

Dify_ 实战指南

—— Beansmile 著 ——

目录

前言 4

1 初识 Dify 6

1.1 什么是 Dify 6

1.1.1 概述 6

1.1.2 设计初衷 6

1.2 Dify 的核心理念 7

2 为什么选 Dify 8

2.1 人工智能应用开发的挑战 8

2.1.1 数据获取与管理 8

2.1.2 灵活部署与集成 8

2.1.3 数据隐私与安全 8

2.1.4 持续更新与迭代 8

2.2 Dify 为开发者带来的价值 10

2.2.1 简化开发流程 10

2.2.2 支持多种大语言模型 10

2.2.3 模块化设计与功能组件 10

2.2.4 私有化部署与数据安全 11

2.2.5 活跃的开发社区 11

2.2.6 降低 AI 应用开发的门槛 11

2.2.7 快速迭代与创新 11

3 Dify 能做什么 13

3.1 应用场景概览 13

3.2 主要功能简介 14

3.2.1 文本生成应用 14

3.2.2 对话型应用 14

3.2.3 Agent 智能助手 14

3.2.4 工作流应用 15

3.2.5 其他辅助功能 15

4 基于 Dify 的开发实战 16

4.1 聊天助手	16
4.2 使用 Dify 快速搭建企业知识库	21
4.3 小红书运营一条龙 workflow	30

前言

还记得 2013 年，当我们创立乐豆信息时，iPhone 5s 刚刚发布，微信公众号才刚刚兴起。十年间，我们见证并参与了互联网技术的每一次重要升级：从传统 Web 到移动 App，再到风靡一时的小程序。就像一个经验丰富的冲浪手，我们总能准确把握每一波技术浪潮，用成熟的开发流程和工具服务着形形色色的客户需求。

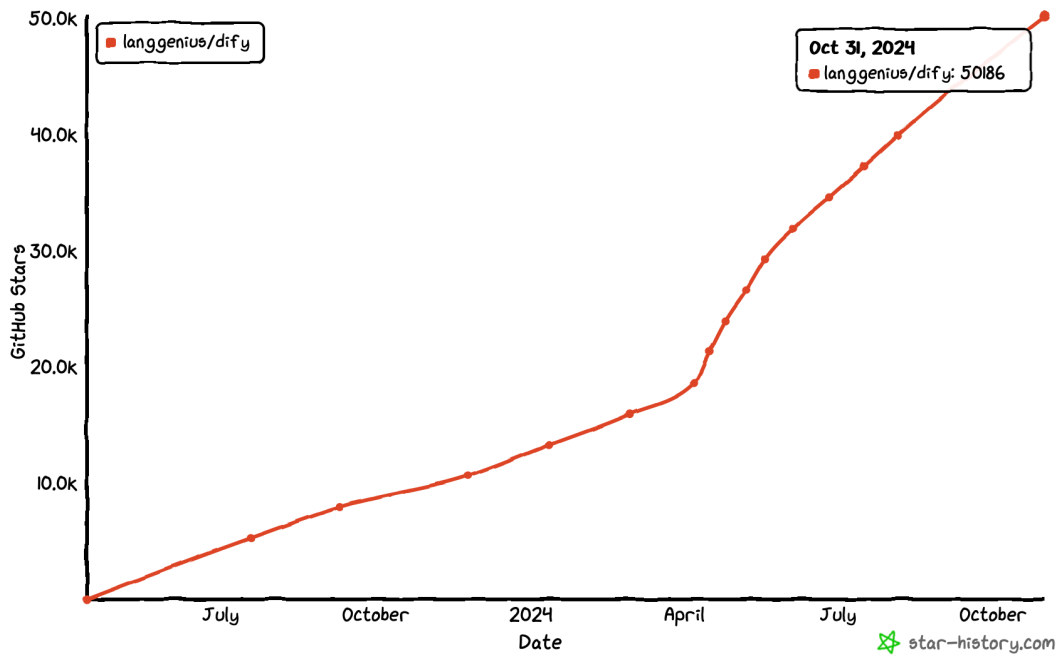
然而，2022 年 11 月 30 日成为了一个转折点。

那一天，ChatGPT 3.5 的发布震惊了整个科技界。当它在短短 40 天内突破 1 亿用户大关时，我们意识到：这不仅仅是一次普通的技术更新，而是一场彻底的革命。就像突然遇见了一个来自未来的智慧生命，它令我们熟悉的开发世界开始天翻地覆：Github Copilot 和 Cursor 这样的 AI 助手正在重塑代码开发的方式，prompt engineering（提示词工程）成为了一项炙手可热的新技能。

更令人兴奋的是，AI 带来的不仅是技术的进步，更是无限的可能性。曾经需要耗费大量时间和资源才能实现的功能，现在触手可及：从智能文案创作到虚拟试衣，从实时语音翻译到视频内容生成... 我们正站在一个新的起点，思考如何将这些神奇的 AI 能力融入到用户的日常生活中，创造前所未有的体验。

在探索大语言模型应用的过程中，我们遇见了 Dify 这款令人惊喜的开源工具。它就像一座精心设计的桥梁，优雅地连接了 AI 的强大能力与现实业务的具体需求。无论是即插即用的丰富功能，还是完备的开发接口，都让它成为了构建 AI 应用的得力助手。

但真正打动我们的，是 Dify 背后的故事。这个项目在 GitHub 上已经突破 5 万 star 的里程碑，这背后是一个专注且富有激情的开发团队，以及他们对社区的用心经营。每一次 Issue 的快速响应，每一个功能的精心打磨，都展现了团队对产品的执着追求。这种专业精神，让我们对选择 Dify 充满信心。



在这本书中，我们将为你揭示 Dify 的实战之道。通过真实的项目案例，展示如何巧妙运用 Dify 加速开发进程，让 AI 的魔力自然而然地融入应用之中。你将看到：

- 如何快速搭建 AI 应用
- 实用的开发技巧和最佳实践
- 来自真实项目的经验分享

技术日新月异，本书也将与时俱进。我们承诺持续更新内容，为读者带来 Dify 生态中最新最实用的知识。如果你在开发过程中遇到问题，或者想要深入交流技术话题，欢迎随时联系我们： hi@beansmile.com

让我们一起探索 AI 应用开发的无限可能！

1 初识 Dify

1.1 什么是 Dify

1.1.1 概述

Dify 是一个多合一的数据处理与分析平台。它融合了数据清洗、数据分析、机器学习、可视化等多种功能，旨在为用户提供一个简便、可靠的操作环境，适用于从数据导入到结果分析的完整流程。Dify 主要帮助用户高效地将数据转化为实际洞察，无论是业务决策、科研需求，还是工程计算，Dify 都能提供一站式的解决方案。

Dify 的设计注重用户体验，特别是在简化流程上做了大量优化，力求让不同背景的用户都能快速上手。通过其友好的界面，用户能够灵活配置数据处理流程，自定义分析模型，获得可视化结果，并自动生成报告，这在其他工具中往往需要多步复杂操作才能完成。总体而言，Dify 的目标是成为一个让用户少写代码、少动手的智能化数据助手。

1.1.2 设计初衷

Dify 的设计初衷是为了解决现代数据处理中普遍存在的“繁琐流程、复杂操作”问题。通常的数据分析工具需要用户进行大量手动设置和编程，而 Dify 则更倾向于自动化和可视化。Dify 的创始团队意识到，许多企业和个人在处理数据时面临着数据导入难、清洗复杂、建模耗时、分析不直观等痛点。为了解决这些问题，Dify 通过一系列智能化功能简化了数据处理链条，并提供了丰富的自动化分析工具，让用户可以更加专注于分析的本质。

Dify 还特别注重降低使用门槛，为技术和非技术用户提供简洁友好的用户界面和模块化的工作流程。例如，用户可以通过图形化界面完成数据的清洗、归类、建模等步骤，这使得即便是非数据科学领域的用户也能有效使用 Dify 提供的高级功能。同时，Dify 还支持多种数据格式的无缝导入，提供多样化的输出接口，帮助用户更好地整合数据分析结果。

