

# 惟实励新 通信行业人工智能数据治理实践指南

## 2025



## 引言

人工智能与数据的演进轨迹始终同频共振。从 20 世纪中叶符号主义开启人工智能的启蒙时代，到联结主义推动数据成为智能核心燃料，再到大模型时代数据与智能的深度融合，近一个世纪的技术迭代中，两者从并行发展的“双轨线”，到相互赋能的“共生体”，最终形成“Data for AI”与“AI for Data”的螺旋式发展格局，深刻重塑了产业形态和发展主线。

数据智能浪潮正带来前所未有的挑战与机遇，通信行业作为信息社会的基石，数智化转型成为新阶段的核心命题。在此过程中，传统的数据治理工作模式和技术工具需要面向人工智能的需求进行一次全面的变革。要求通信运营商在高质量数据集建设和智能化数据治理方面的体系更为完善、技术更为适配、能力更为全面，同时可以相互赋能、深度融合。

本报告立足于国家《“数据要素×”三年行动计划（2024-2026）》与《关于深入实施“人工智能+”行动的意见》的战略指引，紧扣通信行业数据智能发展的核心诉求，深刻剖析了传统数据治理在面对大模型时代的数据需求时的能力短板，阐明了“Data for AI”与“AI for Data”双轮驱动的实践框架：一方面聚焦高质量数据集建设，从业务转型需求出发，明确数据集建设的管理规范、技术路径与评测标准，为模型训练提供坚实数据底座；另一方面深耕智能化数据治理，详解人工智能在数据标准、质量、安全、元数据、资产管理等领域的应用

实践，实现治理效率与数据质量的双重提升。同时，报告通过典型案例，具象化呈现数智融合的落地成效，并创新性提炼可复制、可推广的实践经验。最后，报告前瞻性指出“数智一体化”“人机协同”“内生安全”三大发展趋势，为行业未来布局提供方向指引。

本报告融合了历史演进、理论框架、实践路径与前沿趋势，既立足当下破解行业痛点，也着眼长远引领发展方向，旨在为通信运营商及相关行业的数智化转型提供全面的人工智能数据治理的理论支撑与实操指南，助力行业培育新质生产力，在数据智能浪潮中筑牢核心竞争力。

## 编制说明

本指南的撰写得到了人工智能与数据治理领域多家企业与众多专家的支持和帮助，主要参与单位与人员如下。

### 参编单位：

华为技术有限公司

中国信通院云计算与大数据研究所

中国移动通信集团河南公司

中国移动通信集团福建公司

### 指导组：

王福昌 李伟中 姜春宇 王妙琼

### 参编人员：

蔡火发 尹正 崔一妍 张鹏程 田菁菁 白琳 叶紫光 刘雪婷

曹锐 吴篁 陈哲愚 罗莹 贝高林 楼丕扬 王溪

# 目 录

|                                  |    |
|----------------------------------|----|
| 引言 .....                         | 1  |
| 编制说明 .....                       | 3  |
| 一、 人工智能时代的“智”与“治” .....          | 7  |
| (一) 分与合：人工智能与数据发展历程 .....        | 7  |
| 1. 人工智能发展历程——从“模拟”到“涌现” .....    | 7  |
| 2. 数据发展历程——从“好用”到“用好” .....      | 9  |
| 3. 数智融合浪潮——从“交汇”到“交织” .....      | 10 |
| (二) 智与数：人工智能需要什么样的数据 .....       | 11 |
| (三) 治与智：数据需要怎样的治理，智能如何改变治理 ..... | 13 |
| 二、 通信高质量数据集建设指引 .....            | 17 |
| (一) 人工智能驱动通信业务转型 .....           | 18 |
| 1. 人工智能典型业务场景 .....              | 18 |
| 2. 人工智能典型智能应用 .....              | 21 |
| 3. 人工智能驱动业务转型价值分析 .....          | 22 |
| (二) 通信高质量数据集新需求涌现 .....          | 24 |
| 1. 现有主要数据集基础 .....               | 25 |
| 2. 高质量数据集建设要求 .....              | 26 |
| 3. 高质量数据集建设难点 .....              | 27 |
| (三) 通信高质量数据集建设实践 .....           | 29 |
| 1. 高质量数据集管理 .....                | 29 |
| 2. 高质量数据集技术 .....                | 31 |

