

基于大模型的AI通话分析智能体研究与实现

廖红虹*, 赵文博, 黄莉梅, 许健君, 郭昊淞, 刘剑波

联通(广东)产业互联网有限公司COP运营中心, 广东 广州

收稿日期: 2025年11月7日; 录用日期: 2025年12月9日; 发布日期: 2025年12月16日

摘要

随着国家“数据要素×”战略的深入推进,如何释放企业海量语音通话数据的潜在价值成为关键课题。针对当前企业在语音数据利用中存在的采集渠道分散、处理模式低效及语义挖掘浅层化等瓶颈,本文提出并实现了一种基于运营商合规数据的AI通话分析智能体。该智能体依托联通云犀平台与元景大模型,构建了“智能体实时动态调度CoE (Collaboration of Experts)”引擎和AI通话分析智能体。CoE引擎通过任务规划与多模型混合调度机制,实现了从语音采集、高精度转写、多视角语义理解到结构化价值输出的全链路自动化。AI通话分析智能体实现了基于利益相关者理论的多视角标签体系,以适应不同业务角色的分析需求。对比实验与实证分析结果表明,本文提出的方法在关键意图识别的F1分数上显著优于传统基线模型;在某物流企业的实际应用中,推动订单揽收率从62%提升至92%,投诉处理时效由48小时压缩至15分钟。本文方法为企业通信数据要素的资产化与市场化流通提供了可复制的技术范式。

关键词

数据要素, 语音分析, 智能体, CoE动态调度, 大语言模型, 企业数字化转型

Research and Implementation of AI Voice Call Analysis Agent Based on Large Model

Honghong Liao*, Wenbo Zhao, Limei Huang, Jianjun Xu, Haosong Guo, Jianbo Liu

COP Operations Center, China Unicom (Guangdong) Industrial Internet Co., Ltd., Guangzhou Guangdong

Received: November 7, 2025; accepted: December 9, 2025; published: December 16, 2025

Abstract

With the deepening implementation of the national “Data Elements×” strategy, unlocking the latent value of massive enterprise voice call data has become a critical imperative. Addressing the bottlenecks in current data utilization—such as fragmented collection channels, inefficient processing

*通讯作者。

文章引用: 廖红虹, 赵文博, 黄莉梅, 许健君, 郭昊淞, 刘剑波. 基于大模型的 AI 通话分析智能体研究与实现[J]. 计算机科学与应用, 2025, 15(12): 265-273. DOI: 10.12677/csa.2025.1512342

modes, and shallow semantic mining—this paper proposes and implements an AI-powered voice call analysis agent based on compliant telecom operator data. Leveraging the China Unicom Yunxi Platform and the Yuanjing Large Model, this paper constructs an “Intelligent Agent Real-time Dynamic Scheduling CoE (Collaboration of Experts)” engine and an AI-powered voice call analysis agent. The CoE engine implements task planning and multi-model hybrid scheduling mechanisms. This engine achieves end-to-end automation encompassing voice collection, high-precision transcription, multi-view semantic understanding, and structured value output. The AI-powered voice call analysis agent constructs a multi-view tagging system based on stakeholder theory, which is introduced to adapt to the analytical needs of various business roles. Comparative experiments and empirical analysis demonstrate that the proposed solution significantly outperforms traditional baseline models in the F1 score for key intent recognition. In a practical application within a logistics enterprise, the solution drove an increase in the order pickup rate from 62% to 92% and reduced customer complaint resolution time from 48 hours to 15 minutes. This research provides a replicable technical paradigm for the assetization and market circulation of enterprise communication data elements.

