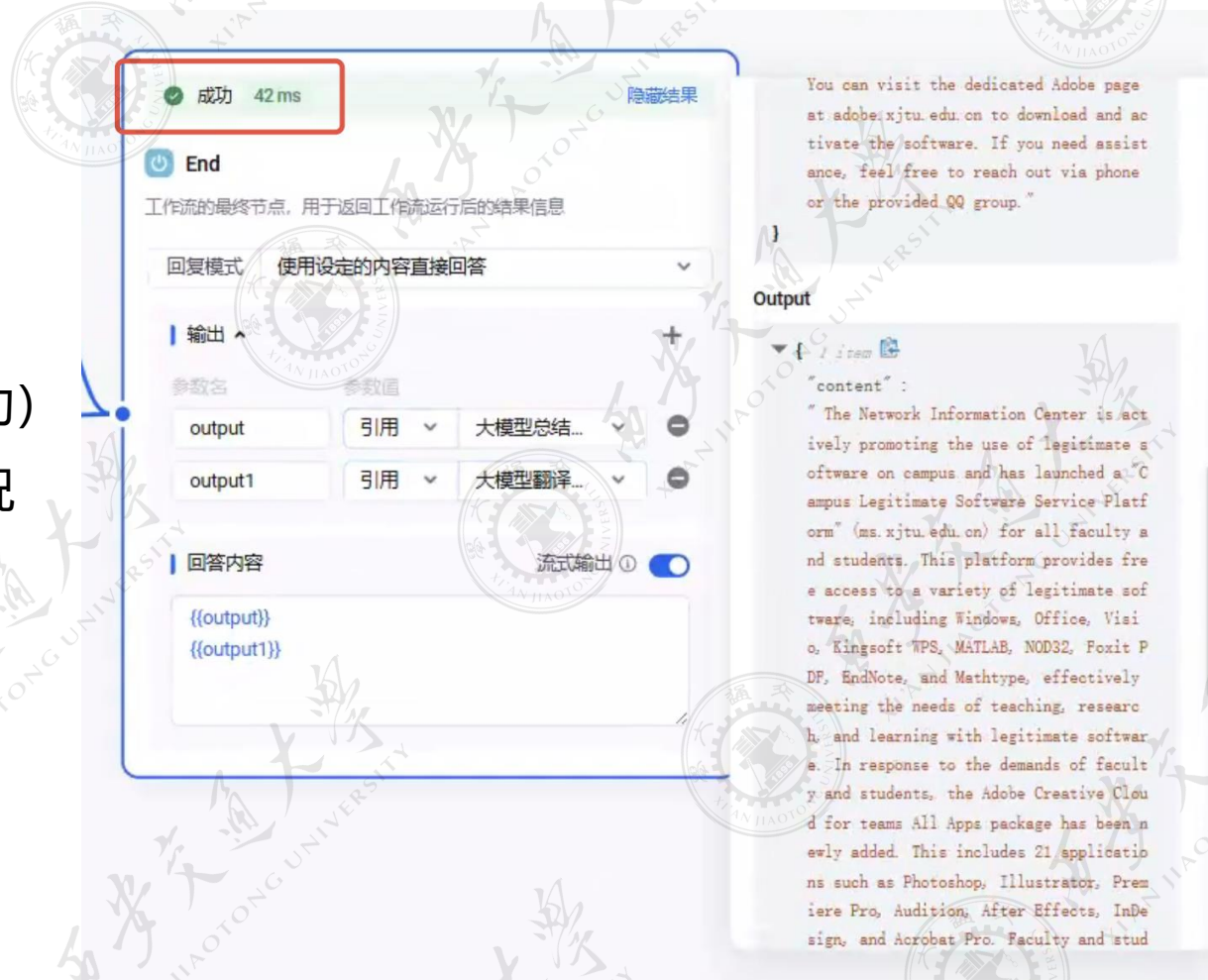
您已经重新编辑了工作流



阶段三：智能体开发

步骤四：调试 workflow——workflow 调试

观察 workflow 各节点执行是否是成功（绿色表示成功）
展示结果：可查看每个节点的输入输出及耗时情况



阶段三：智能体开发

步骤四：调试 workflow——单节点调试运行

节点有 “▶” 按钮表示该节点支持单节点调试。

- 点击 “▶” 按钮，输入该节点的输入参数，进行单节点的调试运行，运行后可查看单节点的耗时、输入和输出



workflow整体调试完成后，需要发布 workflow，发布成功后智能体才能引用

阶段三：智能体开发

创建 workflow 小结

创建工作流的三个环节

流程设计	定义流程：将整个任务分解为多个步骤，明确每个步骤的输入和输出，明确各个步骤的逻辑关系。 设计错误处理机制（按需）：确保在某个步骤失败时有备用方案。
流程搭建	利用平台提供的多种类型节点，搭建工作流。
测试优化	单节点测试：对每个节点进行单独测试，确保其正确性。 工作流测试：测试整个工作流的连贯性和稳定性。