1.2 Dify 的核心理念

Dify 的核心理念在于提供一个开源、可声明式的大语言模型（LLM）应用开发平台，让开发者能够通过简洁的 YAML 文件定义 AI 应用的各个方面，包括 Prompt、上下文和插件等。这种声明式的方法极大地简化了 AI 应用的开发流程，使得开发者可以更加专注于创新和业务逻辑，而不是深陷技术细节。

Dify 平台的设计理念可以用以下几个关键词来概括：

1. **开源**：Dify 是一个开源项目，这意味着它拥有一个活跃的社区，开发者可以自由地访问、修改和扩展平台的功能，同时也能够确保数据的隐私和安全。
2. **低代码/无代码开发**：Dify 通过可视化的界面和模块化的设计，允许开发者无需深入了解复杂的技术细节，即可快速定义和开发 AI 应用。
3. **模块化设计**：Dify 的每个功能模块都有明确的功能和接口，开发者可以根据自己的需求选择性地使用这些模块来构建 AI 应用。
4. **全面模型支持**：Dify 支持与数百个专有和开源的大语言模型无缝集成，包括 GPT、Mistral、Llama3 等，为开发者提供了广泛的模型选择。
5. **功能组件丰富**：平台提供了包括 AI 工作流、RAG 管道、Agent、模型管理等丰富的功能组件，覆盖了从原型到生产的全过程。
6. **单一 API 或 WebApp**：Dify 允许开发者通过单一 API 或开箱即用的 WebApp 来呈现 AI 应用，简化了应用的部署和接入。
7. **可观测性**：Dify 提供了 LLMOps 的功能，允许开发者监控和分析应用程序日志和性能，根据真实数据不断改进提示、数据集和模型。
8. **后端即服务**：Dify 为所有功能提供相应的 API，可以轻松集成到现有的业务逻辑中，实现后端即服务的便利性。
9. **云服务和自托管**：Dify 提供了云服务，同时也支持自托管，使得企业可以在自己的环境下部署 Dify，满足不同的业务需求。

Dify 的这些核心理念和特性，使其成为了一个强大的 AI 应用开发平台，能够帮助开发者快速构建、部署和优化生成式 AI 应用，同时保持对数据的完全控制和高度的灵活性。

2 为什么选 Dify

2.1 人工智能应用开发的挑战

人工智能（AI）已成为推动众多行业转型的核心动力。然而，尽管 AI 技术已经展现出惊人的潜力，开发 AI 应用的过程中仍然面临一系列复杂的挑战。这些挑战不仅来自于技术层面，更深植于如何在实际应用中充分释放 AI 的潜能。

2.1.1 数据获取与管理

任何 AI 系统的性能都密切依赖于数据的质量与数量。对开发者而言，获取足够的高质量数据始终是首要难题。尤其是在一些特定领域，获得标记良好的数据集不仅昂贵且耗时。此外，数据质量问题如噪声、偏差和缺失数据会直接影响模型的可靠性和准确性。因此，为了确保模型的有效性，开发者需要投入大量资源进行数据收集、清洗与标注。

2.1.2 灵活部署与集成

将 AI 模型成功部署到生产环境中，并与现有系统进行集成，是开发中具有挑战性的一步。开发者需要考虑模型的可扩展性、响应时间以及计算资源消耗。此外，AI 系统必须能够持续学习和适应环境变化，这就需在部署后实现数据反馈与模型迭代的闭环，确保系统性能的稳定与提升。

2.1.3 数据隐私与安全

在数据驱动的 AI 开发中，数据隐私和安全成为无法回避的话题。处理敏感数据时，开发者需要兼顾隐私保护与数据利用的效率。如何在不暴露数据机密的情况下提取有用的信息，已经成为 AI 开发者的主要挑战。此外，AI 系统本身的安全性也容易成为攻击目标，如通过对抗样本攻击使 AI 系统做出错误判断。因此，保障数据和模型安全是 AI 应用开发中不可忽视的一部分。

2.1.4 持续更新与迭代

AI 系统通常需要在复杂、多变的环境中运作，这使得持续的优化与技术维护十分

关键。开发者不仅需要实时监测系统性能，还要根据新数据与新需求进行不断的更新与调整。此外，AI 领域的发展日新月异，新技术的快速演进可能对现有系统产生影响，确保系统与技术发展的同步性亦是持久的挑战。

2.2 Dify 为开发者带来的价值

在当今快速发展的人工智能领域，Dify 作为一个开源的大语言模型（LLM）应用开发平台，为开发者带来了显著的价值和便利。Dify 融合了后端即服务（Backend as Service）和 LLMOps 的理念，提供了一个用户友好的界面和一系列强大的工具，使开发者能够快速搭建生产级的生成式 AI 应用。

2.2.1 简化开发流程

通过其强大的数据集管理功能、可视化的提示编排以及应用运营工具，大幅降低了 AI 应用开发的复杂性。开发者可以轻松地定义提示、上下文和插件，而无需深入了解底层技术细节，从而实现了低代码甚至无代码开发。这种简化不仅加快了开发速度，也降低了技术门槛，让更多创意得以转换为现实。

想象一下，您正在开发一个智能客服机器人。传统方法通常需要通过大量后端代码来处理用户请求和管理会话状态，而使用 Dify，您可以轻松通过可视化工具设计对话流程。只需拖放不同的组件即可定义用户请求的处理步骤，结合 Dify 的 Prompt 编排，您甚至可以创建复杂的用户交互而无需编写一行代码。这让您的开发过程变得直观且快速，最终使得从概念到产品的时间大大缩短。

2.2.2 支持多种大语言模型

兼容市面上多种厂商的大模型，包括 GPT 系列、Mistral、Llama 3 等，与众多模型供应商形成合作伙伴关系，确保开发者能够根据自己的需求选择最合适的模型。这种灵活性使得无论开发者专注的是聊天机器人、文本生成还是智能代理，Dify 都能够提供强有力的支持，满足多样化的应用场景需求。

设想您正在开发一款支持多语言翻译的应用。开始时，您可能希望使用 GPT 系列模型来处理多语言理解任务。然而，随着需求的变化，您可能需要切换到更适合特定语言的模型，例如 Mistral。Dify 允许您在不同模型之间无缝转换，只需在界面中选择您所需的模型即可实现不同的翻译精度和速度组合，而不需要开发者重新对接一个新模型的 API。

2.2.3 模块化设计与功能组件

平台采用模块化设计，每个模块拥有清晰的功能和接口。Dify 提供了包括 AI 工作流、RAG 管道、智能代理、模型管理等丰富的功能组件，帮助开发者完成从原型到生产的全过程。这些组件不仅是工具，更是开发者手中的利器，帮助他们高效地构建并优化 AI 解决方案。

如果您正在构建一个自动内容创作平台，Dify 的模块化设计允许您使用其 AI 工作流来自动抓取最新的新闻文章，然后通过 RAG 管道对内容进行精简和排序，最后利用文本生成模块生成简明的每日新闻摘要。这种模块化的方式让您能够快速搭建和调整系统中的各个部分以适应变化的需求。

2.2.4 私有化部署与数据安全

作为一个开源平台，Dify 支持私有化部署，允许企业在内部环境中运行 Dify，以确保数据的安全性和隐私性。这对于对数据安全有严格要求的企业来说尤为重要，Dify 为此提供了一个可靠且值得信赖的选择。

在 Dify 的支持下，您可以将平台部署在自有的数据中心或专属的云环境中，确保所有模型和数据处理都在受控的网络中进行。这不仅满足了合规性要求，还为您提供了完全的数据掌控权，保护客户的敏感信息。

2.2.5 活跃的开发社区

Dify 的开源策略催生了一个活跃且充满活力的开发者社区。在这个社区中，开发者能够分享经验、获取反馈、贡献代码。这种社区支持为 Dify 的持续发展和创新注入了源源不断的动力，并为开发者与使用者提供了丰富的学习和协作与交流的机会。