|                               |    |
|-------------------------------|----|
| 3. 高质量数据集评测 .....             | 33 |
| (四) 河南移动人工智能驱动智慧业务运营实践 .....  | 35 |
| 1. 持续开展数据治理，实现全过程强数 .....     | 35 |
| 2. 人工智能赋能智能治数，提升数据运转效率 .....  | 36 |
| 3. 数据治理驱动智慧业务运营全流程升级 .....    | 38 |
| 4. 业务全链条注“智”，赋能业务效率提升 .....   | 39 |
| 三、 通信智能化数据治理实践 .....          | 41 |
| (一) 人工智能驱动数据标准智能化管理 .....     | 41 |
| 1. 人工智能驱动的数据标准管理带来的变化 .....   | 41 |
| 2. 人工智能驱动的数据标准智能管理价值 .....    | 43 |
| (二) 人工智能协助提升数据质量管理效率 .....    | 44 |
| 1. 人工智能对数据质量管理带来的变化 .....     | 44 |
| 2. 人工智能数据质量智能化管理价值 .....      | 46 |
| (三) 人工智能赋能数据安全智慧化管理 .....     | 46 |
| 1. 人工智能赋能数据安全管理的变化 .....      | 47 |
| 2. 人工智能赋能智能化数据安全智慧化管理价值 ..... | 48 |
| (四) 人工智能使能智能元数据管理转型 .....     | 48 |
| 1. 人工智能使能智能元数据管理变化 .....      | 49 |
| 2. 人工智能使能智能元数据管理价值 .....      | 50 |
| (五) 人工智能帮助提升数据资产管理能力 .....    | 51 |
| 1. 人工智能给企业数据资产管理带来的变化 .....   | 52 |
| 2. 人工智能帮助提升数据资产管理价值 .....     | 52 |

|                             |    |
|-----------------------------|----|
| (六) 福建移动人工智能数据治理业务实践 .....  | 54 |
| 1. 深入开展数据&AI 现状问题分析 .....   | 54 |
| 2. 加强规划，建立多模态数据管理体系 .....   | 55 |
| 3. 数据治理驱动 AI+应用工作举措 .....   | 56 |
| 4. 人工智能原生数据治理使能数智运营实践 ..... | 58 |
| 四、 通信数据智能发展趋势 .....         | 62 |
| (一) “数智一体化” 能力体系 .....      | 62 |
| (二) “人机协同” 技术体系 .....       | 63 |
| (三) 内生安全 .....              | 63 |

## 一、 人工智能时代的“智”与“治”

### （一）分与合：人工智能与数据发展历程

人工智能与数据历史演进的轨迹横跨近一个世纪，从早期的两条平行线，分别受技术突破驱动，塑造各自领域的社会认知；到中期的逐渐交织，受应用和价值驱动，推动技术产业重塑。数据与智能最终走向融合。

#### 1. 人工智能发展历程——从“模拟”到“涌现”

纵观人工智能发展轨迹，其演进伴随技术突破与现实局限的交替，历经“三起两落”。崛起源于核心技术突破，沉寂归因应用落地瓶颈，最终在数据、算力、算法的协同演进中，实现了从“模拟人类逻辑”到“智能自主涌现”的跨越式发展。

#### “一起一落”：从符号主义崛起到第一次人工智能“寒冬”

符号主义核心是通过人工定义规则模拟人类逻辑推理，标志着人工智能从理论构想走向初步实践。作为人工智能的“启蒙时代”，其核心目标是构建能够复刻人类逻辑思维的“逻辑机器”。

但符号主义的崛起未能持续突破现实瓶颈，一系列固有缺陷导致人工智能进入第一次“寒冬”。核心问题集中在三个层面：一是规则局限性，面对复杂、不确定的现实场景，手动编码的规则库难以覆盖所有可能性；二是泛化局限性，领域专家知识的提取、编码成本极高，且难以规模化复制到不同行业；三是需求响应局限性，系统仅能处理结构化、符号化的明确问题，无法应对模糊性、非结构化的现实需求。

投资热潮的退去与应用落地的失败，使得人工智能领域陷入沉寂。这一阶段的困境也让业界意识到，单纯依赖人工规则模拟智能，难以突破人类认知与编码效率的双重限制，人工智能需要新的技术路径。

### **“二起二落”：从联结主义萌芽到第二次人工智能“寒冬”**

第一次寒冬后，人工智能迎来技术路径的分流与复苏。一方面，专家系统在特定领域持续迭代，验证了专用智能系统的商业价值。另一方面，联结主义开始萌芽，核心导向从“人工规则”转向“数据统计学习”，人工智能逐步从“逻辑机器”向“数据机器”转型。