阶段三：智能体开发

步骤五：升级智能体

1. **大模型参数设置**：思维方式选择**function call**，将单次回复最大token数设置成4096，避免输出内容过多被截断。
2. **添加 workflow**：在智能体的技能区域添加 workflow。
3. **编写提示词**：明确调用 workflow 来回答。



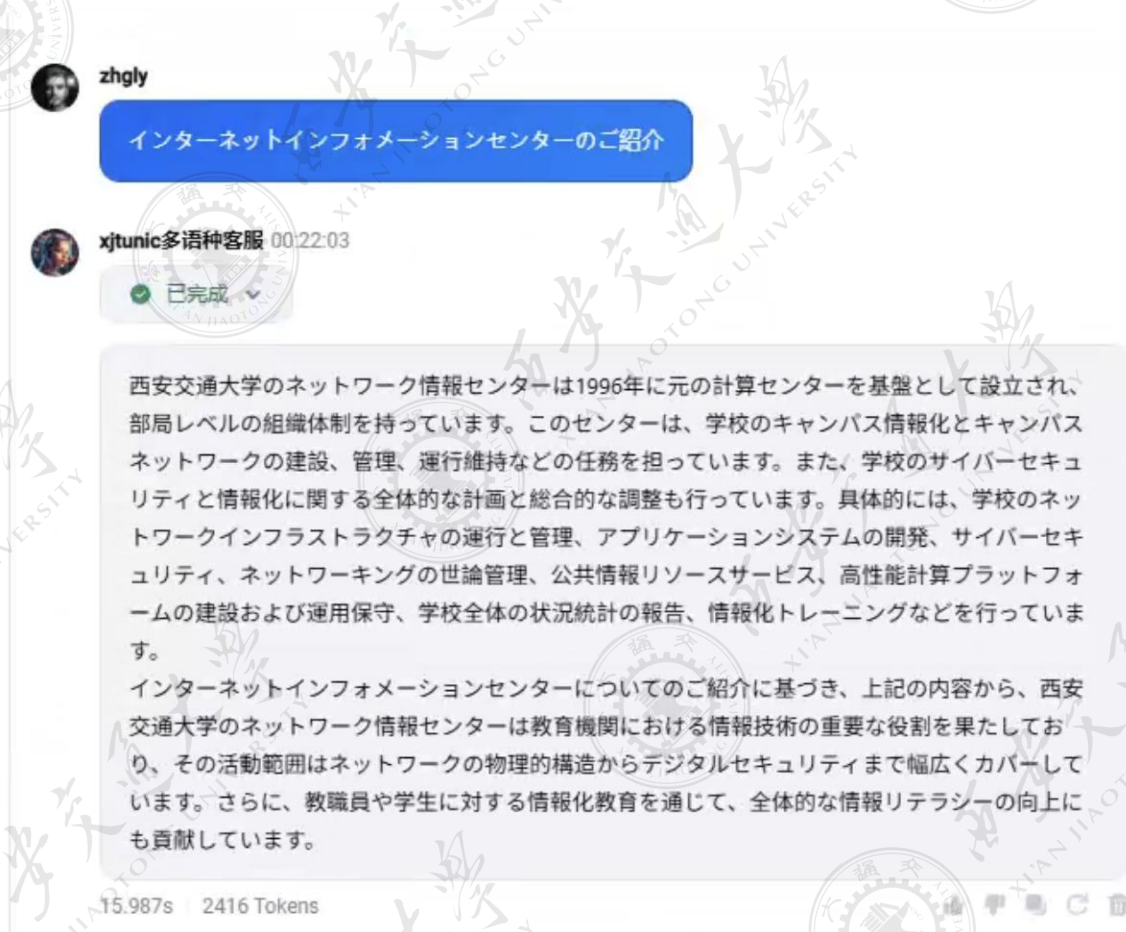
阶段四：测试验证

输入不同语种问题，进行测试

中文示例



外文示例



小结

关键知识点	小结
知识库解析分段	知识文件解析分段的参数（分段规则、解析规则）
知识库召回方式	知识库召回参数（检索方式、召回分段数、相似度设置、知识增强）
工作流	流程设计，并利用多种类型的节点搭建工作流。
工作流调试	完整工作流调试、单节点调试。



大赛报名入口
<https://agent.xjtu.edu.cn/aiic>



人工智能大赛技术答疑群

QQ群号: 2154030379



西安交通大学
XI'AN JIAOTONG UNIVERSITY

谢谢大家!
THANK YOU



网络信息中心



2025年5月