Keywords

Data Elements, Speech Analytics, Intelligent Agent, CoE Dynamic Scheduling, Large Language Model, Enterprise Digital Transformation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

数字经济浪潮正席卷全球，数据已正式成为继土地、劳动力、资本、技术之后的第五大核心生产要素[1]。在这一宏观背景下，中共中央、国务院相继出台《关于构建更加完善的要素市场化配置体制机制的意见》[2]及《“数据要素×”三年行动计划(2024~2026年)》[3]，将数据要素的价值释放提升至国家战略高度。其核心要义在于推动数据从“资源”向“资产”的转变，并深度融入生产、分配、流通、消费等各个环节，以数据流激活和倍增其他要素的价值，成为推动经济社会高质量发展的关键引擎。

在企业产生的海量数据中，语音数据，特别是源自客户服务、市场营销、运营支持等场景的通话录音，无疑是价值密度最高的“黄金数据”之一。作为企业与客户最直接、最即时的交互载体，语音数据不仅包含了显性的业务信息(如客户需求、订单状态、交易意图)，更蕴含着丰富的隐性信号，包括客户情绪、服务态度、潜在不满、风险预警及未被满足的商机[4]。然而，长期以来，这种“黄金数据”的价值却普遍处于“沉睡”状态。

当前，企业在利用语音数据方面普遍面临三大严峻挑战，构成了“数据要素”价值释放的瓶颈：

(1) **数据采集的“碎片化”与“合规性”困境**：企业的通信渠道日益多元，通话数据分散在呼叫中心系统、员工个人手机、VoIP工具等不同“孤岛”中。缺乏统一、全量的采集机制，导致数据源不完整、不一致。同时，如何在保障数据安全与个人隐私合规的前提下进行有效采集，是企业面临的首要难题。

(2) **数据处理的“原始化”与“低效率”瓶颈**：面对海量的非结构化语音流，传统的处理方式严重依赖“人工抽检”与“人工回听”。这种方式不仅耗费巨大的人力成本，且抽样覆盖率极低(通常不足5%)，导致质检结果存在显著的统计偏差，无法客观反映服务全貌。更重要的是，这种事后的“亡羊补牢”的模式，效率低下，无法满足业务实时干预与决策的需求。

(3) **数据挖掘的“浅层化”与“低价值”难题**：即便完成了语音转写，企业也往往停留在简单的关键

词检索层面。传统的质检系统多基于固化规则(Rule-based), 难以捕捉深层语义、上下文关联及复杂的用户意图。企业无法从海量通话中提炼出可用于驱动业务决策的结构化洞察, 导致数据价值挖掘严重不足。

为突破上述瓶颈, 响应国家“数据要素×”战略号召, 本研究聚焦于运营商所掌握的高质量、高覆盖率、强场景属性的语音通信数据, 提出并实现了一种基于大语言模型的 AI 通话分析智能体。该智能体依托中国联通云犀平台[5]提供的合规数据采集能力, 并深度融合联通元景大模型[6]的强大语义理解与推理能力, 旨在构建一个从“语音采集-智能转写-语义理解-标签标注-价值应用”的全链路、自动化、闭环式的数据要素赋能体系。

本研究的核心贡献在于, 它不仅提供了一个技术解决方案, 更探索了一种将非结构化语音数据转化为可度量、可管理、可流通的结构化数据资产的创新模式。该智能体通过技术创新与业务模式重构, 致力于打通语音数据从“沉睡资源”到“增长引擎”的转化链路。本文将系统性地阐述该智能体的分层技术架构、以“智能体实时动态调度 CoE”为核心的创新机制、在物流与金融等行业的实证应用成效, 并深入探讨其作为数据要素市场化配置与商业化推广范例的路径与价值。

2. 相关工作

对通话语音数据的分析与利用, 是一个涉及信号处理、声学建模、自然语言处理(NLP)及业务应用等多学科交叉的复杂领域。其研究与应用历程, 与人工智能技术的发展深度绑定, 大致经历了三个主要阶段。

