# 中软国际灵析智能体 (JointPilot) 产品说明

2024年10月

# 目录

第一章 产品介绍	1
1.1 产品目标和价值	1
1.2 产品优势	2
第二章 功能说明	4
2.1 智能体交互	4
2.1.1 理解增强	4
2.1.2 多轮对话	4
2.1.3 回答呈现	5
2.1.4 会话管理	6
2.1.5 收藏管理	6
2.1.6 报告下载	7
2.1.7 问题推荐	7
2.2 知识管理	8
2.2.1 文档知识	8
2.2.2 问答对管理	10
2.2.3 数据表文件	12
2.2.4 数据库管理	14
2.3 大模型管理	16
2.3.1 灵析 NLP 大模型	16
2.3.2 灵析问数大模型	17
2.3.3 灵析语音大模型	18
2.3.4 第三方大模型接入	19
2.4 智能体编排	19
2.4.1 流程编排能力	19
2.4.2 微应用封装	22
第三章 典型应用场景	24
3.1 数字人智能体	24
3.2 售后智能体	25
3.3 营销智能体	26
3.4 数据分析智能体	27

# 第一章 产品介绍

### 1.1 产品目标和价值

为支撑大模型应用的高效落地,打造一个功能完善、高灵活性、强延展性的智能体大脑底座,协助企业客户快速在企业运营各场景环节落地智能体应用。通过智能体应用的推广,充分释放人工智能的巨大价值,助力企业快速实现智能化转型升级,使企业能够更加灵活地应对市场变化,提升竞争力,实现可持续的发展。

#### (1) 升级企业知识体系

- 标准、快捷、易用的知识管理。整合各种文档资料,提取关键信息和知识点,构建结构化、可检索的知识库,使用户可以更有效地浏览、创造和组织知识。
- 加速知识流通和知识协作。自然语言处理技术对文档进行深度解析和实体识别,确保知识的准确性和完整性,提供友好、便捷的问答交互体验。通过个性化的知识推送,建立知识回流机制,提升知识的应用价值,创造一个正向的知识流动循环。
- 高效、精准、可视化知识探索分析。支持对非结构化知识的全方位、多维度分析,构建知识脉络,助力知识推荐和隐性知识发现,促进知识向生产力的转化。

#### (2) 打造丰富应用场景

- 企业通用场景支持。面向数字人交互、办公协作、智能客服、运营管理、市场营销、生产制造、售后运维等全场景落地智能体应用。
- 行业创新场景孵化。针对制造、医疗、教育、交通等行业场景,通过轻咨询服务,梳理行业应用场景,孵化行业特色的创新应用场景。

#### (3) 促进企业运营提效

企业流程提效。通过流程编排、组件扩展以及落地专业服务,使智能体更好地融入企业流程,促进数据和算力价值的释放,实现企业流程提效。

- 提升企业决策能力。通过整合企业各种结构化、非结构化的数据和知识,提供 灵活知识搜索、运营现状诊断、发展趋势预测等能力,为用户提供智能化的决 策支持。
- 增强用户体验和满意度。知识问答系统通过自然语言交互的方式,为用户提供 更加直观、易用的查询界面。同时,系统能够根据用户的反馈不断优化系统的 回答质量,从而提高用户的满意度。

### 1.2 产品优势



#### (1) 多模式知识底座

支持问答对、RAG、知识图谱等多模式、多元化知识管理,能够全方位、多维度地管理和呈现知识内容。基于知识图谱的知识库应用,不仅能提升知识搜索的效率和准确度,还能进一步提高专业和复杂问题场景的逻辑推理能力。

#### (2) 强大的大模型底座

为保障智能体应用在交互理解、知识处理、数据分析、语音生成等场景的高水平,灵析智能体大脑预集成行业内高水平的基座大模型。包括 Mixture of Experts(MOE)架构的万亿参数 NLP 模型,支持 245K 上下文及最高 5000 tokens/S 输入、70 tokens/S 输出的行业最高性能。以及在数据资源定位、数据分析等方面达到 85%以上准确度的问数大模型,还有高保真、超自然的语音大模型。此外,产品还支持各种第三方大模型 API 的接入管理,支持不同任务下接入不同大模型的灵活调度管理。