“数据驱动”展现出潜力，但技术发展再次遭遇瓶颈。核心制约因素集中在两点：一是算力制约，网络传输和芯片工艺制约了神经网络的规模扩张；二是数据制约，因场景和技术局限，当时的数据量级与多样性不足，无法支撑复杂模型的训练。

由于模型泛化能力有限，业界对人工智能的过高期待与实际应用效果之间的差距，商业价值未达预期，导致投资的收缩。这一阶段的沉寂并非技术方向彻底失败，反而让研究者认识到，人工智能发展对“数据、算力、算法”协同的迫切需求，为后续深度学习的爆发埋下伏笔。

### **“第三次兴起”：从大模型爆发到跨越式发展“前夜”**

随着大数据技术成熟、算力指数级提升以及算法突破创新，人工智能迎来第三次崛起，核心导向从“专用智能”转向“通用智能”，实现了从“数据驱动”到“智能涌现”的质变，发展主线从“模型为中心”到“数

据为中心”。技术方面，参数规模化、多模态融合、提示工程等实现跨越式发展，模型不再局限于特定任务的优化，而是展现出上下文学习、思维链推理等“涌现”能力。

这一阶段，人工智能实现了规模化产业应用，成为“颠覆性生产力”，推动社会生产生活方式的深度变革。

## **2. 数据发展历程——从“好用”到“用好”**

而数据技术发展，同样伴随技术突破，经历了三代技术变革。

**第一代数据技术（DT1.0 时代）**以数据存储和查询为目标。在这个阶段，介质革命和存储范式革命，解决了海量数据的物理存储和高效管理之间的矛盾，数据成为独立处理对象，数据技术以事务处理为主，支撑业务贯通。数据进入“**存得住、查得到**”时代。

**第二代数据技术（DT2.0 时代）**以数据资源开发利用为目标，支撑数据分析、治理工作，实现数据驱动的业务决策。在这个阶段，互联网迎来爆炸式发展，人类社会产生的数据量级、生成速度呈现指数级增长，“大数据”概念产生，利用数据进行信息分析和知识挖掘需求增长，大规模/高弹性存算能力、半结构化/非结构化数据处理技术，以及数据分析相关技术都有长足进展。数据进入“**存得多、算得快**”时代。

**第三代数据技术（DT3.0 时代）**以数据要素价值释放为目标。在这个阶段，大数据的蓬勃发展，推动全球数据存储与处理能力形成规模化、平台化能力，如何进一步萃取数据价值并进行价值释放成为核

心任务。数据此时完成了从技术导向到价值导向的蜕变，数据资源化、数据资产化、数据要素市场化成为新发展路径。数据进入“供得出、流得动、用得好”时代。

我国数据产业在此时期发展迅猛，党的十九届四中全会（2017）将数据增列为一种生产要素，以“数据二十条”（2020）、《“数据要素×”三年行动计划（2024-2026）》（2023）为代表的制度出台，明确了数据要素市场化配置的正式顶层设计，奠定了我国数据产业发展的基础制度方向，推动数据要素价值释放。

技术方面，围绕数据资产管理领域的方法论包括数据管理能力成熟度评估模型（Data Management Capability Maturity Assessment Model, DCMM）、数据研发运营一体化（Data Development and Operations, DataOps）、数据编织、数据工程、数据运营、数据估值、知识工程等领域研究，以及围绕数据流通的包括隐私计算、区块链、可信数据空间等领域研究，正在书写数据发展史的崭新篇章。

### 3. 数智融合浪潮——从“交汇”到“交织”

符号主义崛起后，数据通过人工处理，在系统架构中转化为规则库、知识库，成为逻辑推理的依据，也是算法的验证基础。明显交汇的时间点在1970年和1980年，前者关系型数据库的出现推动了专家系统向不同行业的横向发展，后者是人工智能和数据同时进入商业化和规模化发展的重要节点。

联结主义萌芽后，数据从幕后进入舞台中心，统计方法、机器学习

习、深度学习、大模型的技术跃迁，依赖于大数据技术的突破，也反哺了数据技术的发展热情。明显交汇的时间点在 2000 年，机器学习算法的创新，带动人工智能技术走出第二次“寒冬”。同时，大数据浪潮推动数据存储、处理、应用等方面的技术跃迁，提供更专业的人才保障，为强化学习、深度学习提供了充足、好用的“燃料”基础。