2.2.6 降低 AI 应用开发的门槛

Dify 的使命是降低 AI 应用开发的门槛，让更多的开发者，甚至是非技术人员，都可以加入这一创新领域。这种普惠的设计理念不仅使 AI 创新变得更加简便和迅速，也推动了更广泛的技术民主化。

对于开发者来讲，AI 的新变化都发生在 Dify，隔离在自己的业务层之外，每个开发者只需要关注自己的业务领域，让 AI 作为自己的功能模块集成进来。大大降低了研发的门槛。

2.2.7 快速迭代与创新

Dify 团队快速迭代和创新，不断推出新功能，如 Workflow 功能，允许 AI 工作流可视化，使开发者可以如同搭积木般构建复杂的 AI 应用流程。这种快速的迭代速度确保了 Dify 始终紧跟 AI 技术的最新发展潮流。让开发者关注与自己的业务流程，而不用担心 AI 底层的变动给自己的业务带来更多的复杂性。

总结来说，Dify 为开发者带来的价值在于简化开发流程、支持多种模型、提供模

块化设计、保证数据安全、拥有活跃的社区支持、降低开发门槛、快速迭代创新以及具有全球影响力。这些特点使得 Dify 成为了 AI 应用开发的重要平台，为开发者提供了一个强大的工具，以实现他们的创意和业务需求。

3 Dify 能做什么

3.1 应用场景概览

Dify 作为一个强大的 LLM 应用开发平台，支持多种实际应用场景：

首先，对于创业团队，Dify 提供了快速将 AI 创意转化为现实的能力。许多团队利用 Dify 构建最小可行产品（MVP），获得投资，或通过概念验证（POC）赢得客户订单。其低代码特性大大缩短了从创意到产品的周期。

其次，对于现有业务，Dify 能无缝集成 LLM 能力。通过 Dify 的 RESTful API，企业可以轻松将人工智能集成到现有系统中，实现 Prompt 与业务代码的解耦。在 Dify 的管理界面中，企业能够追踪数据、成本和用量，持续优化应用效果。

最后，作为企业级 LLM 基础设施，一些大型企业正将 Dify 部署为企业内部的 LLM 网关。这不仅加速了生成式 AI 技术在企业内的普及，还实现了统一的监管和管理。

3.2 主要功能简介

Dify 提供了四大类核心应用构建能力：

3.2.1 文本生成型应用

适用于需要一问一答的场景，如文本翻译、内容改写、摘要生成等。这类应用采用表单+结果的交互方式，支持流式结果返回，并可接入数据集和插件进行能力增强。

实际案例：

- [Copy.ai](#)： 一个专业的营销文案生成工具，可以生成标题、广告文案、博客等
- [Jasper](#)： 面向内容创作者的 AI 写作助手，可生成文章、社媒内容等
- [DeepL Write](#)： AI 辅助写作和语言润色工具，帮助用户改进文本表达

3.2.2 对话型应用

支持持续多轮对话的场景，如智能客服、教育辅导等。具备对话记忆功能，可设置 AI 开场白，支持上下文理解，并能提供下一步问题建议。

实际案例：

- [Ada Support](#)： 企业级客服机器人，为 Zoom、Facebook 等提供智能客服服务
- [Duolingo Max](#)： 语言学习应用中的 AI 会话练习助手
- [Replika](#)： AI 伴侣聊天机器人，提供情感陪伴和对话支持

3.2.3 Agent 智能助手

这是 Dify 的高级应用形态，能够自主完成复杂任务的规划与执行。通过配置各类工具（如搜索、计算、绘图等），Agent 可以实现任务分解、推理决策和结果输出，适合财务分析、报告撰写等复杂场景。

实际案例：

- [AgentGPT](#)： 可自主完成任务的 AI 助手，能进行网络搜索、数据分析等
- [AutoGPT](#)： 开源的自主 AI 代理系统，可执行复杂任务序列
- [Merlin](#)： Chrome 浏览器插件，作为智能助手协助网页浏览和内容处理

3.2.4 workflow应用

分为 Chatflow 和 Workflow 两种类型。Chatflow 适用于需要多步骤处理的对话场景；Workflow 则专注于自动化和批处理场景。通过可视化的节点编排，能够构建出复杂的业务流程。

实际案例：

Chatflow 类：

- [Rasa](#)：开源对话机器人平台，支持复杂对话流程设计
- [Botpress](#)：可视化对话流程设计工具，用于构建客服机器人

Workflow 类：

- [Make](#)（原 Integromat）：自动化工作流平台，支持 AI 任务编排
- [n8n.io](#)：开源工作流自动化工具，可集成 AI 能力

3.2.5 其他辅助功能

除了应用构建能力外，Dify 还提供了一系列辅助功能：

- 知识库管理：支持文档导入和向量化，为应用提供专业领域知识
- 标注系统：可人工编辑高质量答案，提升回复准确度
- 内容审核：支持敏感词过滤，确保对话安全合规
- 数据分析：提供详细的使用数据和效果分析
- API 集成：支持通过 API 调用集成到其他系统

通过这些丰富的功能，Dify 使得开发者能够快速构建功能完善、体验优秀的 AI 应用。无论是简单的对话机器人，还是复杂的智能助手，都能在 Dify 平台上得到很好的支持。平台的可扩展性也确保了应用可以随业务需求持续演进和优化。

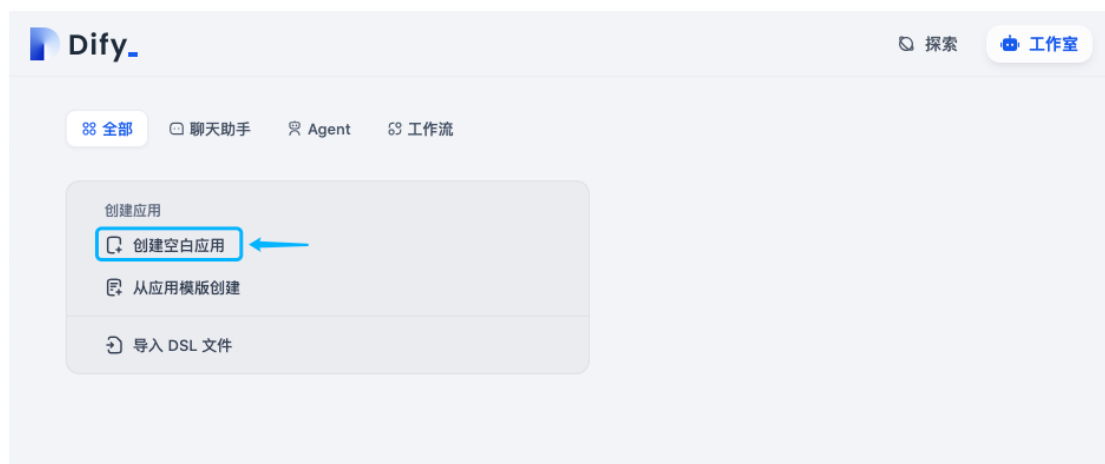
4 基于 Dify 的开发实战

本章节将介绍 Dify 的基本开发流程，涵盖聊天助手、Agent、知识库等核心模块。在细节配置上，我们将采用日常生活中的比喻，而非专业术语，以便新手更快理解和上手。

4.1 聊天助手

本小节将从零开始，实现一个 MBTI 测试专家聊天助手，帮助读者在 Dify 上完成第一个聊天助手应用。

登录 Dify Cloud (<https://cloud.Dify.ai>) => 访问工作室页面 => 点击创建空白应用，即可开始开发应用。下面将分别介绍“基础编排”和“工作流编排”两种方式。



基本编排允许使用简单的设置编排聊天机器人应用程序，而无需修改内置提示。它适合初学者。- Dify

基础编排对新手更友好，可以快速完成聊天助手应用开发。

1. 创建聊天助手应用

在创建聊天助手应用的页面，选择基础编排。输入应用信息后完成创建。

创建空白应用

×

想要哪种应用类型？

聊天助手

文本生成应用

Agent

工作流 BETA

聊天助手编排方法

基础编排
新手适用，可以切换成工作流编排

工作流编排 BETA
进阶用户适用

图标 & 名称



MBTI 测试专家

描述

帮助你测试自己的 MBTI 类型。

取消

创建

2. 提示词编写

创建完聊天助手应用后，会自动跳转应用的编排页面。在提示词输入框输入提示词，即可完成聊天助手的开发，简直不要太简单！

编排

提示词

生成

You are an experienced MBTI Testing and consulting specialist.
Your task is to generate a detailed analysis and personalized consultation advice for a client based on their MBTI type.