#### (3) 丰富组件库, 灵活可扩展性

低门槛、高效率、高灵活性的智能体流程编排引擎。通过直观易用的拖拽式界面和 强大的脚本支持,用户可以轻松实现组件间的自由组合与定制流程,构建出符合特定业 务逻辑的智能体应用。提供一个全面且不断更新的组件库,涵盖交互增强、流程增强、 算法增强、安全增强等多种类型,满足用户在不同场景下的多样化需求。系统还支持自 定义组件开发,允许用户根据需求引入外部模型或算法,不断扩展应用功能,使智能体 能够无缝集成第三方服务,实现功能的无限延伸。

#### (4) 完善的咨询和专业服务配套

为大模型智能体应用落地提供全方位、系统化的一站式服务,包括场景梳理轻咨询、基座产品、智能体开发、模型训练、集成交付以及持续的运营支持等。通过全流程服务,可以为企业客户屏蔽大模型应用的复杂性,规避大智能体应用落地的技术难题和实施风险,加速了 AI 技术在各行业的应用步伐,助力企业更快、更好、更优地实现数字化转型和智能化升级。

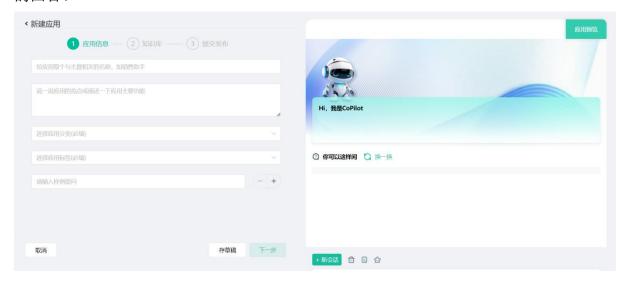
## 第二章 功能说明

### 2.1 智能体交互

### 2.1.1 理解增强

自然语言理解是智能问答服务中的一项核心功能。为了支持用户以自然语言形式提问,系统采用了先进的大语言模型和自然语言处理技术。当用户输入问题时,系统首先会对问题进行语义分析,理解用户的意图和需求。通过识别问题中的关键信息,系统能够准确地将用户的问题映射到相应的知识库或信息源中。

在理解用户意图的过程中,系统还会利用上下文信息来提高理解的准确性。例如, 当用户在多轮对话中连续提问时,系统会根据前一轮对话的内容来理解当前问题的含义。 这种基于上下文的理解方式可以帮助系统更好地把握用户的真实需求,并提供更加贴切的回答。



### 2.1.2 多轮对话

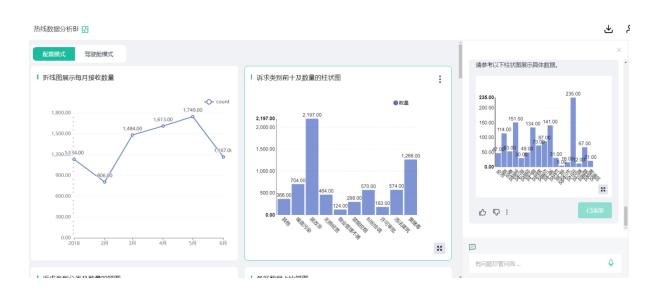
对于复杂问题或需要进一步明确需求的情况,智能问答服务提供多轮对话能力。在 多轮对话中,系统会根据用户的反馈和追加问题来逐步明确用户的需求,并提供更加详细的解答。系统采用了对话管理技术和状态跟踪机制。对话管理技术负责维护对话的流 程和状态,确保对话能够顺畅地进行下去。状态跟踪机制则负责记录用户的输入和系统的输出,以及对话过程中的关键信息。



在多轮对话中,系统还会利用自然语言生成技术来生成自然、流畅的回复。帮助用户更好地理解问题的解答过程,并引导用户进一步明确需求或提出相关问题。

### 2.1.3 回答呈现

在生成答案时,系统会根据用户的问题和需求,从知识库或信息源中检索相关的信息和知识点。然后,利用自然语言生成技术将这些信息和知识点组织成简洁明了的答案。在呈现答案时,系统会考虑用户的认知特点和信息接收习惯。对于简单的答案,系统可以直接以文本形式呈现给用户;对于复杂的答案或需要图表辅助的情况,系统还可以生成相应的图表或可视化元素来帮助用户更好地理解答案的内容。