当前，数据与人工智能两条交织的历史发展曲线已趋近重合，“AI for Data”“Data for AI”两条路线的螺旋式发展成为主线，人工智能的数据集建设进入黄金发展窗口期，智能化的治理和分析决策已经普及。在可以预见的数据智能浪潮中，我国加速产业和政策布局，出台《高质量数据集建设指引》《关于深入实施“人工智能+”行动的意见》等制度文件，为数据与人工智能深度融合指明方向。未来，多种数据智能的融合应用需求将重塑技术发展路线和社会价值框架。

## **（二）智与数：人工智能需要什么样的数据**

总体看，人工智能“三起两落”的发展脉络中，符号主义崛起后和联结主义萌芽后两个时期所需的数据类型有着根本性的不同，这塑造了数据处理、数据存储、数据应用等方面截然不同的两个时期。

符号主义崛起后，用户提出的具体问题，需要系统基于知识库进行推理来回答。这一时期，以处理结构化、符号化的规则集和知识为主，主要包括事实数据（实例、属性）、规则数据（逻辑推理规则）、概念数据（知识表示）。在数据存储方面，主要依托内存和磁盘为介质，以知识库为结构存储，包括规则库、事实库、本体库、语义网络

等形式。在数据使用方面，数据通过推理引擎进行使用，有“规则驱动推理”和“目标驱动推理”两种典型方式，部分系统已经具备冲突解决、解释、不确定性处理等功能。

联结主义萌芽后，正式进入“数据驱动”时代，人工智能转向神经网络结构和学习机制的研究，以及智能体与环境的交互学习。数据需求和处理方式发生根本性变革，可以分为机器学习、强化学习-深度学习、大模型三个时期。

机器学习期，需求聚焦从历史数据中学习规律，解决有明确输入输出的监督预测任务、无监督结构发现任务、半监督/弱监督任务，追求预测精度和泛化能力。此阶段所需的数据类型包括，**大规模数据集（结构化为主）、特征数据、标签数据、测试和验证数据等**。数据存储方面，主要依赖关系型数据库以及文件系统。数据处理方面，以统计分析和特征工程为核心，主要工作流程包括数据清洗与预处理、特征工程、模型训练与调参。数据使用方面，训练好的模型本质上是一个 预测/决策函数，使用方式通常包括接口调用、离线预测/批量预测等。

强化学习 - 深度学习期，开始出现处理复杂的视觉、自然语言处理等需求，需要理解高维、非结构化数据，同时关注智能体在与动态环境的持续交互中学习最优策略等需求。此阶段所需的数据类型包括，**大规模标注数据集、未标注/弱标注数据、经验数据/轨迹数据/交互数据、多元特征等**。数据存储方面，主要依托分布式文件系统、非关系

型数据库等。数据处理方面，和机器学习期相比，需要数据增强、数据标注等新的处理流程，主要面向端到端学习、深度神经网络训练、强化学习训练等相关流程。数据使用方面，通常用于支持包括复杂感知推断、端到端应用、在线决策等应用。

大模型期，开始面向构建具有极强语言理解、生成、多任务泛化能力的系统，需要模型能够胜任开放的复杂对话、创作、知识问答、代码生成等高难度任务，并表现出一定程度的上下文学习、思维链、“涌现”能力，应用目标更加开放、通用。此阶段所需的数据类型包括，超大规模语料库、海量多模态数据、指令微调数据、偏好排序/对齐数据、高质量清洗后的数据等。数据存储方面，以超大规模分布式/云对象存储、数据湖、向量数据库等为主。数据处理方面，与强化学习-深度学习期相似，部分需要数据合成，以嵌入无监督预训练、微调、提示工程、分布式训练等相关流程。数据使用方面，需要支持包括检索增强生成（RAG）、智能体（Agent）、“人机交互”等应用需求。

### **（三）治与智：数据需要怎样的治理，智能如何改变治理**

人工智能发展的“三起两落”，带来数据规模、数据类型的变化，同样也对数据治理的工作模式和技术发展产生深刻影响，呈现出阶段性特点：一是从治理对象看，研究重点从“结构化数据”到“非结构化数据”；二是从治理目的看，治理重点从“质量”到“场景”；三是从治理技术看，投入方向从“劳动密集”到“智能密集”。

