AiDD AI+研发数字峰会
AI+ Development Digital summit
第5届

改造 Dify 实现生产可用的 AI Agent 应用落地

陈迪豪 | 顺丰科技

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 · 敦煌站
K+ 思考周®研习社
时间：2025.08.29-30

K+峰会 · 上海站
K+ 金融专场
时间：2025.10.17-18

K+峰会 · 香港站
K+ 思考周®研习社
时间：2025.11.25-26

K+峰会详情

AiDD峰会

AiDD峰会 · 上海站
AI+研发数字峰会
时间：2025.05.17-18

AiDD峰会 · 北京站
AI+研发数字峰会
时间：2025.08.08-09

AiDD峰会 · 深圳站
AI+研发数字峰会
时间：2025.11.28-29

AiDD峰会详情

演讲嘉宾

陈迪豪

顺丰科技 AI 技术平台高级工程师

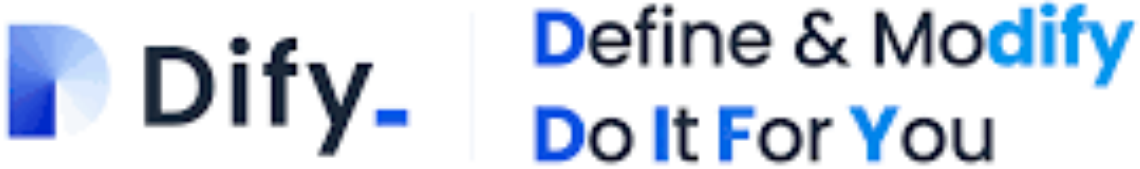
目前担任顺丰科技 AI 技术平台高级工程师，负责顺丰集团 AI 和大模型基础架构功能，曾任第四范式平台架构师和 OpenMLDB 项目 PMC，过去在小米担任云深度学习平台架构师以及优思德云计算公司存储和容器团队负责人。活跃于分布式系统、机器学习相关的开源社区，也是HBase、OpenStack、TensorFlow、TVM 等开源项目贡献者。

目录
CONTENTS

PART 01

介绍 Dify 开发平台

介绍 Dify 开发平台 - What



- 简介：Dify是一个低代码AI平台，旨在帮助用户更快捷地构建AI驱动的应用程序。
- 特点：模块化设计、易于集成、提供多种AI服务接口。
- 优势：低代码开发、高效部署、面向业务需求的AI支持。

介绍 Dify 开发平台 - Why

1. 低代码开发，缩短开发周期

- Dify支持低代码开发，用户只需少量编码即可构建复杂的AI应用，极大地缩短了开发周期。
- 可视化的界面设计和API集成工具使得开发流程更为简洁，即便没有深厚的编程背景也可以快速上手。

2. 模块化架构，支持灵活扩展

- Dify的模块化设计使平台更具扩展性，用户可以按需选择和组合AI模块，适应各种应用场景。
- 平台支持不同AI模型的接入和切换，能满足不同类型的任务需求，如文本处理、图像分析、数据预测等。