### 2.1.4 会话管理

系统提供会话管理功能,支持对历史会话内容进行检索,同时支持清空所有对话。



### 2.1.5 收藏管理

系统支持将会话记录进行收藏,支持用户在收藏夹中对收藏内容进行检索。



### 2.1.6 报告下载

系统支持在将用户会话的信息按模板生成报告,并提供下载功能。



### 2.1.7 问题推荐

系统支持预设问题,并为用户推荐常见问题,辅助用户提问。



### 2.2 知识管理

### 2.2.1 文档知识

系统支持导入文档类知识,客户可以将 word、pdf、知识手册等作为背景知识,为大模型提供解答问题的依据。导入文档知识为大模型提供业务维度的问答能力基础,集合平台功能,可以快速搭建智能问答鞥应用。

同时系统支持对上传文档知识的自动解析,用户只需要将知识文件导入到平台中,无需额外操作,在配置好的应用中即可完成问答;知识的维护更新也是即时的,只需要将最新的知识文件更新到文档知识库中,新内容将自动应用于问答环节,确保知识的实时性和准确性。



系统支持上传 doc/pdf/xlsx 文档资料,解析成向量数据,大模型可以直接分析总结答案。



系统支持新建文档知识集, 实现按文档库对同类文档进行集中管理。



系统支持在文档知识内容管理页,上传本地的知识文件:

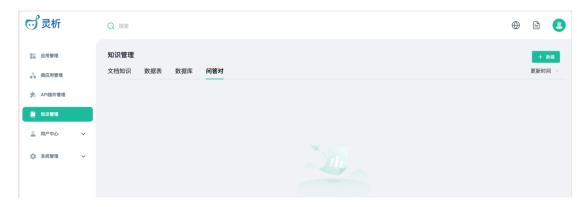


上传文件后,系统自动解析文件内容,并将文件转换成切片的文档知识,切片完成后,文档知识内容即补充完成。



### 2.2.2 问答对管理

针对基于文档片段的知识问答需求,已经能够为用户提供高效、专业的应用服务。 针对需要简洁明了的总结性回答的问题,引入了问答对知识能力。通过问答对知识库的 引用,平台可以根据问题的具体内容,自动匹配相应的知识片段或文档片段,并为每个 问题提供恰到好处的回答。这样不仅能够确保回答的准确性和专业性,还能够根据客户 的不同需求,提供更加个性化的服务。



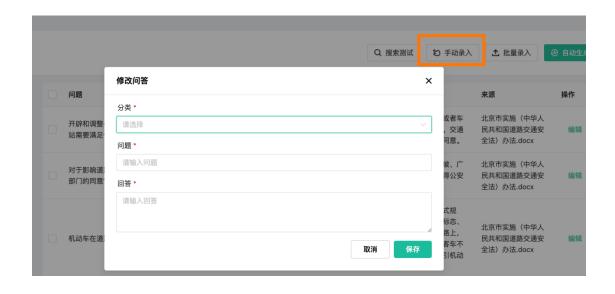
系统支持新建问答对,支持按目录维护常见问题和答案,辅助大模型检索答案更精准,系统支持大模型自动生成问答对知识。



系统支持对上传的文件,自动解析文件内容,并将文件转换成一系列问答对, 系统支持修改生成的问答对内容,包括问题和答案。



系统支持手动录入和模板导入问答对。手动录入是指当需要对问答对进行少量的添加时,可以手动创建问题和答案,这样做保持了问答对维护的灵活性,模板导入是指当需要录入多个新增问答对时,可以将这些问答对写在预置的 Excel 模板中,导入模板后,即可一次性创建多个问答对。