2.1. 传统语音处理与关键词检索

在深度学习浪潮之前, 语音数据分析的研究主要集中在基础的语音识别(Automatic Speech Recognition, ASR)技术。早期的 ASR 系统多依赖于高斯混合模型(GMM) [7]和隐马尔可夫模型(HMM) [8]等统计方法, 这些方法在特定词表和安静环境下尚可取得一定效果, 但在高噪声、多口音、大词汇量的真实通话场景中, 其识别准确率往往难以满足商用需求。在应用层面, 受限于 ASR 的精度和 NLP 技术的匮乏, 通话分析主要停留在基于关键词提取的“检索”阶段[9]。例如, 呼叫中心使用该技术来检索是否出现“投诉”、“退款”等特定词汇, 其分析维度单一, 且无法理解上下文语义, 价值释放极为有限。

2.2. 基于深度学习的智能语音分析

2010 年以后, 以深度神经网络(DNN)为代表的深度学习技术为声学建模带来了革命性突破[10]。后续的 RNN、LSTM 及其变体, 极大地提升了 ASR 系统在复杂声学环境下的鲁棒性与准确率[11][12]。随着 ASR 技术的成熟, 研究重点转向了对转写后文本的自然语言处理。

在这一阶段, 智能质检系统(Smart Quality Inspection)成为语音分析最成熟的商业化场景[13][14]。这些系统开始应用 NLP 技术进行意图识别、情感分析和话题聚类。例如, 通过情感分析模型判断客户的满意度或愤怒程度, 通过意图识别模型判断客户来电的核心目的。然而, 这些系统多基于浅层模型或独立的、针对特定任务(如情感、意图)训练的小模型。它们虽然优于纯规则引擎, 但仍存在两大局限: 一是泛化能力差, 模型对业务场景的细微变化(如新话术、新产品)非常敏感, 需要持续的、高成本的模型重训; 二是能力孤立, 缺乏对长篇对话上下文的综合理解能力, 难以实现“全流程”的深度分析。

2.3. 大语言模型(LLM)驱动的智能体范式

近年来, 以 Transformer 架构[15]为基础的大语言模型(LLM)的出现, 彻底重塑了 NLP 领域的技术范式。LLM 凭借其庞大的参数量、海量数据的预训练以及强大的上下文理解(In-context Learning)、语义表征和零/少样本(Zero/Few-shot)推理能力, 为通话数据的深度分析提供了前所未有的机遇。

LLM 不再是“单点”解决 ASR、情感、意图等割裂的任务，而是有能力对整段对话进行端到端的综理解与生成[16]。这催生了“智能体(Agent)”这一新概念。智能体不仅能“理解”对话内容，还能“行动”——例如，自动生成通话摘要、提炼关键实体、评估服务质量、标注风险等级，甚至自主调用外部工具(如 CRM 系统)来完成特定任务[17]。

2.4. 本研究的定位与创新

尽管 LLM 展现了巨大潜力，但将 SOTA (State-of-the-art)技术转化为可靠、高效、可落地的企业级解决方案，仍面临诸多挑战。传统智能质检工具显然已无法满足数据要素化时代对数据深度挖掘与多维赋能的需求。

本研究的独特性与创新性主要体现在以下几点：

(1) 数据来源的根本性突破：以往研究多基于企业内部呼叫中心数据，本研究依托运营商的“工作手机号卡”，实现了对企业(尤其是外勤、销售等岗位)与客户真实业务通话的全量、合规采集，从源头保证了数据要素的高质量与完整性。

(2) 技术架构的范式跨越：本研究方案区别于传统基于规则或浅层模型的质检工具，构建了一个以自研“元景大模型”为核心 AI 引擎，并辅以“智能体实时动态调度 CoE”实现 AI 任务分解与多模型混合调用机制。实现了从“单点质检”到“全流程智能分析”的跨越，能够动态编排 ASR、NLP、多维建模等多种 AI 能力，以适应复杂多变的业务场景。

(3) 价值定位的全面升级：本研究的目标并非仅为企业提供“降本增效”工具，而是构建一个“语音数据资产池”。通过将非结构化的通话转化为结构化、标签化的数据资产，使之成为可检索、可分析、可驱动决策，乃至可流通交易的数据产品，践行了“数据要素×”的核心理念。