3. 高效部署，快速上线

- Dify集成了简化的部署工具，使得AI应用可以更快速地上线。
- 提供云端部署和本地部署选项，支持企业根据需求灵活选择部署环境，有效应对数据安全和隐私要求。

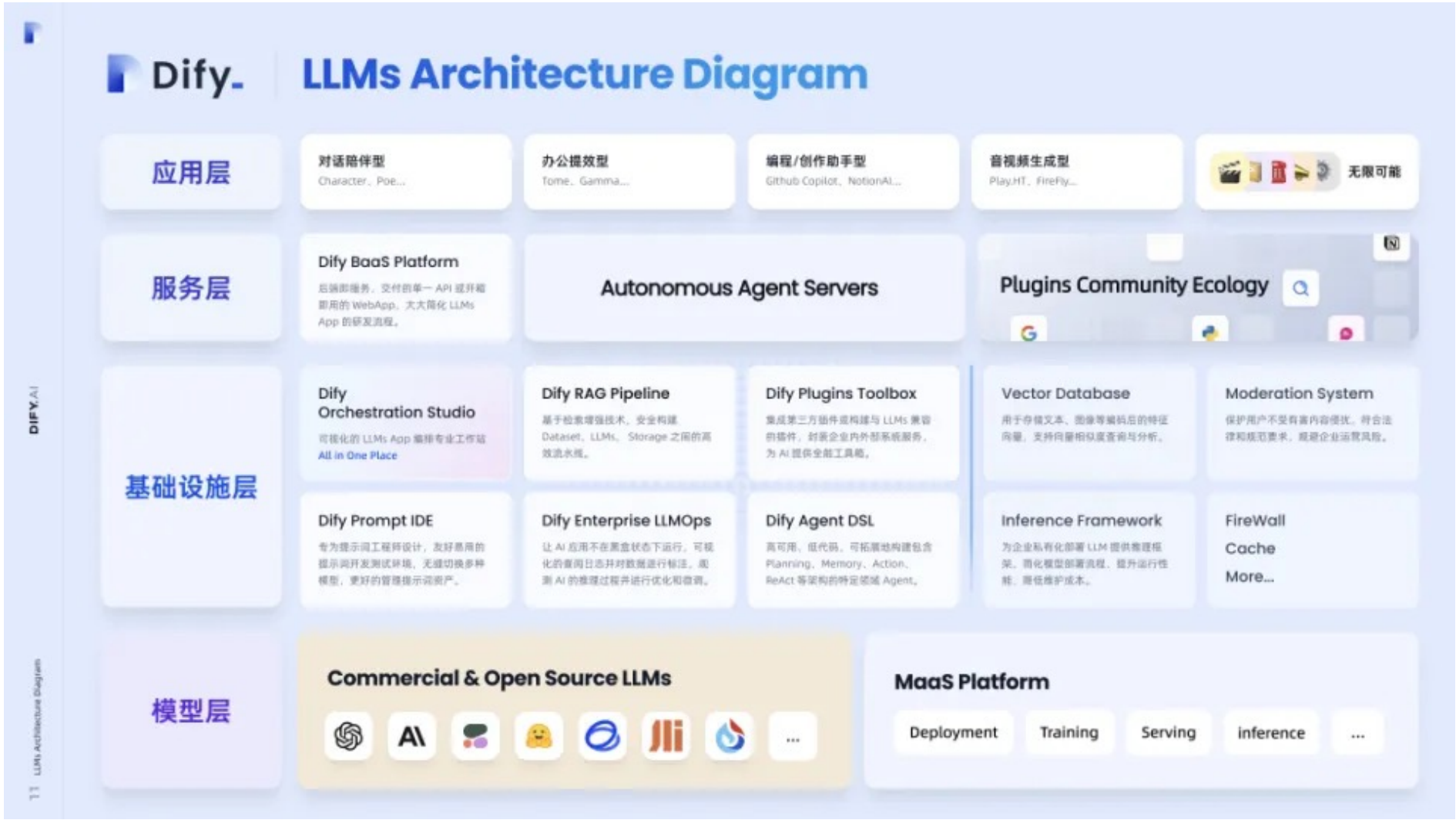
4. 多模型支持，满足多样化需求

- Dify兼容主流的AI模型，并支持自定义模型的集成，用户可以在平台内调取不同的模型服务，以实现从数据分析到自然语言处理的多样化需求。
- 模型的灵活替换和升级使平台更具适应性和前瞻性。

5. 节省成本，提高效率

- 低代码的开发模式和便捷的模型管理工具大幅降低了开发和维护成本。
- 通过自动化流程和AI驱动的智能应用，Dify帮助企业提高了工作效率，减少了人工干预，节省了运营成本。

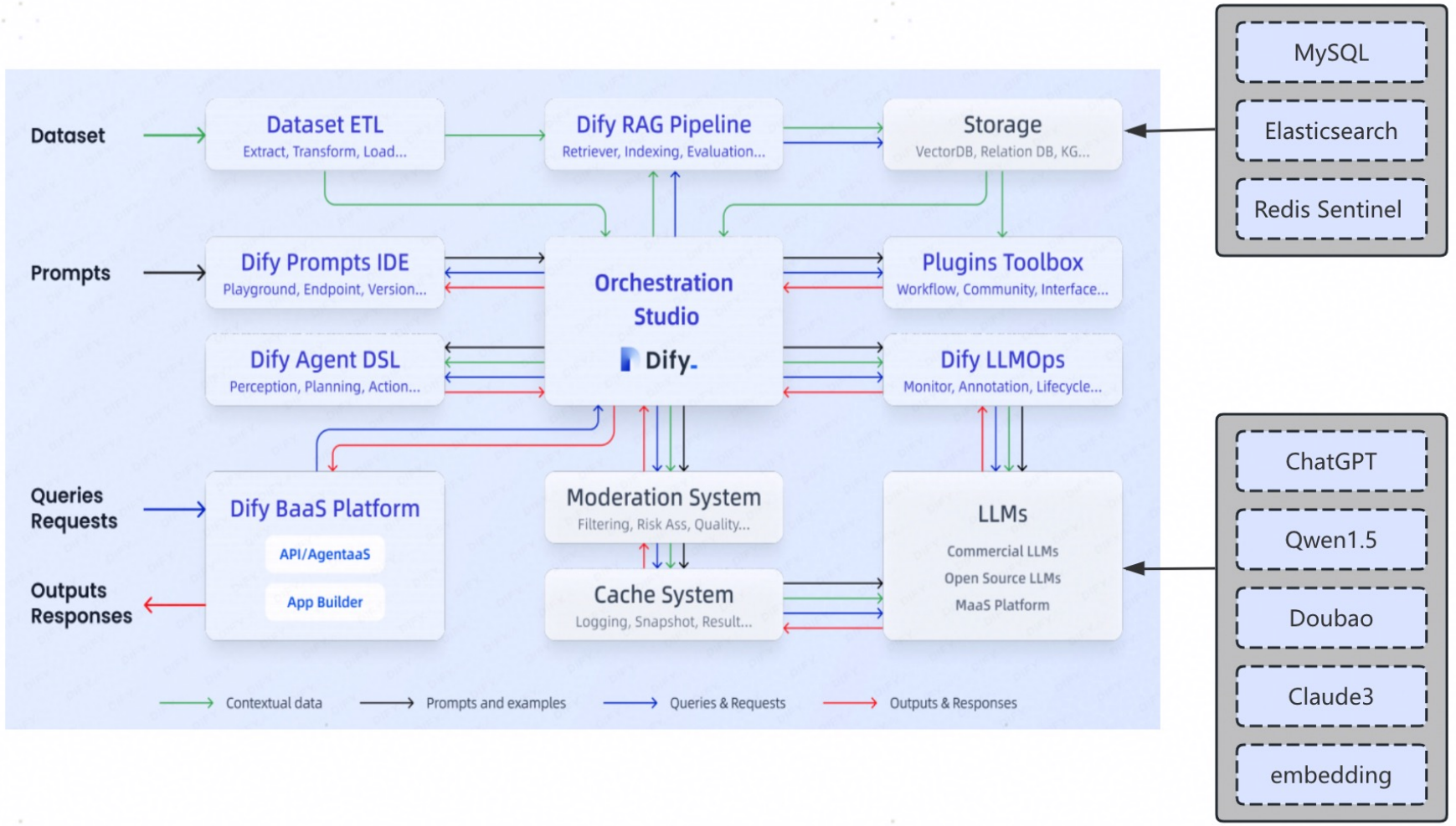
▶ 介绍 Dify 开发平台 - How



介绍 Dify 开发平台 - But?