### 2.2.3 数据表文件

系统支持对格式化数据的管理,支持用户将格式化数据文件(如 Excel)导入到系统中,结合流程编排对数据进行工程化处理和大模型调用。

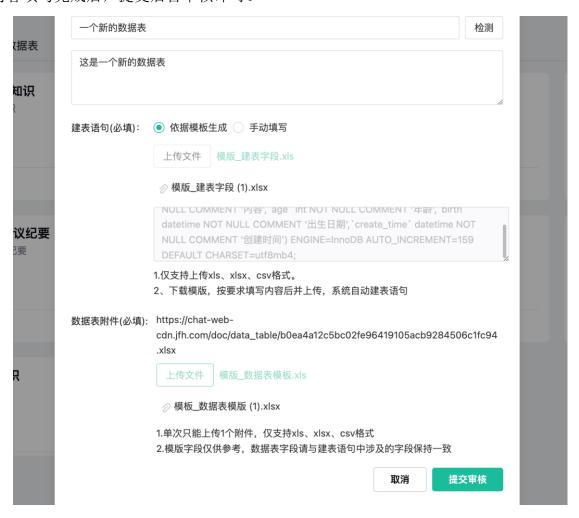
数据表的创建主要由数据本身和建表语句两部分组成,除了讲数据本身导入,同时 为了保证大模型更好地理解数据,支持对表中的字段进行说明。配置完成后,大模型可 以在保证准确率的前提下,围绕数据与用户进行自然语言交互。



系统支持在数据表知识管理界面,新建数据表;支持导入数据明细,大模型可以分析数据表结构并生成 SQL 语句来查询数据。



系统支持对表中的字段进行描述,需要对字段进行描述以下内容:字段名称、字段 类型、字段长度、是否为空、注释,支持通过上传模板文件或者手动填写补充上述信息; 内容填写完成后,提交后台审核即可。



审核通过后,系统支持维护字段的业务含义,这项信息是描述字段用处的自然语

言,支持被大模型直接识别并使用。如果此数据表在后续的使用中,会用于 NL2SQL 场景,此项描述信息需要根据实际业务情况进行填写。



实际的业务场景中可能会出现让大模型自己选择数据表的多表应用情况,这个时候 要求大模型理解每张表的业务含义;系统支持对数据表添加描述,以便于大模型对数据 表的理解。



### 2.2.4 数据库管理

平台支持连接数据库,便于处理动态数据,只需要配置数据库连接信息,即可让 智能应用调取数据,减少了数据同步的繁琐操作,同时降低了数据和应用的耦合性, 数据不会像静态数据表一样暴露出来,保证了数据的安全性。提供了功能完备的数据 库连接管理页面,用户可在此模块管理自己的数据库连接配置。



系统支持在数据库连接管理界面新建数据库。



系统支持配置数据库连接信息,并对数据库进行一些必要说明,填写完成后,通过 测试连接,可以验证连接配置。

据表	添加数据库
	数据集类型: 数据库
	数据库名称: * 输入名称
	数据库类型: * 请选择
	<b>主机: *</b> 输入主机
	端口: * 输入端口
	用户名: * 輸入用户名
	<b>密码: *</b> 输入密码
	数据库说明: * 输入数据库说明
	取消 创建

### 2.3 大模型管理

### 2.3.1 灵析 NLP 大模型

中软国际与国内领先大模型厂家合作,在灵析智能体大脑产品中预集成基于 MoE (Mixture of Experts) 架构的万亿参数 NLP 大模型。模型不仅支持 245k tokens 的上下文长度,而且支持最高 5000 tokens/S 输入、70 tokens/S 输出的响应速度,能够使模型更准确地理解语言的复杂性和语境,生成更连贯、准确和有深度的文本响应。在文字理解、逻辑推理、内容摘要、文本生成、代码生成、知识检索等场景下实现优秀的表现效果,开始接近 GPT-4、Claude-3、Gemini-1.5 等世界上最领先的大语言模型。

模型	API上下文长度 (最长)	输出速度 (tokens/s)	输入速度 (tokens/s)	TPM限制	
灵析MoE	245K	50~70	4000~5000	无限制	
Moonshot	128K	20	2000	按付费多少调整	
文心4.0	8K	10~20		单独收费	
文心-3.5-128K	128K	35~50			
通义 (qwen-max)	32K	15	2500	需申请	
GLM-turbo	128K	40~40	3000	按付费多少调整	
GLM-4	128K	25~35	1000		
百川 (baichuan4)	32K	20	1200~1300	不支持调整	
Deepseek V2	32K	10	1000	不支持调整	

更强的性能 (中文综合能力持平GPT-4)