综上所述，本研究在融合前沿大模型技术、创新运营商数据应用模式以及探索数据要素化实现路径上，进行了系统性的尝试与实践。

3. AI 通话分析智能体架构与实现

本智能体系统采用分层设计，自下而上分为数据层、模型能力层、智能体层和应用层，确保了系统的可扩展性、高可用性与智能化水平。详细架构如图 1 所示。

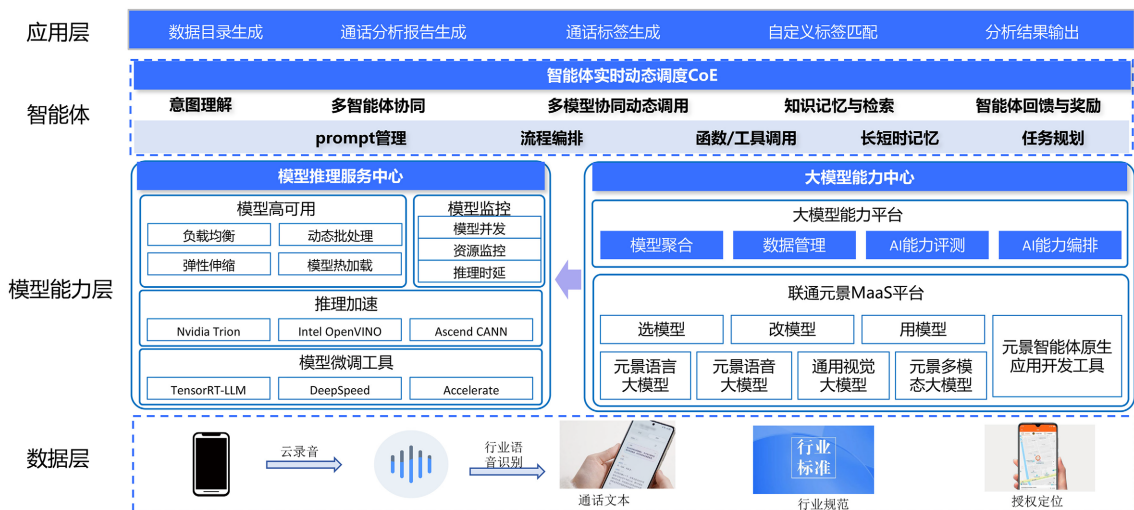


Figure 1. The architecture for the proposed AI-powered voice call analysis agent
图 1. 本文提出 AI 通话分析智能体架构

3.1. 数据层：高质量数据要素的合规采集

数据是智能体的基石。本方案依托中国联通云犀平台的“工作手机号卡”作为数据采集入口，实现对员工与客户真实业务通话的全量、自动化采集。方案的合规性通过技术与流程双重保障：**(1) 流程授权**：在员工“工作手机号卡”的入网协议及企业服务合同中，明确告知数据将被用于 AI 辅助分析与质检，并获取相应授权；**(2) 技术加密**：采集到的原始语音数据及元数据(如通话时长、双方角色等)，在传输、存储、处理的全链路中均采用高强度加密，并对客户敏感信息(如身份证号、银行卡号)进行自动化的“去标识化”处理，确保数据安全与隐私合规。