- 元数据存储实现方式单一
- RAG能力有限，不支持最新检索算法
- 向量数据库支持不全
- 原生部署不支持高可用
- 不支持删除 message 接口
- 社区版不支持模型服务负载均衡
- ……

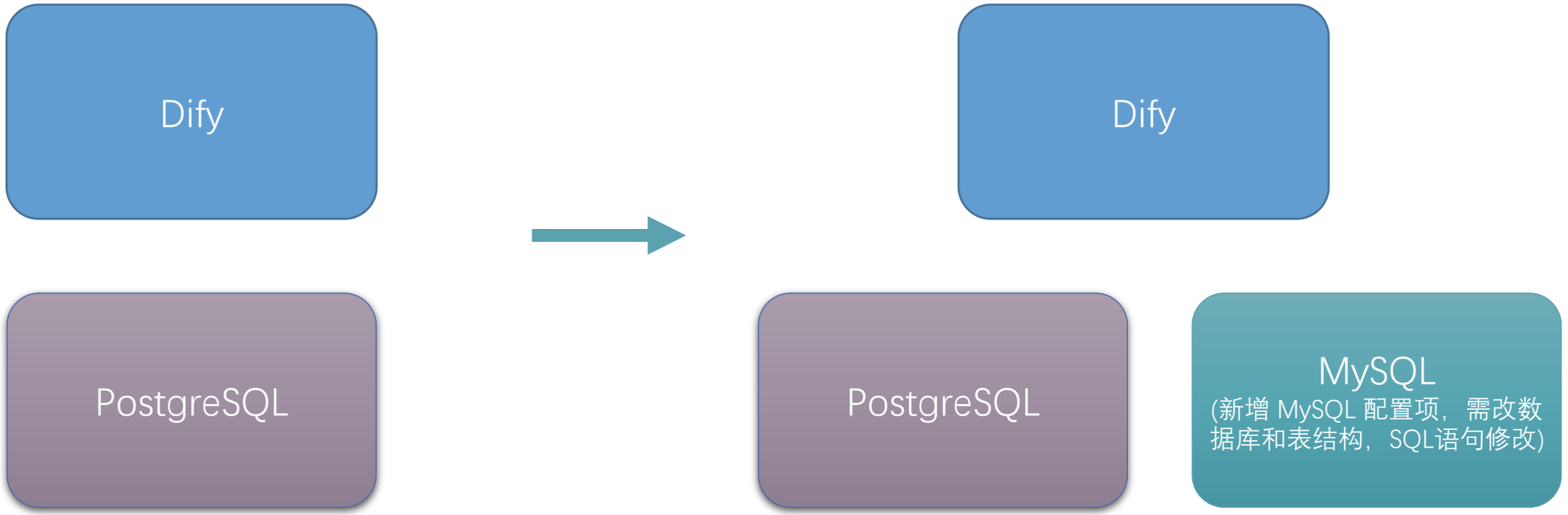
介绍 Dify 开发平台 - 顺丰集成版



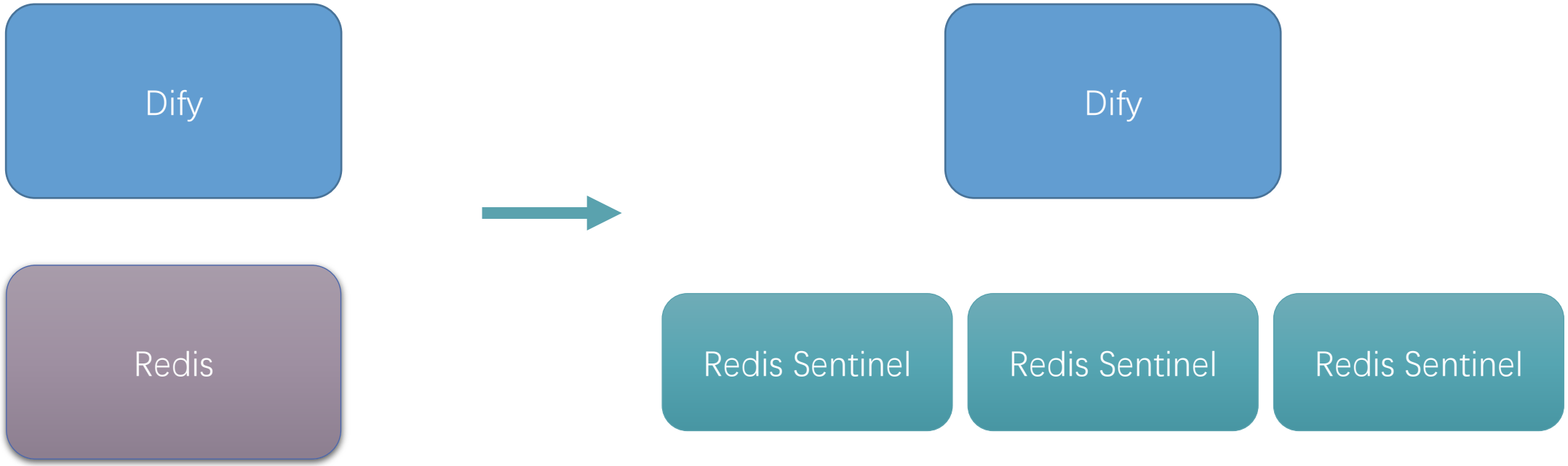
PART 02

改造 Dify 开发平台

改造 Dify 开发平台 - 部署增强

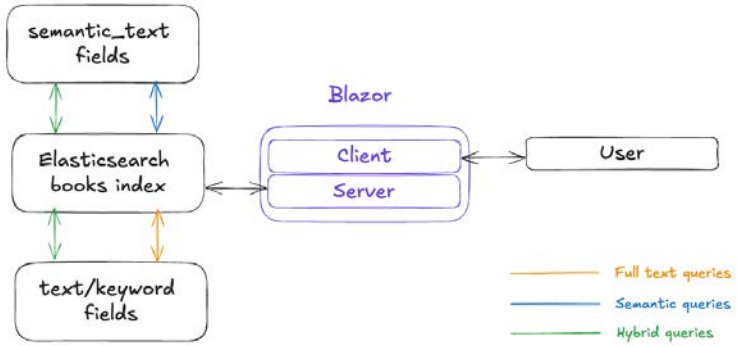


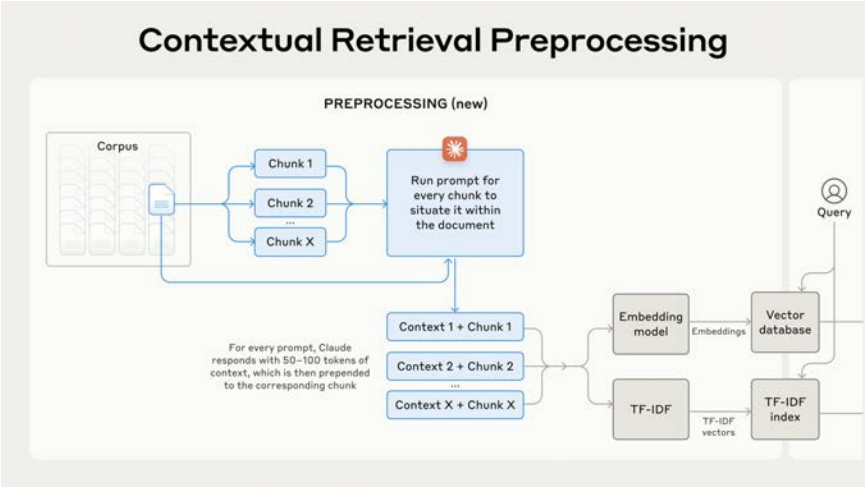
改造 Dify 开发平台 - 部署增强



改造 Dify 开发平台 - 检索增强

- 新增 ElasticSearch 8支持，支持混合检索
- 新增 ES 相关配置，支持 ES Vector 组件
- 支持 Contextual Retrieval (#8776)
- 支持 GraphRAG / LightRAG (#6019)





改造 Dify 开发平台 - 集成内部服务

集成私有化大模型

- 开源大模型
- 微调领域模型
- 商业模型服务

Dify 内置工具

- 向量混合检索工具
- 数据平台表接口查询服务

Dify 接口改造

- 修改 LLM 节点输入获取方法
- 新增删除 messages API

改造 Dify 开发平台 - Ongoing

- 更多 RAG 算法集成：GraphRAG、LightRAG 等
- 多 Agent 调度系统集成：OpenAI Swarm、MetaGPT 等
- 生产特性开发：高可用、负载均衡的模型服务支持等
- 产品化集成：对标 HiAgent 和 Coze 等商业产品
- 更多应用和流程集成：
  - 数据增强工作流
  - 全平台软件助手
  - 混合云推理服务
  - ……

PART 03

落地软件机器人 Agent 场景

落地软件机器人 Agent 场景 - 背景

RPA（Robotic Process Automation，机器人流程自动化）是一种通过软件机器人（bot）来模拟人类用户在应用程序中执行规则化任务的技术，旨在简化流程、提高效率并减少人为错误。随着大语言模型（LLM）的快速发展，将 RPA 与大模型集成，能够进一步扩展其应用场景，实现更智能的自动化。