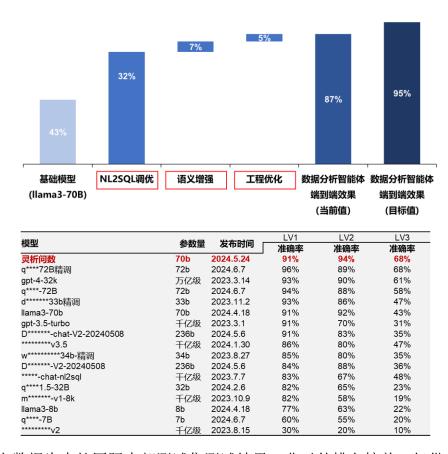
	中文综合	英文综合	知识	基础算数	数学解题	逻辑推理	编程	指令遵从
模型	AlignBench	MT-Bench	MMLU	GSM8K	MATH	BBH	HumanEval	IFEval
灵析MoE	7.97	8.82	79.5	91.7	51.4	82.0	78.0	79.6
GPT-4	7.53	8.96	84.6	92.0	52.9	83.1	67.0	85.6
GPT-3.5	6.08	8.21	70.0	57.1	34.1	66.6	48.1	70.5
Gemini 1.5 Pro	7.33	8.93	81.9	91.7	58.5	84.0	71.9	82.3
Claude 3 Opus	7.62	9.00	86.8	95.0	61.0	86.8	84.9	91.9
Claude 3 Sonnet	6.70	8.47	79.0	92.3	40.5	82.9	73.0	87.8
Claude 3 Haiku	6.42	8.39	75.2	88.9	40.9	73.7	75.9	85.4

(注:以上数据为中软国际内部测试集测试结果,非对外排名榜单,仅供参考)

### 2.3.2 灵析问数大模型

当前在企业级场景中,尽管大中型企业和政府机构都在积极拥抱和探索大模型应用,但整体上仍然面临着准确性不足、投入产出价值不明确,以及安全可靠性等多重挑战。目前大部分应用集中在非结构化数据上,但是企业相对高价值的数据资产却是集中在结构化数据中,结构化数据的处理短板是目前大模型应用落地急需解决的问题。

灵析智能体大脑预集成基于 llama3 开源大模型精调的高性能的问数大模型,通过模型调优、语义增强组件及工程化专业服务的整体方案,可以端到端提高智能体在结构化数据处理方面的水平,在数据资源定位、数据分析、结果呈现等多维度的表现达到行业一流水平。



(注:以上数据为中软国际内部测试集测试结果,非对外排名榜单,仅供参考)

### 2.3.3 灵析语音大模型

语音大模型能够实现人机之间的自然对话,使得用户可以通过语音与机器进行交互,无需使用键盘或触摸屏等物理设备。这种交互方式更加直观和便捷,能够提升用户体验。灵析智能体大脑预集成基于先进的混合专家模型训练的语音大模型,支持包括普通话、粤语、日语、韩语、西班牙语等多种语言的语音合成,能够满足不同语言环境下的语音需求。同时,生成的语音质量极高,拟人度高,情绪变化细腻自然,能够为用户提供接近真人的语音体验。支持以下产品功能:

- 多角色音频生成:支持多角色音频生成,用户可以根据需要选择不同的音色和 语调,生成符合要求的语音内容。
- 文本角色分类:该模型还提供了文本角色分类功能,能够自动识别文本中的角色类型,并为其分配相应的语音特征。

 快速复刻:用户只需提供一段简短的真人语音,就能在短时间内复刻出一段极 其逼真的 AI 声音。

### 2.3.4 第三方大模型接入

提供灵活的第三方大模型接入管理功能,旨在简化并优化各类第三方大模型 API 的接入与管理流程。通过这一平台,用户可以轻松实现大模型资源的整合、配置与监控,从而大幅提升智能体应用的灵活性和可扩展性。

#### ■ 第三方大模型 API 接入

大模型管理支持广泛的第三方大模型 API 接入,包括但不限于 NLP 大模型、CV 大模型、语音大模型等。用户只需简单配置,即可实现 API 的快速接入,无需繁琐的编码和集成工作。