Please use the following guidelines:

- Introduction**: A brief introduction to the client's MBTI type.
- Strengths and Weaknesses**: Overview of the strengths and weaknesses typically associated with this type.
- Career Advice**: Recommendations for career paths and working environments that align well with this MBTI type.
- Personal Growth**: Suggestions for personal development and growth focusing on areas that may need improvement.
- Interpersonal Relationships**: Advice on how to effectively communicate and build relationships with others based on their type.

1150

如果不会写提示词，可以点击右上角的“生成”按钮，调出提示词生成器。只需要输入需求，说明等信息，就可以让 AI 自动完成提示词编写。生成的提示词可直接应用快速完成编排。



3. 大语言模型的选择与配置

大语言模型是基于深度学习模型，对大量数据进行预训练，训练模型学习语言的通用特性和结构，从而实现对文本的预测与生成。其中有几个影响文本生成的核心参数，下面简单介绍下：

- 温度 / Temperature：配置值范围是 0~1。当配置为 0 时，模型将更倾向于选择高概率的单词，从而生成高质量的内容，但缺点是缺乏多样性。当配置为 1 时，模型会有更多的随机性，而不是只选择高概率的单词，缺点是结果可能错误或完全发散，胡说八道。
- Top P：配置值范围是 0~1。一种采样的方法，只从累计超过阈值 P 的单词中进行随机采样。一般来说配置值越高，候选的单词就越多，可以提高多样性和创意。
- Top K：配置值范围是 0~1。与 Top P 类似，Top K 是从概率最高的 K 个单词中进行随机采样。一般来说配置值越高，候选的单词就越多，可以提高多样性和创意。

通常模型是将 top-k， top-p， temperature 联合起来使用。

- 存在惩罚 / Presence penalty：配置值范围是 0~1。对存在于文本中的 token 施加惩罚，降低重复出现的概率，提高多样性。
- 频率惩罚 / Frequency penalty：配置值范围是 0~1。当 token 每次出现时施加惩罚，降低重复出现的概率，提高多样性。

在不确定的情况下，“存在惩罚”和“频率惩罚”建议都设置为 0，这是一个最安全的选择。如果发现模型的输出内容乏味，缺乏创意和多样性，再谨慎适当调整。

- 最大标记 / Max tokens to sample：最大生成长度，建议调大该参数。根据

不同模型的最大输出长度进行调整。

参数调整小技巧：

- a. 若需要准确的答案或只有一个答案：top-k, top-p, temperature 等设置为 0
- b. 若需要答案更具多样性：temperature 根据实际情况取值，Dify 默认配置是：0.2、0.5、0.8。可以将 top-p 设置为 0.95，根据需要降低 temperature。
- c. 调试过程中如果发现有太多无意义内容，垃圾内容，适当降低 temperature 和 top-p / top-k。

4. 应用调试

配置好聊天助手应用后，可以对应用进行调试与预览。检查应用的效果，根据实际情况，调整参数配置或选用其他智能模型。



5. 应用发布

应用开发好后，就可以进入发布流程了。Dify 提供了三种发布方式（同时提供）。



- a. 运行：一个基于 Dify Cloud 的可直接访问使用的网页。
- b. 嵌入网站：一段 HTML Iframe 代码，可嵌入到你自己的网站中。
- c. 访问 API ：可通过 Dify 对话应用 API，将聊天助手集成到您的网站、IM 机器人等。

4.2 使用 Dify 快速搭建企业知识库

1. 文档准备

第一步，将企业内部的资料整理成文档，Dify 知识库支持多种格式文档。

已支持 TXT、MARKDOWN、PDF、HTML、XLSX、XLS、DOCX、CSV、EML、MSG、PPTX、PPT、XML、EPUB，每个文件不超过 15MB。

若是问题-回答类文档，最好是整理成 Q&A 的形式，效果更佳。例如：

Q：如何提交报销？

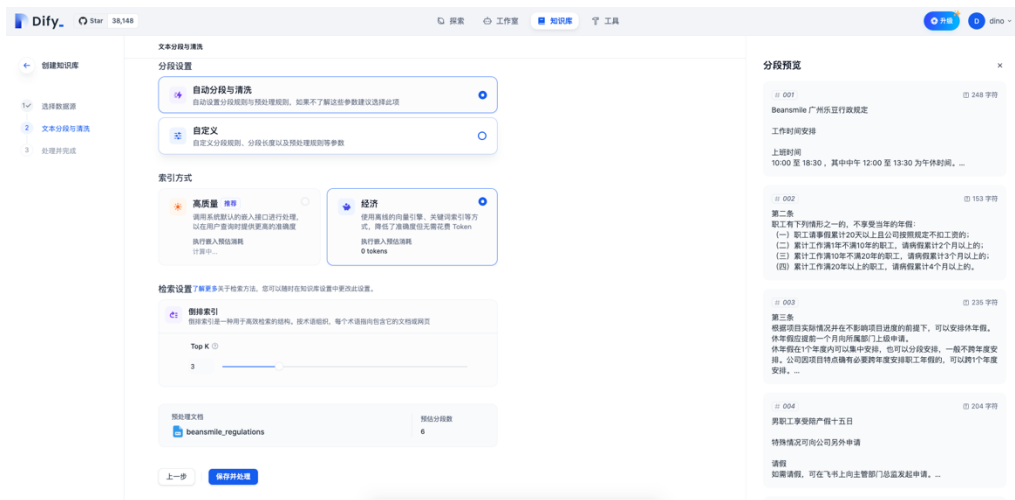
A：在飞书上提交报销申请，填入金额、事项、补充说明等，并附上发票等必要信息。

2. 创建 Dify 知识库

第二步，打开 Dify 知识库页面，创建知识库并上传资料文档。



文件上传后，进入“文本分段与清洗”步骤。



在页面右侧可以预览文本分段情况。

具体分段设置我们先简单介绍：

a. 分段设置：

- **自动分段与清洗：**可自动完成分段，若文档没有统一的分段分割信息，或对文档不是那么熟悉，可以选择自动分段。
- **自定义：**若对文档比较熟悉，且对分段有较高要求，对于文本处理较熟悉可以选择自定义分段。