**符号主义崛起后**，数据类型集中于结构化、符号化的规则集、事实库与概念数据，数据规模小、格式规整，决定了数据治理呈现“人工主导、规则驱动、场景专用”的特点。

**联结主义萌芽后**，数据类型从结构化扩展到非结构化、多模态，数据规模呈指数级增长，数据治理核心转向“规模化管控、全流程质量优化、动态安全防护”，形成了与机器学习、强化学习-深度学习、大模型发展适配的治理体系。

**机器学习期**，以统计学习为核心，数据类型以大规模结构化数据为主，含特征数据、标签数据、测试 / 验证数据，核心目标是提升数据与模型的适配性，治理特点呈现“半自动化、统计驱动、聚焦特征优化”。该阶段的关键治理流程围绕数据的预处理与特征优化展开，核心流程包括数据清洗、特征工程、标签规整，其核心逻辑是通过半自动化手段提升数据的统计有效性，为模型训练提供高质量输入。

**强化学习—深度学习期**，需处理高维非结构化数据（图像、语音、文本等），数据需求聚焦高质量、高多样性，治理特点呈现“自动化工具辅助、全流程管控、质量与安全并重”。该阶段的关键治理流程包括数据标注、数据增强、数据合成，核心是通过自动化工具提升数据规模与质量，填补稀有场景数据缺口。

**大模型期**，数据类型涵盖超大规模语料库、多模态数据、指令微调数据，核心目标是实现数据的“高纯净度、高多样性、高安全性、高适配性”，治理特点呈现“全流程智能化、跨域协同、价值驱动”。

该阶段的关键治理流程围绕多模态数据、数据清洗、指令微调数据治理、知识管理展开。当前，多数企业已经能够将半自动化或自动化数据治理工具嵌入数据全生命周期的管理流程中，正式进入智能化数据治理的全链路升级阶段。

智能化数据治理是借助人工智能等自动化、智能化技术和手段开展数据工作，提高数据管理与应用的效能。充分发挥人工智能在语义理解、逻辑推理、智能生成等方面的优势，通过深度学习、大模型、回归模型、知识图谱等大小模型的技术搭配，提升对多模态数据的处理、理解及应用，降低运营成本、提升工作效率，并通过人工交互等技术降低数据使用门槛，提升数据资源利用率。

智能化数据治理的探索工作正在深入快速展开，未来将从全链条“AI+数据治理”“人机协同”两个方面重点发力。在全链条“AI+数据治理”方面，已经可以看到智能化数据治理在数据质量、数据标准、元数据、数据安全、数据模型、数据资源等方面的落地实践，应用场景极为丰富，可以提高原有数据管理工作的效率，也可以构建起新的管理能力和管理模式。在“人机协同”方面，通过特定场景的“数字员工”赋能千行百业，充分发挥智能体的思维、逻辑、知识以及场景任务执行力，构建“人机协同”的工作新范式。

人工智能与数据治理的交织发展也为通信运营商数智化转型提供了双轮驱动力。数据治理通过构建高质量的数据基础，为人工智能算法训练与应用提供了重要支撑，而人工智能则通过自动化、智能化

技术优化数据治理流程，提升数据处理效率与价值挖掘能力。此外，通信运营商通过数据治理构建的统一数据平台，为 AI 应用提供了数据底座，而 AI 的嵌入则加速了数据要素价值转化，将助力运营商从传统电信服务向智能信息科技服务转型。

数据治理与人工智能这种双向赋能交织发展关系不仅提升了通信运营商的运营效能，还为其开拓新的智能服务、智能云化网络、行业数智化解决方案等新增长领域提供了技术保障，成为其数智化转型的关键加速器。

## 二、 通信高质量数据集建设指引

在人工智能时代，数据与人工智能关系日益紧密，人工智能技术的深入发展，使得人工智能将深入通信运营商各类业务领域生产流程，由此带来数据管理方式也将更加复杂多变，非结构化数据管理提上日程，高质量数据集建设成为必选。

大模型技术发展大幅降低了算力需求和算法瓶颈，使得通信运营商企业可以较低成本开展人工智能实践及应用。在企业人工智能应用过程中，高质量的数据成为智能化表现的核心差异化因素，通过深入开展企业数据治理形成高质量数据集和良好的数据伦理，积极基于通用大模型技术开展 L1/L2/L3 级别运营商领域大模型的推理、增训工作，将推动运营商在企业办公、市场营销、客户服务、网络运维、资源管理、安全合规等领域快速应用人工智能，助力运营商企业更好地利用数据智能驱动业务创新，实现更加可持续的高质量发展。