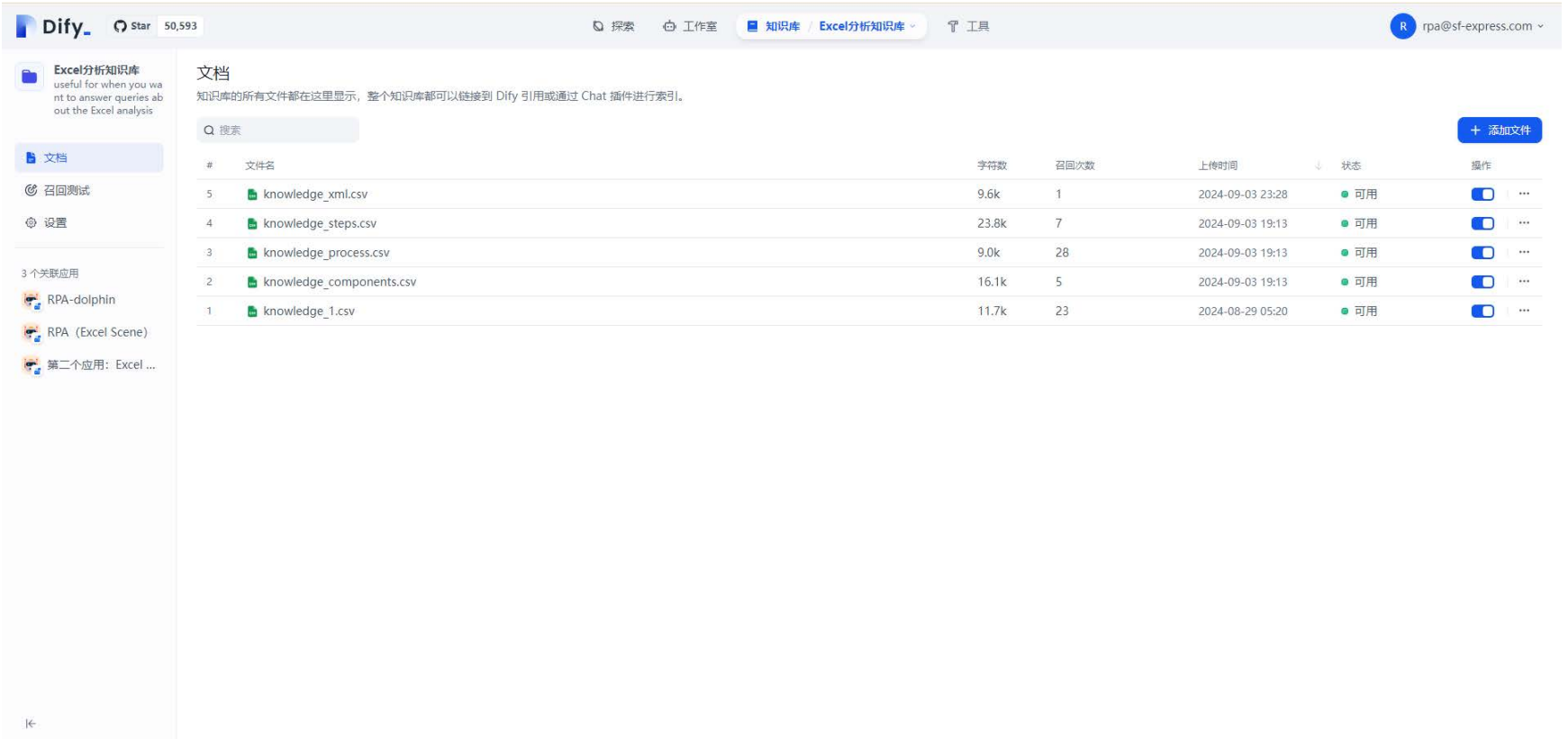
落地软件机器人 Agent 场景 - RPA vs Workflow

特性	RPA (机器人流程自动化)	工作流 (Workflow)
定义	模拟人类用户操作的自动化技术	一系列任务或活动的自动化处理
应用场景	高度重复、标准化的任务，如数据输入和迁移	复杂的业务流程，如审批流程、项目管理
技术实现	通过软件机器人与用户界面交互	通过规则和逻辑构建自动化流程
灵活性与可扩展性	灵活性较低，主要适用于标准化任务	更具灵活性和可扩展性，适应动态业务需求
用户参与	通常是完全自动化，用户干预较少	涉及多方参与者，支持协作和沟通
监控与分析	关注机器人的性能和执行状态	提供全面的监控和分析能力，对整个业务流程进行分析
集成能力	能够集成现有应用程序和系统，不需要重构	通常需要与业务系统、数据库、API 等集成