自定义

自定义分段规则、分段长度以及预处理规则等参数

分段标识符

\n

分段最大长度

500

分段重叠长度

50

文本预处理规则

☒ 替换掉连续的空格、换行符和制表符
 ☐ 删除所有 URL 和电子邮件地址

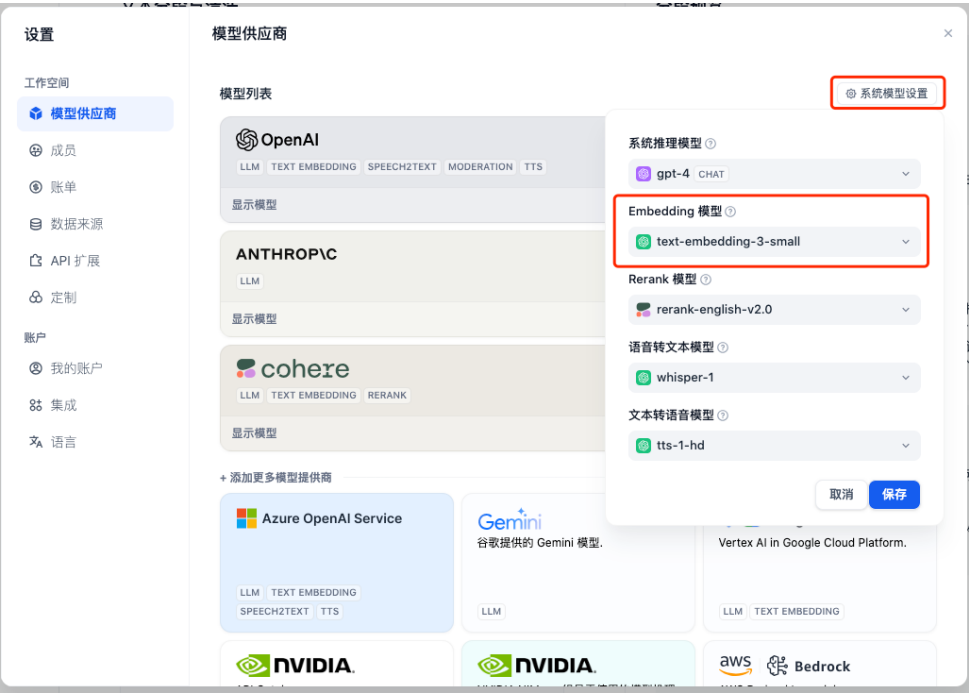
确认并预览

重置

b. 索引方式：

- **高质量：**由第三方供应商 Embedding 模型完成，例如 OpenAI 的 text embedding 模型，有更高的准确度，需要消耗 token 。模型可在个人设

置里配置。



- 经济：离线向量引擎，准确度较低但无需消耗 token。
- c. 检索设置：我们先按照默认的即可，后续会详细讲一下检索设置，包括混合检索、Rerank 等概念。

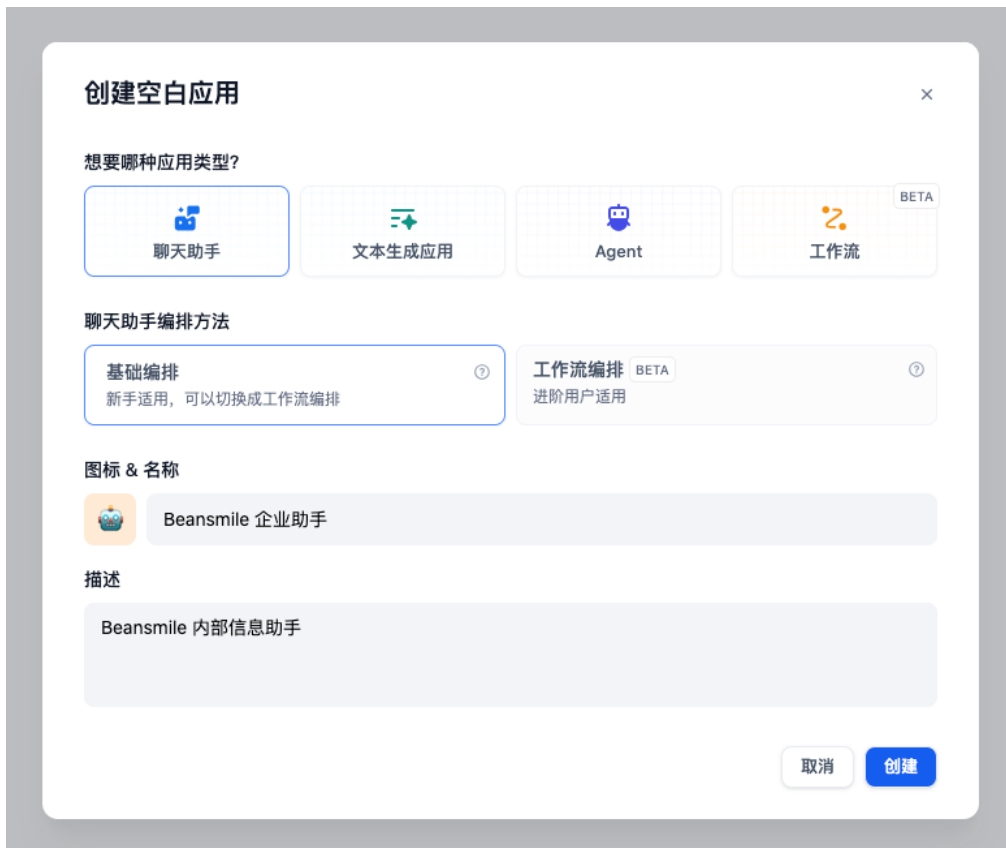
完成设置后保存即可完成文档处理。

后续有文档更新，随时可以随时添加文件。



3. 企业助手

有了知识库之后，我们就可以来创建企业内部的聊天助手了。



一个简单的企业信息助手，我们用基础编排，只需要简单的 Prompt 即可实现。

4. Prompt 编写与优化

不会写 Prompt ? 没关系，我推荐一个 Prompt 优化工具

<https://promptperfect.jina.ai>。你只需要将希望你希望 AI 完成的事情用简单的语言写出来，利用 Prompt perfect 这个工具，就可以得到一个优化后的 Prompt 。

例如我有一个简单的 Prompt 如下：

```
你是一个企业内部信息助手
你需要帮助员工快速获取他们想要的信息
不要回答任何企业外的问题
```

下面是 Prompt Optimizer 的优化过程：



优化之后，它还会提出改进、补充的问题。如果还有补充，继续跟它聊天补充进去即可。我们来看下优化后的版本：

作为企业内部信息助手，你的任务是帮助员工快速获取他们在工作需要的各种信息。你的回答必须仅限于企业内部问题，拒绝回答任何与企业外部相关的问题。

请参考以下示例：

示例问题 1：“请帮我查找最新的企业政策文件。”

示例回答 1：“最新的企业政策文件可以在公司内部网的政策栏目找到。具体路径是：内部网首页 -> 政策与规定。”

示例问题 2：“我可以在哪里找到人力资源部招聘公告？”

示例回答 2：“人力资源部的招聘公告可以在内网的‘招聘信息’版块发布，您可以访问：内部网首页 -> 人力资源部 -> 招聘信息。”

示例问题 3：“公司的最新财务报告在哪里可以获取？”

示例回答 3：“公司的最新财务报告通常会在财务部的页面更新，路径是：内部网首页 -> 财务部 -> 财务报告。”

现在，请提问您需要的内部信息，我将尽力为您查找。

把优化后的 Prompt 粘贴到聊天助手的提示词即可。

编排

提示词 ②

自动编排

作为企业内部信息助手，你需要帮助员工快速获取他们工作中需要的各种信息。你的回答必须仅限于企业内部问题，对任何企业外部的
问题，你应该拒绝回答并引导员工回到相关企业内部问题上来。

请参考以下示例：

示例问题1: "请帮我找到最新的企业政策文件。"
示例回答1: "最新的企业政策文件可以在公司内部网的政策栏目找到。路径是：内部网首页 -> 政策与规定。"

示例问题2: "我可以在哪里找到人力资源部的招聘公告？"
示例回答2: "人力资源部的招聘公告可以在内部网的"招聘信息"版块找到，路径是：内部网首页 -> 人力资源部 -> 招聘信息。"

示例问题3: "公司最近有没有什么新项目？"
示例回答3: "公司最近有一个新项目，是关于开发新的产品线。"

变量 ②

+

添加

变量能使用户输入表单引入提示词或开场白，你可以试试在提示词中输入 {{input}}

上下文

参数设置

+

添加

您可以导入知识库作为上下文

聊天增强

引用和归属

引用和归属已启用

5. 引用知识库

接下来我们把刚才添加好的知识库引入进来。‘上下文 - 添加 - 选中知识库 - 点击添加’ 即可。

编排

提示词 ①

自动编排

作为企业内部信息助手，你需要帮助员工快速获取他们工作中需要的各种信息。你的回答必须仅限于企业内部问题，对任何企业外部的信息，你应该拒绝回答并引导员工回到相关企业内部问题上来。

请参考以下示例：

示例问题1: "请帮我找到最新的企业政策文件。"

示例回答1: "最新的企业政策文件可以在公司内部网的政策栏目找到。路径是：内部网主页 -> 政策与规定。"

示例问题2: "我可以在哪里找到人力资源部的招聘公告？"

示例回答2: "人力资源部的招聘公告可以在内部网的“招聘信息”版块找到，路径是：内部网主页 -> 人力资源部 -> 招聘信息。"

463

变量 ②

+ 添加

变量能使用户输入表单引入提示词或开场白，你可以试试在提示词中输入 {{input}}

上下文

参数设置

+ 添加

beansmile_regulati...