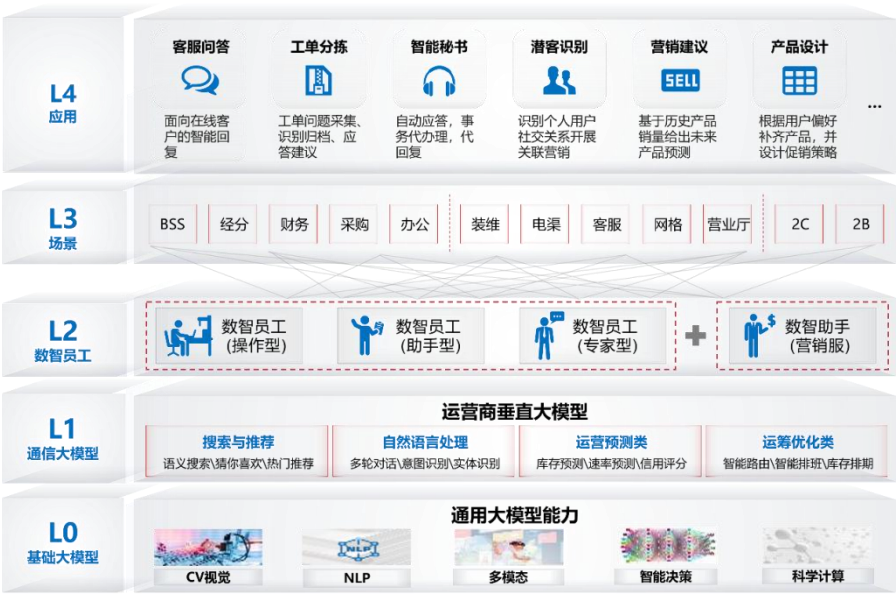


图 1：运营商大模型典型应用场景

## （一）人工智能驱动通信业务转型

目前通信运营商在企业内部办公协同、市场营销、客户服务、合同管理、供应链管理、网络运维等重点场景，均存在较强烈的大模型应用需求。部分领先运营商优先聚焦办公、营销、客服类等高频、海量、刚需场景，再逐步深入网络运营运维、资源管理等生产场景，进行了较多探索实践，有力地促进了业务转型。

通信运营商实现“用人工智能”走向“用好人工智能”转变，其主体应用原则一方面是场景选择先易后难，办公、营销、客服场景容易，网络运维生产场景复杂实现较难；另一方面是技术选择从简单到复杂，先推理后微调/增训/强化学习，运营商各场景主要使用大模型技术及工具诉求如下：



图 2：运营商 AI+应用场景主要使用的大模型技术

### 1. 人工智能典型业务场景

通信运营商可以在人工智能驱动下，结合企业业务发展情况，优先针对企业办公、市场营销、客户服务、网络运维、资源管理及安全合规等业务场景，开展业务转型探索实践。

- **企业办公场景：**办公类的场景的智能化转型相对比较容易实现，其特点是单场景的智算算力消耗比较小，但端侧客户使用量比较大，通过大模型+RAG 方式，通信运营商可以快速部署办公助手，

在企业办公场景智慧问答、公文撰写，财经助手，审计助手开展智能化应用实践。

➤ **市场营销场景：**市场营销类场景包括运营商市场、政企、商客类的业务营销，其业务智能化转型细分场景主要包括：一、借助智能应用，市场营销业务人员能够快速地去查阅知识库，了解客户的套餐产品，以及相关的技术，业务类的问答；比如通过构建智慧营销助手，通过 RAG 构建私域业务知识向量库，以场景为维度聚合展示业务相关知识，包括业务说明、资料等，能做到业务知识一问即答，信息搜索一查即达；二、通过业务智慧化改造，业务人员在客户业务办理上提供便捷的路径，直接能跳转到办理的业务流程中，通过 RPA 的方式通过接口登陆到业务系统中，比如描述业务需求，即可帮助销售人员办理云、政企专线等主要业务；三、通过智能化应用，如面向政企领域，客户经理面向客户界面，比如构建客户经理分身，运营商客户有问题优先找客户经理分身，客户经理分身无法解决的，再由客户经理进行解决。客户经理通过客户经理助手来搜索业务问答，尽量满足客户的需求，提升客户满意及营销效益；四、在市场营销客户洞察、产品匹配、策略配置、渠道匹配及营销执行等方面也可以使用大模型技术深度介入市场营销流程，构建智能化应用，实现营销效率、效益及质量的提升。