3.2. 模型能力层：强大的 AI 引擎支撑

模型能力层是智能体的“大脑”，由模型推理服务中心和大模型能力中心构成。

● **大模型能力中心**：基于联通元景 MaaS 平台[6]，提供选模型 - 改模型 - 用模型全栈大模型研发体系，构建联通特色的大模型工具链范式，一站式大模型应用开发；提供自定义智能体 + workflow 开发工具，实现低门槛应用开发。平台工具链支持：**(1) 在选模型方面**：提供模型对比、模型推荐、模型排行等工具，对大模型进行效果对比，用于用户进行基本模型的选择。**(2) 在改模型方面**：提供模型管理工具。支持用户基于基础大模型进行预训练、SFT 训练、纠偏等训练方式，支持模型压缩、模型评估。提供完善的数据管理及数据处理工具，支持多版本数据集管理、数据标注；在数据处理方面支持数据体检、数据清洗、数据增强、数据回流等功能。**(3) 在用模型方面**：提供零代码智能体开发工具，零代码使用智能体工具可快速创建智能体及插件使用和发布。提供画布式 workflow 开发工具，支持 HTTP、知识库、MCP、分支器、代码等常用节点，节点类型持续扩增。提供 MCP 广场功能，支持部分 MCP 通过 SSE 方式接入智能体活 workflow 中使用。

● **模型推理服务中心**：基于联通元景 MaaS 平台，提供模型微调技术支撑、模型推理加速及软硬件适配(尤其是国产芯片的适配)、模型高可用、模型应用监控与审计等模型推理服务。

3.3. 智能体层：动态协同与任务编排

智能体层核心是“智能体实时动态调度 CoE (Collaboration of Experts)引擎”。该引擎能够根据业务场景的实时变化，动态决策并优化调度 ASR 转写、NLP 理解及多维建模等 AI 能力组合。它具备多智能体协同、长短时记忆、知识检索与流程编排等能力，实现了从“语音采集 - 内容理解 - 价值输出”的全流程智能化与自动化，确保了分析任务的高效、精准执行。

CoE 引擎它的工作原理如图 2 所示，其核心步骤如下：

步骤 1：接收并解析 MCP 上下文请求

CoE 系统接收一个符合预设模型上下文协议(MCP)的请求。该 MCP 上下文包是一个结构化的数据对象，至少包含：

context_id：本次交互的唯一标识。

user_profile：用户身份、偏好、权限等信息。

session_history：历史交互记录。

goal：本次任务的明确目标描述(例如，“查询语音订单状态并预警潜在风险”)。

available_tools：一个描述所有可用“专家”能力的列表。每个专家能力(或称工具)都以标准格式描述，包括 *tool_id*，功能描述，输入/输出模式(schema)等。

步骤 2：中央规划模型进行任务规划

将接收到的完整 MCP 上下文包输入至一个作为系统“大脑”的中央规划模型(本文为联通元景 72B 大语言模型)。该规划模型执行以下操作：

意图理解: 深刻理解 goal 字段中的用户意图。

任务分解: 基于对 available_tools 的理解, 将复杂目标分解成一个或多个可执行的子任务步骤。

生成执行计划: 生成一个结构化的执行计划(plan), 并将其写入(或更新)到MCP 上下文包中。该计划明确了每一步骤需要调用的工具和预期的中间结果。

步骤3: 调度并执行专家工具

系统的调度执行模块根据 MCP 上下文包中的 plan, 执行当前步骤:

选择工具: 从 available_tools 中选择计划要调用的目标专家工具。

参数生成: 指示规划模型(或一个独立的模型)为调用该工具生成必要的输入参数。

执行调用: 安全地调用目标专家工具的 API 接口, 并获取返回的执行结果。

步骤4: 更新并迭代演化 MCP 上下文

将上一步获取的执行结果作为观察信息(observation), 更新至 MCP 上下文包中。这一步至关重要, 它记录了任务执行的中间状态。

步骤5: 循环迭代直至任务完成

将更新后的 MCP 上下文包再次送回中央规划模型(返回步骤2)。规划模型根据新的观察信息, 评估任务进展, 并决定下一步行动:

如果任务未完成, 则重新规划, 继续生成下一步的行动指令。

如果任务已完成, 则生成最终的响应结果。

这个“规划-执行-观察-再规划”的循环, 构成了对 MCP 上下文的持续迭代演化, 直至目标达成。

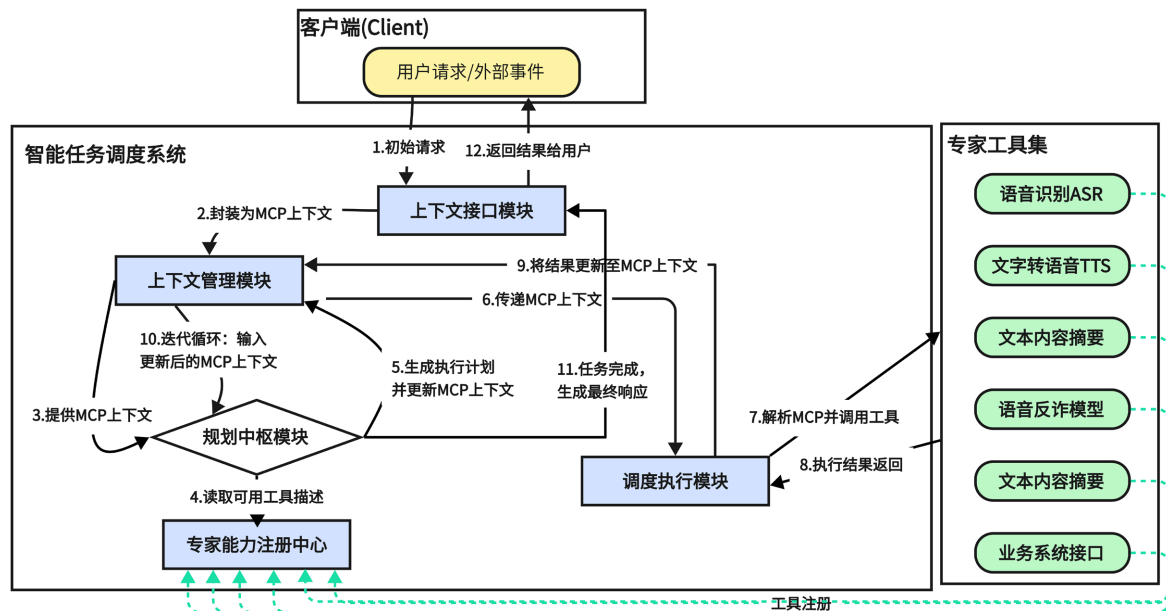


Figure 2. The architecture for the proposed CoE agent

图 2. 本文提出 CoE 智能体架构

3.4. 应用层: 场景化价值输出

应用层直接面向企业用户, 提供可视化的功能模块与数据服务, 包括通话标签生成、自定义标签匹

配、分析报告生成与结果输出等。企业可根据自身业务需求，灵活配置分析维度与标签体系，将分析结果无缝嵌入 CRM、质检等现有业务系统，实现数据价值的闭环。

4. 核心创新与机制突破

4.1. 技术创新：基于多视角标签体系的场景化适配

本方案基于联通云犀平台和元景 MaaS 平台，研制了“智能体实时智能调度 CoE 技术底座”。在此基础上，将“利益相关者理论”工程化，构建了“多视角分析框架”。

传统质检系统通常只有一套固定的、面向“管理者”的“合规”标签。本文研究发现，同一段通话对于不同角色的价值是不同的。例如，对于同一段物流客服通话：

对管理者(视角 1)：智能体重点分析“服务态度”、“是否使用标准话术”、“是否出现‘虚假签收’风险词”。

对营销部门(视角 2)：智能体重点分析“客户是否提及竞品”、“客户是否表达了新的寄件需求”、“客户画像(如高价值客户)”。

对员工(视角 3)：智能体可自动生成“优秀话术总结”和“待改进点”，作为赋能工具。

通过 CoE 引擎对元景大模型的灵活调度，结合自定义标签库，实现了对同一数据源的多维度、多场景的价值挖掘，极大提升了数据要素的复用价值。

4.2. 模式创新：数据要素化与流通机制探索

本文提出的 AI 通话分析智能体应用形成了以运营商为主导的“运营商 + 企业”合作型数据采集新范式。该智能体实现了将通话数据结构化、标签化，沉淀为可被检索、分析、决策驱动的“语音数据资产池”。原始的、非结构化的 *audio.wav* 文件，通过 CoE 引擎输出的结构化 JSON 对象，输出为标准化的、可机读的 *analysis_result.json*，其中包含了 *{intent: "投诉", emotion: "愤怒", tags: ["虚假签收", "超时"], summary: "..."}* 等字段。这些以“分析标签包”等形式存在的标准化数据产品，为数据要素的市场化配置提供了实践样本。