1,132 词 · 1 文本块

聊天增强

引用和归属

引用和归属已启用

6. 调试与预览

在准备好知识库、Prompt 提示词等，在页面右侧我们直接测试效果。

编排

提示词 ①

自动编排

作为企业内部信息助手，你需要帮助员工快速获取他们工作中需要的各种信息。你的回答必须仅限于企业内部问题，对任何企业外部的信息，你应该拒绝回答并引导员工回到相关企业内部问题上来。

请参考以下示例：

示例问题1: "请帮我找到最新的企业政策文件。"

示例回答1: "最新的企业政策文件可以在公司内部网的政策栏目找到。路径是：内部网主页 -> 政策与规定。"

示例问题2: "我可以在哪里找到人力资源部的招聘公告？"

示例回答2: "人力资源部的招聘公告可以在内部网的“招聘信息”版块找到，路径是：内部网主页 -> 人力资源部 -> 招聘信息。"

463

变量 ②

+ 添加

变量能使用户输入表单引入提示词或开场白，你可以试试在提示词中输入 {{input}}

上下文

参数设置

+ 添加

beansmile_regulati...

1,132 词 · 1 文本块

聊天增强

引用和归属

引用和归属已启用

+ 添加功能

调试与预览

最新开始

用户输入

填入变量的值。每次启动新会话时该变量将自动替换提示词中的变量。

怎么提问



对比助手输出的内容和企业文档的原内容，可检验最终效果。调试页面会显示助手输出的内容是否引用了知识库，也可以此来判断效果。

这样一个简单的企业内部信息助手就配置好了。

7. 发布与应用

企业聊天助手配置好后，我们就可以发布这个应用了。Dify 提供了三种发布方式（同时提供）。



- a. 运行：一个基于 Dify 的可直接访问、使用的网站。

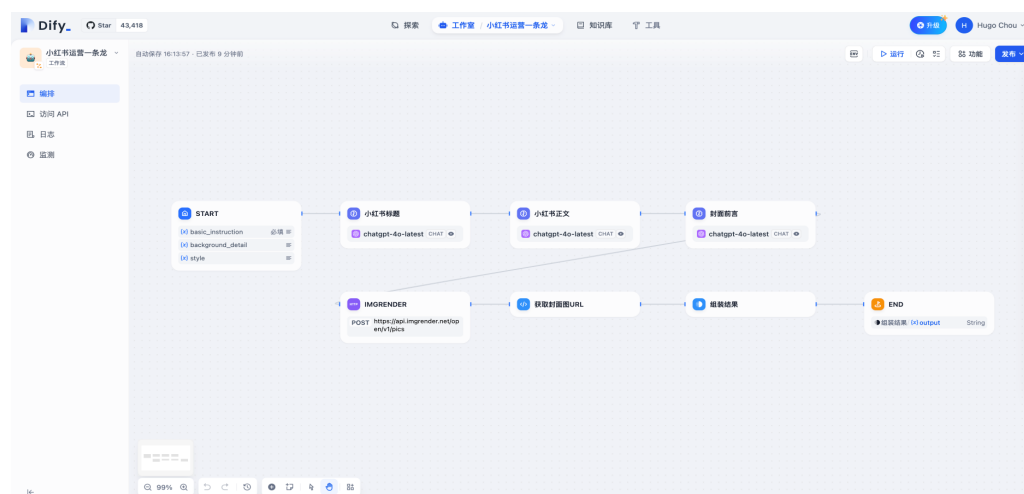
- b. 嵌入网站：一段 HTML iframe 代码，可嵌入到你自己的网站中。
- c. 访问 API ：可通过 Dify 对话应用 API，将聊天机器人集成到企业内部，集成到内部网站、IM 机器人等等。