➤ **客户服务场景：**利用大模型优化智能客服，实现客户体验提升

及个性化推荐、客户服务效能提升及客服工单质检效率提升，比如：一是面向终端用户提供更优体验，通过人工智能让客户能体验更好，通过识别用户意图，自动登单帮客户办理或退订相关业务，减少等待及操作；二是面向坐席提升效能，当呼叫呼入后，用户咨询坐席人员，可以快速的将用户的意图识别，同时调用知识库给用户回答问题，解决用户的问题后实现自动录单提升效率；三是面向质检提供准确率，通过ASR+大模型+人工判断来实现“人工智能质检”，能做到全量质检，覆盖质检量，提升效率，同时降低人工成本。

➤ **网络运维场景：**通过人工智能分析网络流量、信令数据、用户使用及网络设备状态，实现网络资源动态调度、故障预测与自动修复。同时利用大模型技术进行网络性能监控、故障预测与自愈、容量规划。例如：基于深度学习的异常检测模型可提前识别基站故障风险，实现预测性维护。

➤ **资源管理场景：**结合运营商网络资源、成本资源及业务运营数据，通过构建资源管理领域大模型，开展运营商网络资源深化管理，优化通信基站能耗、数据中心资源分配及基础设施维护计划；在通信运营商数据中心、边缘计算节点等基础设施管理中，借助人工智能辅助能耗优化、负载均衡与资源分配，降低网络运营运维成本，提升智能化水平。

➤ **安全合规场景：**通过人工智能检测网络攻击、欺诈行为及数据

泄露风险，保障通信运营商用户隐私与网络安全，应用大模型技术识别异常通信行为，防范垃圾短信、骚扰电话、伪基站及 SIM 卡盗用等安全威胁，保障用户通信安全与网络可信环境。

2. 人工智能典型智能应用

基于上述业务场景，通信运营结合大模型基础模型、自身数据及大模型数据工程、知识工程有能力，可以开展诸如典型智能应用包括：

- 办公助手应用：开发及部署办公助手，在企业办公场景智慧问答、公文协助，财经助手，审计助手开展智能化应用。
- 营销智能体应用：利用 NLP、数据工程、RAG 知识工程、提示词工程等大模型技术，构建市场智慧营销智能洞察、智慧分析、智能问答及应用编排等智能应用能力，实现市场营销相关客群人工智能分析、商机人工智能挖掘、策划自动生成、产品智能推荐、复盘报告自动生成等智能应用。

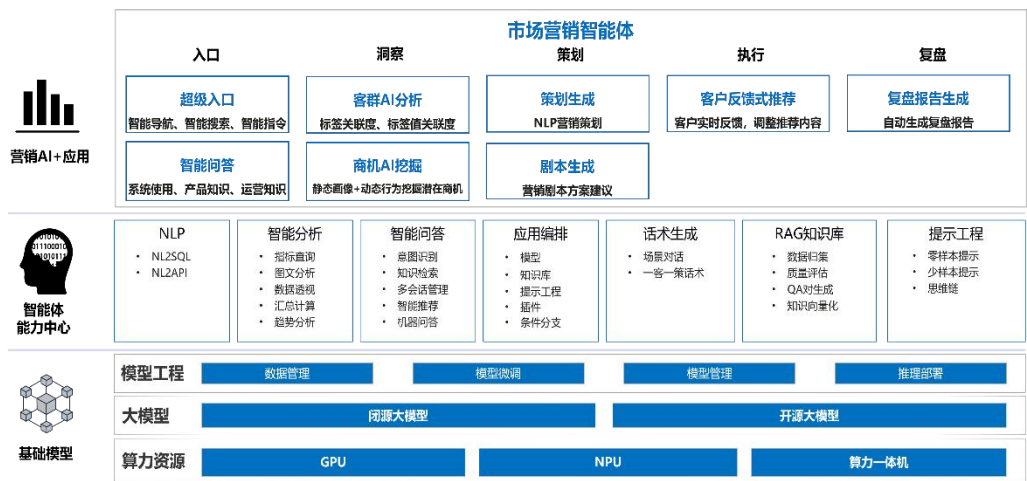


图 3：运营商市场营销智能体

- 智能客服应用：构建智能客服应用，实现 7×24 小时语音/文本交

互，解决用户咨询与投诉，降低人工成本。基于自然语言处理（NLP）技术，支持多轮对话理解与意图识别，实现首呼解决率（FCR）超过 70%，提升客户服务效率，降低客户服务成本。