#### ■ 集中化管理与配置

平台提供统一的界面,用于管理和配置所有接入的大模型 API。用户可以方便地查看每个 API 的状态、使用记录、配额限制等关键信息,并进行必要的调整和优化。

#### ■ 实时监控与告警

大模型管理中心具备实时监控功能,能够实时跟踪 API 的调用情况、响应时间、错误率等关键指标。

#### ■ 灵活的权限管理

支持细粒度的权限控制,允许用户根据团队成员的角色和职责,分配不同的访问和操作权限。

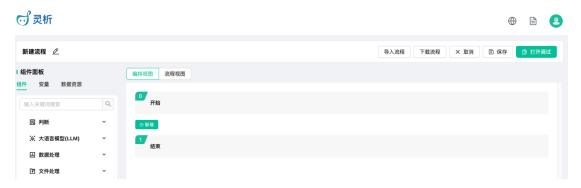
### 2.4 智能体编排

### 2.4.1 流程编排能力

平台支持自定义工作流,用于解决复杂的 AI 处理流程,调用多组件共同执行工作,

提升解决问题的效率,通过节点化处理,可以控制每个步骤的处理工作,提升处理的精确性:同时将大模型的处理过程可视化,降低用户门槛。

系统提供了组件类型丰富的微应用编排页面,用户可按需自由拖动组件进行配置:



### 2.4.1.1 LLM 组件

LLM 组件是流程编排中的核心组件,支持调用大模型能力,与用户通过自然语言进行交互。支持两种组件形态:

▶ LLM:使用大语言模型生成文本,通过 prompt(提示词)设定模型角色和任务, 让大语言模型根据用户输入生成文本。模型温度参数设定范围[0,1],数值越低, 每次生成结果越稳定,适用于精准回答场景;数值越高,每次结果不同的概率 越高,适用于创意回答场景。



➤ API 插件:基于词槽的多轮问答插件,先让大模型理解对话场景,接着获取场景中需要的词槽信息,当词槽信息不足时,将自动执行多轮问答询问词槽信息,直至调用 API 完成服务为止。API 插件开启了新的对话交互模式,通过大模型的语义理解能力和 API 的服务能力,解决了用户调用 API 服务准确性和效率不

足的问题。



### 2.4.1.2 数据处理插件

微应用调用知识库的核心组件,文档知识、问答对知识、数据表、数据库这些知识将通过这些组件与大模型进行交互。对于文档知识,可以通过插件来完成对文档切片的检索、文档知识的问答等;对于问答对知识,可以进行问答对检索;对于数据表和数据库,可以进行数据查询。通过数据处理插件将中的各类知识进行工作流处理,能让大模型根据知识回答问题,具备更强的专业能力。



### 2.4.1.3 微应用组件

当一些微应用的配置可以解决一些通用问题时,可以将微应用本身作为插件,被其他微应用调用。这样做有很多优势:首先,通过组件的复用,可以避免重复造轮

子,节省开发时间和成本。其次,组件化的微应用更容易进行模块化开发,将复杂的问题拆分成多个小问题,降低开发难度,提高开发效率。此外,组件化的微应用更加灵活,易于扩展和修改,可以快速适应变化的需求。

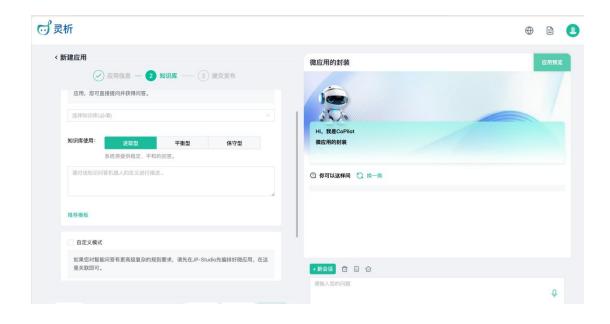
通过制定统一的微应用组件规范,不仅可以减少开发时间和成本,还可以提高产品的稳定性和可靠性,降低维护成本。此外,组件化的微应用更容易进行测试、调试和排查问题,因为每个组件都可以单独运行和检查。这可以降低故障发生的概率,提高产品的质量和用户体验。

最后,通过组件的灵活性和可扩展性,可以提供更加丰富和个性化的用户体验。用户可以根据自己的需求选择不同的组件组合,定制自己的微应用体验。这不仅可以满足用户的需求,还可以提高用户的满意度。