落地软件机器人 Agent 场景 - RPA with Dify

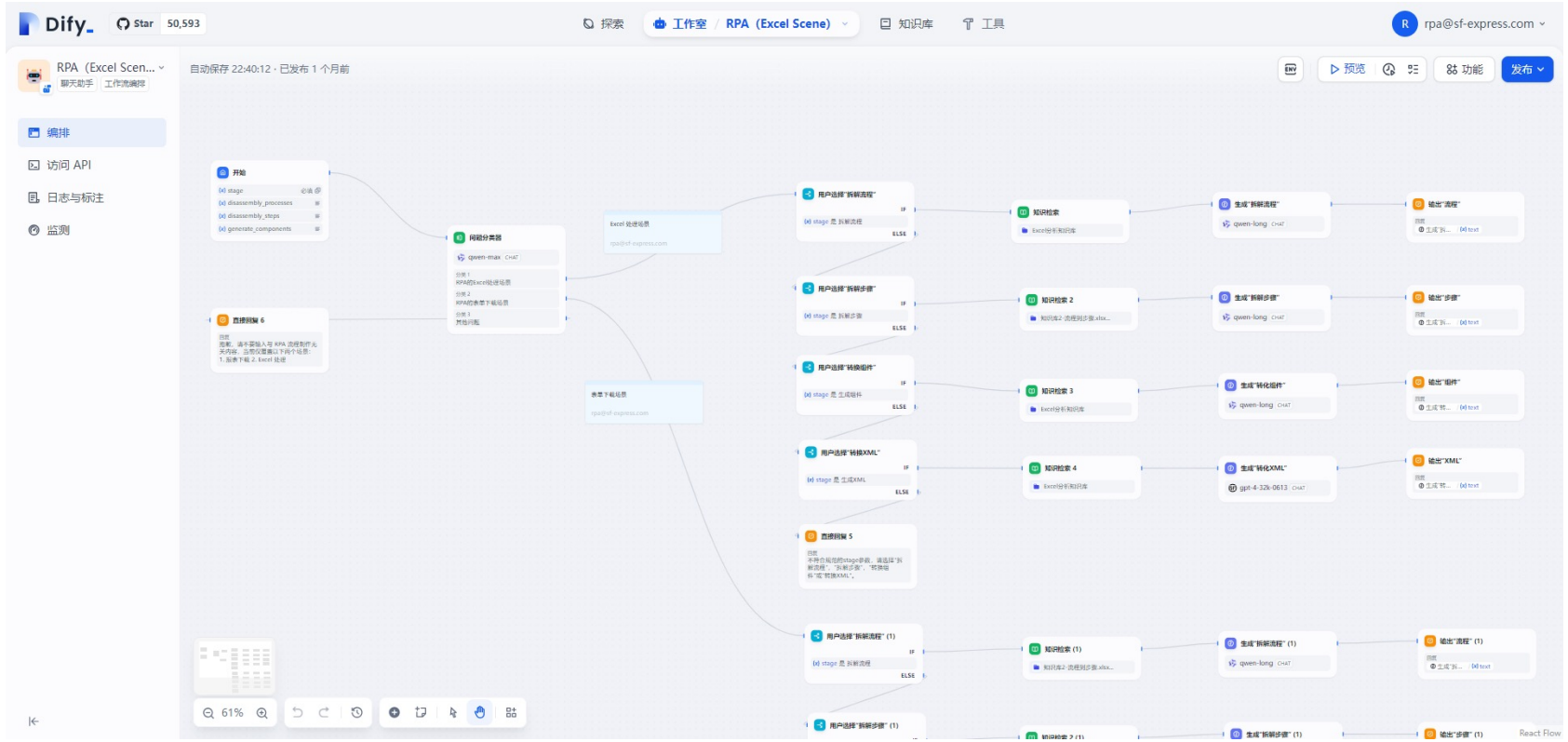
- 相比于传统工作流，Dify 大模型能力可以实现意图识别、代码生成等功能
- 通过知识库引入让大模型生成结果更加可用
- 通过微调模型让意图识别准确率和性能更好
- 通过流程拆解、步骤拆解、代码生成等多步实现人工参与调优
- 最终实现端到端自然语言到 RPA 流程的 AI 应用落地

落地软件机器人 Agent 场景 - RPA with Dify



- 知识库搭建，引入 RPA 领域知识

落地软件机器人 Agent 场景 - RPA with Dify



- 多场景工作流，实现多步任务拆解

PART 04

落地数据平台 Agent 场景

落地数据平台Agent场景 - 背景

1. 数据处理与分析

- **数据清洗与预处理**：利用 Dify 的 RAG（检索增强生成）功能，从各种文档格式（如 PDF 和 PPT）中提取文本，进行数据清洗和预处理。
- **数据分析与报告生成**：通过 Dify 的模型集成和提示词设计，自动分析大规模数据集，生成结构化的分析报告，辅助决策。

2. 智能问答与知识库构建

- **企业知识库问答**：将企业内部知识库与 Dify 集成，构建智能问答系统，提升员工和客户的查询效率。
- **客户服务自动化**：利用 Dify 的 AI 智能体功能，自动处理客户咨询，提供实时、准确的服务。

3. 数据驱动的业务流程自动化

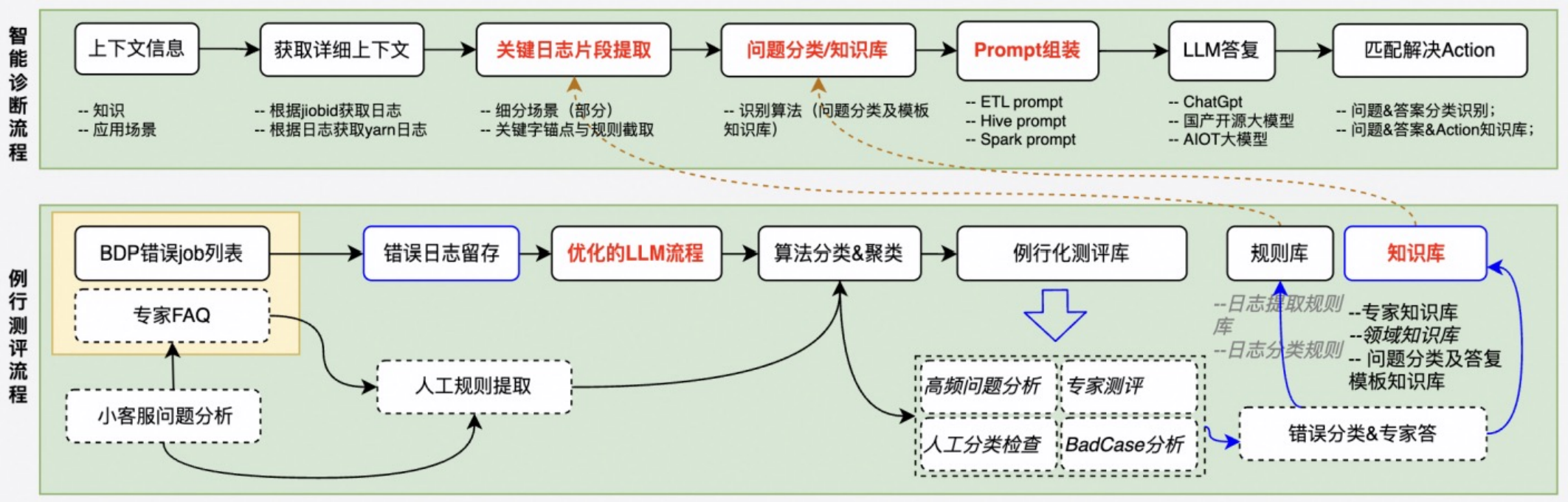
- **自动化决策支持**：结合 Dify 的 AI 智能体和工作流编排功能，实现基于数据的自动化决策流程，提升业务效率。
- **业务流程优化**：通过 Dify 的可视化工具，分析和优化业务流程，减少人工干预，降低成本。