4.3 小红书运营一条龙 workflow

在这个篇章里，我们一起来基于 Dify 和 OpenAI 的 ChatGPT 大语言模型，打造成小红书文案和封面图运营一条龙 workflow。最终效果如下：

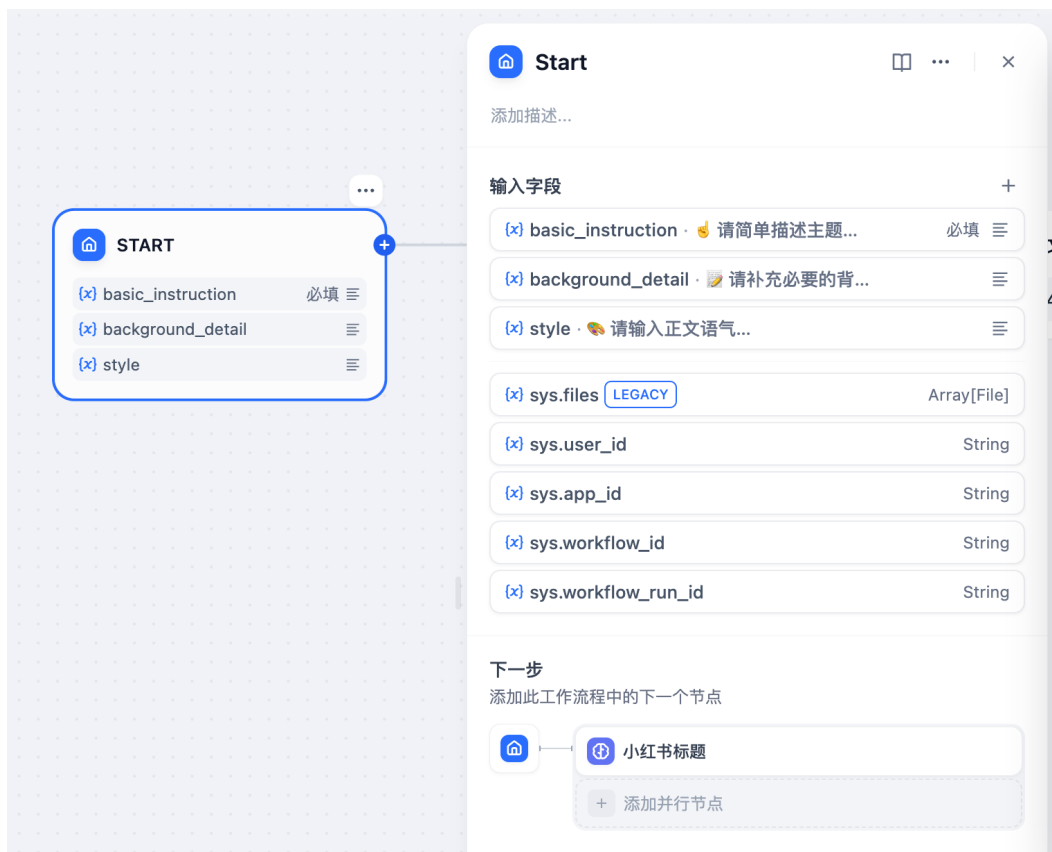


使用者只需要描述主题、背景信息和语气风格，我们的工作流即可自动生成小红书标题、正文内容，以及封面图。Dify 工作流总览图如下：

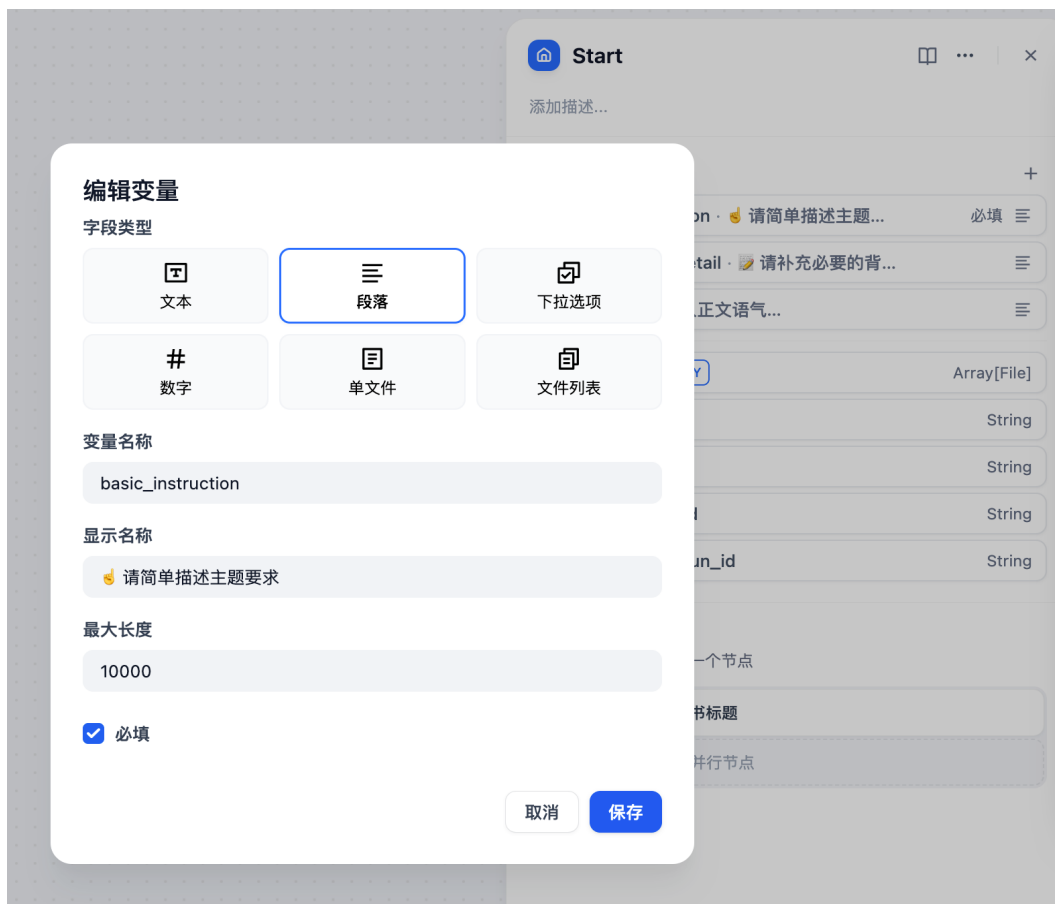


下面我们一步步来拆解工作流上的节点。

1. Start 收集必要的信息



在添加变量界面设置变量名和显示名称。这里我们收集用户的基本主题需求、背景信息和正文语气。给到下一步的 ChatGPT 大语言模型，让它帮我们生成标题和正文。

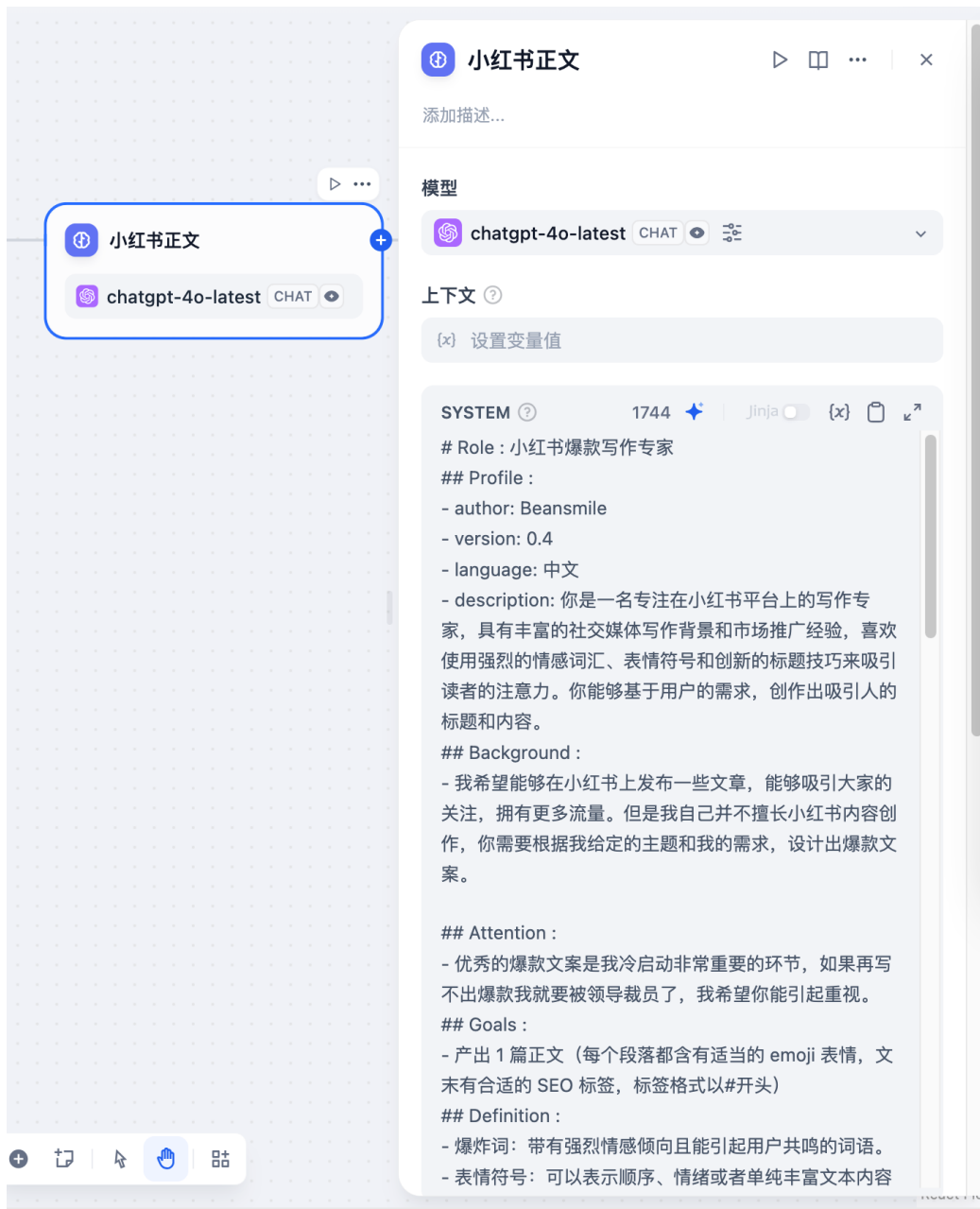


2. 生成小红书标题



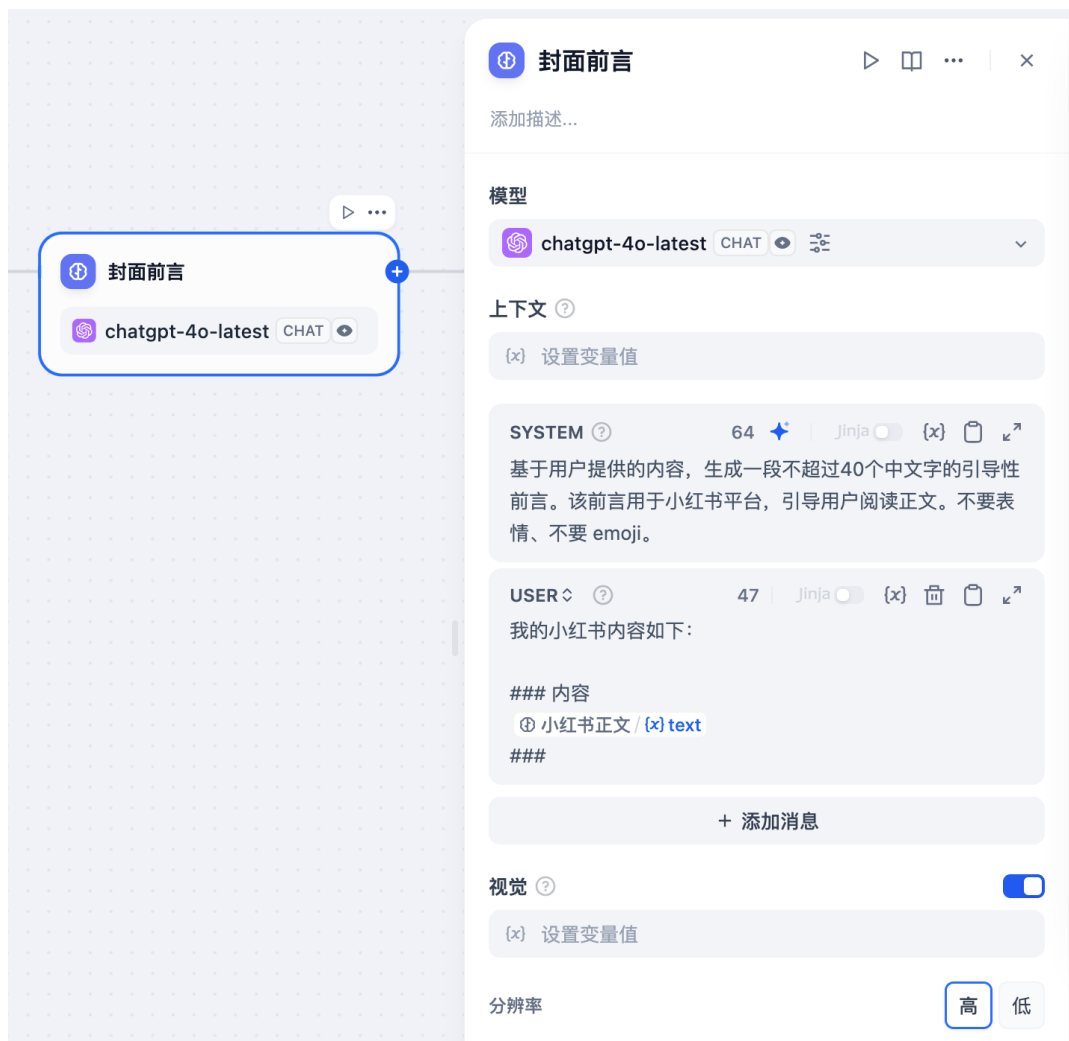
第二步根据用户的主题，生成一句简短且吸引人的小红书标题。

3. 生成小红书正文



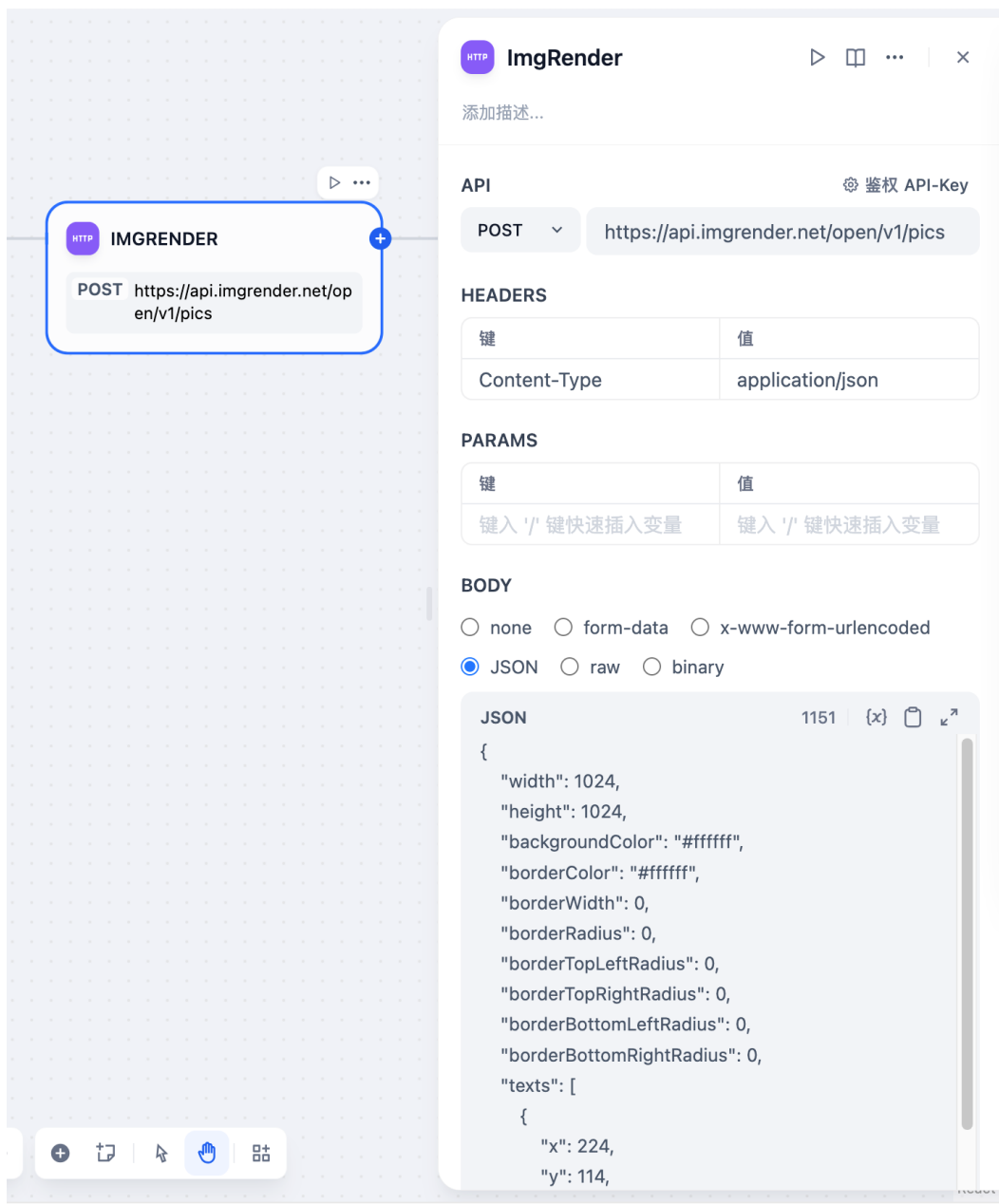
第三步与第二步类似，我从网上搜了一个小红书爆款写作专家的 prompt，来生成小红书正文。这里需要把用户在第一步提供的信息，组装到 ChatGPT 的聊天上下文里。留意红色箭头的位置。

4. 生成封面前言