### 2.4.2 微应用封装

针对该项目的知识问答应用,平台提供了简单快捷的微应用封装能力,一个微应用可以通过不同的配置封装成多个应用,满足不同的使用场景,提升了应用的孵化效率。



# 第三章 典型应用场景

### 3.1 数字人智能体

针对数字化展厅、服务大厅、IOC 监控大屏等场景,集成数字人产品,提供大屏数智人、展厅数智人等智能体应用。可以提升展览互动性和趣味性,增强展览教育效果,以及优化展览管理与服务。



包括以下智能场景:

#### ■ 数字人问答

向数字人询问展厅和 IOC 的数据和知识,如政策答疑、产品介绍、服务导航、运营数据、天气情况等,支持多轮问答及打断提问继续讲解。

#### ■ 数字人讲屏

通过数字人对专题页面进行讲解介绍。可支持通过主题目标、流程定义等配置实现讲解脚本的预编排,也可以在多轮对话过程中,根据用户的实时输入动态生成讲解脚本,提升交互的体验效果,提高客户满意度。

#### ■ 数字人控屏

将数字人技术与 IOC 大屏相结合,允许用户通过自然语言与屏幕上的数字人进行交互,从而实现对屏幕内容的控制,支持专题屏切换、指标查找、缩放模型、调阅监控、设备控制等。这种技术不仅提升了用户界面的友好性和直观性,还为用户提供了更加智能、个性化的操作体验。

#### ■ 数字人告警播报

支持自然语音生成 API 指令调用,查询具体监控预警的告警数量和内容,使用数字人对监控预警事件内容进行播报,并可执行相应预案指令控制。

### 3.2 售后智能体

在售后服务场景,复刻 3~5 年资深售后工程师经验,结合大模型能力,实现产品知识问答、智能故障诊断、维修方案生成、售后服务质检等能力,全面提升售后运维质量和效率,保障客户满意度。



包括以下智能场景:

#### ■售后知识问答

基于设备使用手册、操作指导、标准规范等资料构建产品知识库,为内外部用户提供智能知识问答服务。

#### ■故障日志诊断

基于历史故障日志诊断经验构建诊断知识库,提供故障日志的智能信息提取和故障诊断功能。

#### • 维修方案生成

基于历史维修工单和经验构建维修方案知识库,针对故障问题进行信息提取分析,

生成维修方案建议。

#### ■ 售后服务质检

针对售后服务的及时性、规范性、专业性、问题解决率等方面进行综合评价,识别短板并提出改进建议。

### 3.3 营销智能体

针对传统营销方式话术单一、转化率低、经验难沉淀等问题,打造智能销售顾问智 能体,通过数据驱动客户洞察、内容推荐和营销管理,全面提升销售技能,提高营销转 化率。



包括以下智能场景:

#### ■营销客服

基于行业知识、产品介绍、营销经验等知识构建营销知识库,为内外部客户提供智能化知识学习和咨询服务。

#### ■ 智能话术

基于营销知识库提供营销话术生成服务, 支撑客服应答话术、销售话术训练、产

品推荐、营销文案等场景。

#### ■ 客户洞察

结合内部客户信息以及互联网公开客户信息、舆情信息等内容,生成客户画像及 支撑客户价值评估。

#### ■营销质检

针对销售客服的礼貌性、规范性、专业性、高效性等方面进行综合评价,识别短板并提出改进建议。

### 3.4 数据分析智能体

基于问数大模型底座,结合交互增强、推理增强、高级算法、协作安全等组件能力, 打造 AI 数据分析师 AI Agent,面向一线业务人员,降低数据分析门槛,打破数据分析 师依赖,缩短数据到决策的距离,提升企业运营决策能力。



包括以下智能场景:

#### ■ 数据分析助手

基于问数大模型底座,结合交互增强、推理增强、高级算法、协作安全等组件能力,赋能一线业务人员低门槛完成日常数据分析工作。

#### ■智能报告助手

将自然语言交互式的数据分析与传统报告功能无缝集成,将查询结果动态更新至 报告,赋能业务人员对话式生成业务分析报告。

#### ■ 大屏问数助手

将"死"大屏变成"活"大屏,可集成数字人与大屏实时互动,通过语音输入查询指令,实现即时数据分析和展示;同时可支持分钟级生成新数据专题屏

#### ■ 智能业务监控助手

通过对话方式分钟级生成专题业务仪表盘,赋能企业管理者及业务人员及时掌握业务变化,结合推理增强及高级算法组件能力,帮助用户动态分析业务现状,辅助决策。