5. 应用成效与实证分析

5.1. 实验评测：技术性能对比分析

为验证本 AI 通话分析智能体在核心技术指标上的优越性，我们设计了对比实验，以证明其相较于传统方案的先进性。

5.2 应用成效与价值分析

AI 通话分析智能体已在物流、金融、政务等多个行业的头部企业成功落地，验证了其显著的降本、提质、增效价值。

(1) 评测数据集：我们从某物流合作伙伴的真实通话数据中，随机抽取 2000 条通话录音。由 3 名专业的标注人员进行人工精标，构建了评测基准数据集。该数据集覆盖了“网点揽收及时率、沟通合规率、客诉风险指数”3 个关键场景的 35 项指标。

(2) 基线对比方案：1) **规则 + 关键词：**模拟传统质检系统。采用 ASR 转写后，使用预定义的关键词(如“投诉”、“签收”)进行硬匹配。2) **ASR + 小模型：**模拟主流 AI 质检方案。采用 ASR 转写后，使用独立的、基于 BERT 的 NLU 模型分别进行意图识别和情感分析。

(3) 评测指标：我们重点关注系统在关键信息提取上的 F1 分数(F1-Score)，该指标兼顾了精确率

(Precision)和召回率(Recall)，是评估此类系统的核心标准。

(4) 实验结果与分析

Table 1. Experimental results

表 1. 实验结果

评测任务	评测指标	规则 + 关键词	ASR + 小模型	AI 通话分析智能体
ASR 转写	词错误率(WER)	--	98.2%	98.2%
网点揽收及时率	F1-Score	--	62%	92%
沟通合规率	F1-Score	43%	82.5%	95%
客诉风险指数	F1-Score	62%	75%	95%

从上表 1 可见，规则 + 关键词在需要理解上下文的沟通合规率等标签上，召回率极低，导致 F1 分数不理想。ASR + 小模型表现较好，基本上处在可用的水平，但在复杂意图和上下文关联上仍有不足。本文提出的智能体方案，凭借元景大模型的强大上下文理解和推理能力，以及 CoE 引擎的动态调度，在所有关键标签的 F1 分数上均显著优于两个基线方案，证明了其在技术上的先进性。

5.2.1. 降本增效：重塑运营成本结构

通过全自动的语音转写与智能标注，智能体替代了传统的人工质检模式，实现了 100% 全量通话覆盖，将人工从重复性劳动中解放出来，直接降低了人力成本。在某物流项目中，服务质控效率实现了超过 200% 的提升。

5.2.2. 服务提质：驱动精细化管理与风险预警

智能体能够自动识别“虚假签收”、“服务超时”等关键行为，并生成员工表现、客户画像及风险预警等多维度数据报告，助力企业管理层实现了从“事后复盘”到“事中干预”的精益化管理转变。在某物流的实践中，客户投诉处理时效由 48 小时压缩至 15 分钟，投诉率大幅下降。

5.2.3. 业务增长：释放数据要素商业价值

通过深度分析客户通话中透露的需求与偏好，智能体能够精准挖掘潜在商机。例如，在金融行业，系统可识别客户预算、理财偏好，辅助优化营销策略，推动服务部门从“成本中心”向“价值中心”转型。

6. 结论与展望

本研究通过严谨的技术评测(第 5.1 节)和广泛的行业实证(第 5.2 节)，证明了该方案在技术先进性与商业价值上的双重闭环。它以运营商真实数据为基石，通过全链路智能处理，将沉睡的语音数据转化为驱动企业高质量发展的核心资产。实证结果表明，本研究在提升运营效率、优化服务质量、强化风险控制等方面成效显著。