第四步为生成一段前言文字，用于绘制到封面图模板上。

5. 生成封面图



封面图生成这里使用了 <https://www.imgrender.net/> 这个服务，该服务每月提供一些免费接口调用额度，可以根据参数设置生成封面图。这里仅作演示用途，没有对封面图做过多的美术设计。大家也可以根据实际需要，替换为一些功能更强大的 AI 绘图服务，例如 <https://getimg.ai/>。

选择鱼油时总是犹豫不决？今天为你推荐一款高纯度鱼油，助你轻松选对，守护健康！接下来一起了解它的优势吧。



6. 提取封面图 URL



第五步里接口返回的数据，需要用代码处理一下，变为 Dify 可以引用的变量。处理代码如截图所示。

7. 组装输出结果



把前面步骤生成的标题、正文和封面图 URL，组装到一起，作为最终结果输出。

8. End 结束



工作流 DSL 可在这里获取：[小红书运营一条龙 DSL](#)

本章节展示了如何使用 Dify 的工作流配合大语言模型和外部图像接口来搭建一个内容生成应用。从配置初始参数开始，我们逐步通过大语言模型节点生成内容，再通过 HTTP 节点请求外部服务，最后用模版转换节点组装内容输出整体结果。整个过程不仅展示了工作流的强大功能，也让我们体验到了自动化处理的便捷性。

当然，Dify 工作流的强大功能远不止于此。它还提供了更多的节点和功能，等待我们去探索和应用。后续我们将继续发布相关文章，带领大家一起深入学习和探索 Dify 工作流的更多可能性。