4. 多模态数据处理

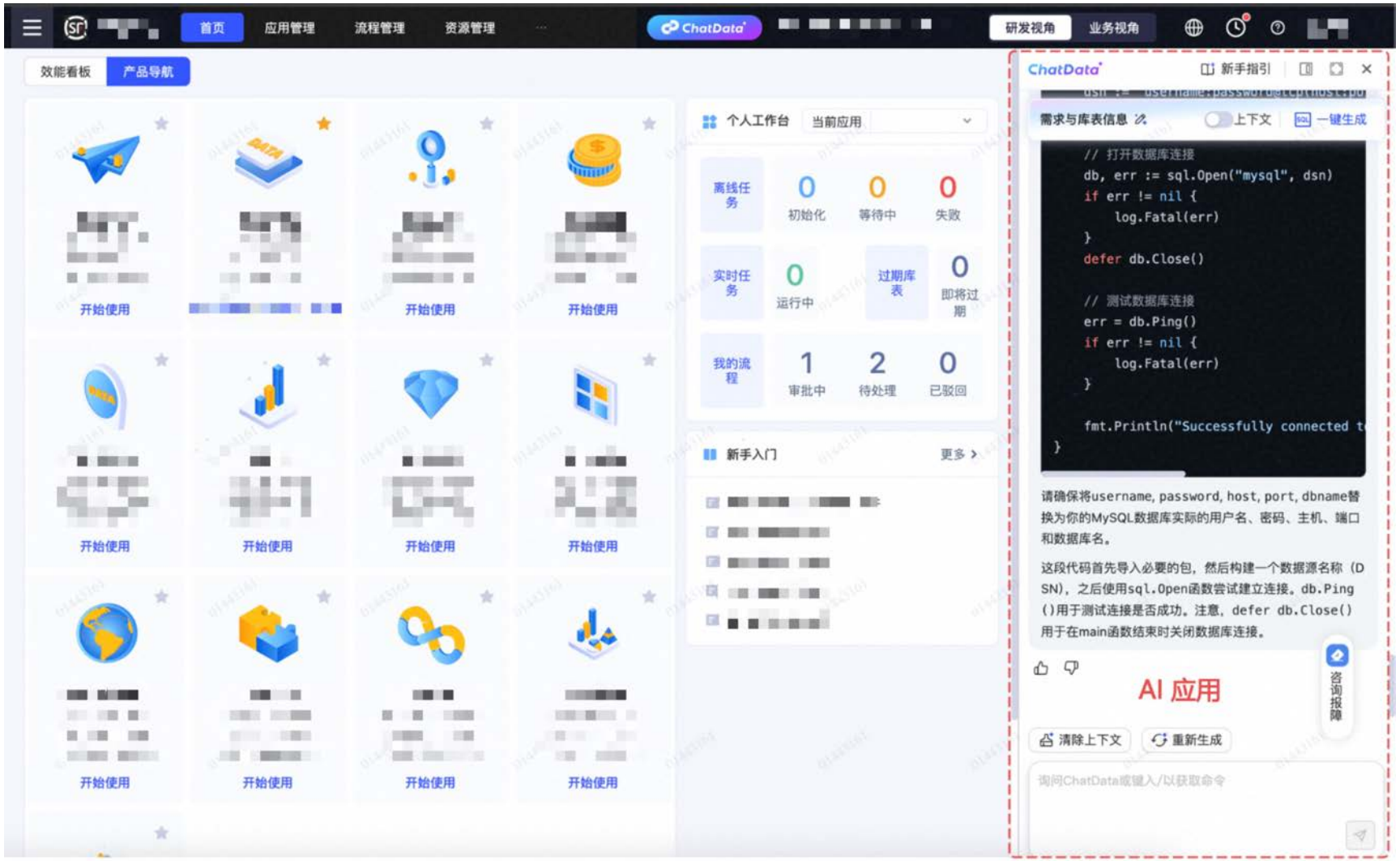
- **文本与图像数据融合**：利用 Dify 的模型支持，处理文本和图像等多模态数据，实现综合分析。
- **多语言数据处理**：Dify 支持多语言模型，适用于全球化业务的数据处理需求。