展望未来，AI 通话分析智能体具备广阔的推广前景。一方面，其标准化的技术架构与场景化的定制能力，使其具备跨行业、跨企业的强复制性；另一方面，其探索的数据资产化与流通机制，为数据要素市场的健康发展提供了有益借鉴。随着技术的持续迭代与应用场景的不断深化，AI 通话分析智能体有望成为推动千行百业数智化转型的重要基础设施。

基金项目

2024 年中国联通(广东)产业互联网有限公司云犀实时动态 AI 引擎研究项目(Y91R240EGH0003)。

参考文献

- [1] 中华人民共和国国家发展与改革委员会. 加快构建数据基础制度体系, 促进数据要素市场培育[EB/OL]. https://www.ndrc.gov.cn/xxgk/jd/jd/202212/t20221219_1343665.html, 2022-12-20.
- [2] 中华人民共和国中央人民政府. 中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见[EB/OL]. https://www.gov.cn/zhengce/2020-04/09/content_5500622.htm, 2020-04-09.
- [3] 中华人民共和国国家数据局. 十七部门关于印发《“数据要素×”三年行动计划(2024-2026年)》的通知[EB/OL]. https://www.cac.gov.cn/2024-01/05/c_1706119078060945.htm, 2024-01-05.
- [4] 芦宇. J 公司智能语音业务的竞争战略研究[D]: [硕士学位论文]. 北京: 对外经贸大学, 2024.
- [5] 中国联通政企在线. 联通云犀-COP 平台[EB/OL]. <https://gec.10010.com/sc/product/1003181>, 2025-11-17.
- [6] 中国联通元景大模型 MaaS 平台. 元景大模型 MaaS 平台介绍[EB/OL]. https://maas.ai-yuanjing.com/doc/pages/216543011/#_1-中国联通元景大模型_maaS_平台, 2025-11-17.
- [7] Yu, D. and Deng, L. (2015) Automatic Speech Recognition a Deep Learning Approach. Springer-Verlag.
- [8] Chavan, R.S. and Sable, G.S. (2013) An Overview of Speech Recognition Using HMM. *International Journal of Computer Science and Mobile Computing*, **2**, 233-238.
- [9] 郭晓哲, 彭敦陆, 张亚彤, 等. GRS: 一种面向电商领域智能客服的生成-检索式对话模型[J]. 华东师范大学学报(自然科学版), 2020(5): 156-166.
- [10] Mohamed, A., Dahl, G. and Hinton, G. (2010) Deep Belief Networks for Phone Recognition. *NIPS Workshop on Deep Learning for Speech Recognition and Related Applications* 2010, Whistler, 2010, 1-9.
- [11] Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., *et al.* (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, **29**, 82-97. <https://doi.org/10.1109/msp.2012.2205597>
- [12] Sun, Y., ten Bosch, L. and Boves, L. (2010) Hybrid HMM/BLSTM-RNN for Robust Speech Recognition. In: *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 400-407. https://doi.org/10.1007/978-3-642-15760-8_51
- [13] 王颖, 李承桓. 基于语音分析的智能质检系统设计[J]. 中国新通信, 2021, 23(12): 59-61.
- [14] 李如雄. 基于语音分析的智能质检系统设计[J]. 自动化与仪器仪表, 2017(6): 114-116.
- [15] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems* 2017, San Diego, 2017, 6000-6010.
- [16] Li, Y., Yu, J., Zhang, M., Ren, M., Zhao, Y., Zhao, X., *et al.* (2024) Using Large Language Model for End-To-End Chinese ASR and Ner. arXiv: 2401.11382.
- [17] Mohammadi, M., Li, Y., Lo, J. and Yip, W. (2025) Evaluation and Benchmarking of LLM Agents: A Survey. *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, Toronto, 3-7 August 2025, 6129-6139. <https://doi.org/10.1145/3711896.3736570>