落地数据平台Agent场景 - 智能错误归因

- 智能错误归因分析



落地数据平台Agent场景 - 智能错误归因



落地数据平台Agent场景 - SQL 生成和自动查表

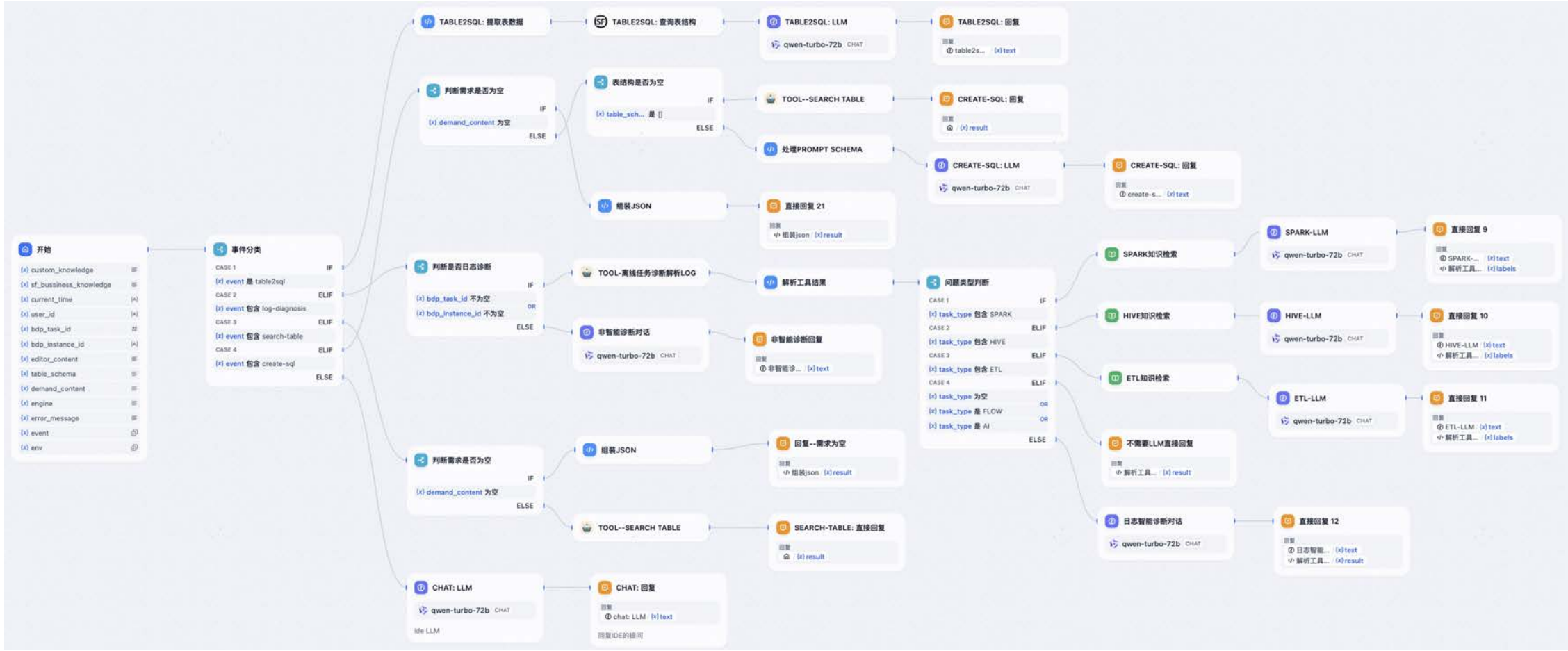
SQL 生成

- **自然语言转 SQL**：用户只需以自然语言描述所需的数据查询，大模型会解析该描述并生成相应的 SQL 查询。
- **复杂查询支持**：大模型能够处理复杂的查询逻辑，包括多表联接、聚合函数、条件过滤等，用户只需提供业务逻辑即可。
- **动态查询生成**：根据输入的参数动态生成 SQL 查询，支持灵活的查询需求，如排序、分组和限制结果集。

自动查表功能

- **表结构理解**：大模型可以解析数据库的表结构信息（如表名、字段名及其数据类型），帮助用户更好地理解数据库 schema。
- **海量表查询**：根据数据库的表属性，从数十万的表中查找业务关联的表内容，通过粗排和精排实现查表准确率和召回率。
- **智能建议**：在用户输入查询时，模型可以根据当前数据库结构提供相关表和字段的智能建议，提高查询的准确性和效率。

落地数据平台Agent场景 - SQL 生成和自动查表



PART 05

Dify 实践总结

Dify 实践总结

- Dify 在顺丰内部有海量的应用场景（数百个在线应用，几十万对话数）

开发流程	未使用 Dify 平台	使用 Dify 平台	性能提升
开发应用前&后端	集成和封装 LLM 能力，花费较多时间开发前端应用	直接使用 Dify 的后端服务，可基于 WebApp 脚手架开发	80%
Prompt Engineering	仅能通过调用 API 或 Playground 进行	结合用户输入数据所见即所得完成调试	25%
数据准备与嵌入	编写代码实现长文本数据处理、嵌入	在平台上传文本或绑定数据源即可	80%
应用日志与分析	编写代码记录日志，访问数据库查看	平台提供实时日志与分析	70%
数据分析与微调	技术人员进行数据管理和创建微调队列	非技术人员可协同，可视化模型调整	60%
AI 插件开发与集成	编写代码创建、集成 AI 插件	平台提供可视化工具创建、集成插件能力	50%

Dify 实践总结

- 未来将集成顺丰统一的模型广场
- 通过插件增加顺丰内部所有 API 市场
- 打破公有云和私有云限制，充分利用 vGPU 池化技术
- 多 Agent 调度支持，尤其是 OpenAI Swarm 深度集成
- ……

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 · 敦煌站
K+ 思考周®研习社
时间：2025.08.29-30

K+峰会 · 上海站
K+ 金融专场
时间：2025.10.17-18

K+峰会 · 香港站
K+ 思考周®研习社
时间：2025.11.25-26

K+峰会详情

AiDD峰会

AiDD峰会 · 上海站
AI+研发数字峰会
时间：2025.05.17-18

AiDD峰会 · 北京站
AI+研发数字峰会
时间：2025.08.08-09

AiDD峰会 · 深圳站
AI+研发数字峰会
时间：2025.11.28-29

AiDD峰会详情

AiDD AI+研发数字峰会 第5届
AI+ Development Digital summit

利用AI技术深化计算机对现实世界的理解

推动研发进入智能化时代