➤ **故障预测应用：**基于通信运营商网络运维故障、工单、资源等历史数据与实时监测，提前识别网络瓶颈或故障节点，通过集成大模型，实现网络告警压缩、根因分析与自动化处置，通过人工智能模型将告警数量减少 80%，故障定位时间缩短至分钟级。

➤ **财经助手应用：**针对传统财经知识问答依赖人工维护问答库、仅能回答预设范围内的问题、仅能回答预设范围内的问题等短板及问题，通过构建财经领域 RAG 知识库，在公司差旅政策、报销标准，税务政策查询等细分场景实现多路径知识自动召回及智能问答，问题回复准确率 90%+。

### 3. 人工智能驱动业务转型价值分析

人工智能不仅是技术工具，也是运营商战略转型的核心驱动力。通过提升效率、优化体验、数据赋能、业务创新、增强韧性和履行社会责任，运营商能够实现从传统通信服务商向信息服务科技公司的转身，抓住数字化时代的增长机遇。实践证明，人工智能转型结合平台工具、数据治理、组织变革和生态协作，可以最大释放其全部潜力，为运营商各类业务转型带来了显著的业务与战略价值。

➤ **提升运营效率与降低成本**

人工智能通过自动化和智能化流程显著优化运营商的运维效率，减少人工干预。例如，在某运营商的数字智能运维项目中，人工智能技术将问题分析时间从 60 分钟缩短至 15 分钟，处理效率从 85% 提升至 95%。此外，人工智能驱动异常检测和根因分析（如数字孪生技术）能主动识别网络问题，避免被动响应，从而降低运营成本，人工智能转型可使企业从繁琐任务中解放，提升员工创造力。

### ➤ 优化客户体验与满意度

利用人工智能，通过实时数据分析和智能推荐，显著提升客户体验。例如，在运营商自智网络实践实现了“零业务影响感知”和“零人为干预”，确保用户服务连续性。人工智能还能精准预测用户需求，例如通过智能客服提供个性化服务，减少客户等待时间。此外，人工智能在新通话场景中的应用（如人工智能代聊、摘要生成）进一步增强了用户体验。

### ➤ 数据驱动决策与业务创新

运营商通过整合人工智能平台与大数据平台能力，开展数据治理，挖掘数据价值，支持精准营销和业务创新。例如，中国移动通过人工智能协同大数据，利用“场景、数据、流程、平台、工具”五要素开展企业数智化转型，提升“大价值”，而中国联通则构建智能数据中台，实现“数据驱动”业务增长，此外人工智能还能协助商业智能（BI），提升数据分析效率，增加业务管理决策能力。

### ➤ 推动业务模式转型与增长

人工智能赋能运营商从“连接提供商”向“技术解决方案”转型。例如，某 T 运营商将人工智能、云与网络连接融合，拓展企业数字化转型市场，人工智能驱动的转型路径可支撑运营商政企业务增长，如智能云网、5G 专网等。此外，人工智能在视觉智能领域、智慧家庭领域的应用为运营商开拓新市场空间。

### ➤ 增强网络韧性和安全性

人工智能技术提升网络的自适应和自修复能力，保障业务连续性。例如，某 S 运营商的智能网络安全管理系统通过人工智能排除网络故障，人工智能 WAN 解决方案则通过智能感知和运维升级，增强网络安全性，运营商需通过人工智能和数据保护构建“业务韧性”，如防勒索攻击的隔离区保护，确保核心业务零中断。

### ➤ 实现可持续发展与社会责任

人工智能助力运营商优化能源消耗，推动绿色网络建设。如某运营商基于人工智能驱动的 OXC 和 OTN 光电融合方案打造低能耗全光城市；此外，人工智能在社会服务中的应用（如医疗影像分析、智能驾驶辅助、智慧家庭中心）使运营商成为社会数字化转型的赋能者，履行社会责任的同时开拓商业价值。

## （二）通信高质量数据集新需求涌现

高质量数据集是人工智能发展的基础，高质量数据集建设已被提升至战略高度，《国家数据基础设施建设指引》《高质量数据集建设

指引》等政策明确要求建设行业高质量数据集，推动“人工智能+”行动落地。2025 年国家数据局召开的高质量数据集建设工作启动会进一步强调，数据与人工智能的深度协同将激发产业创新的倍增效应。

运营商作为国家数字经济的核心骨干企业，需在政策指导下加速人工智能布局。同时，运营商自身面临数字化转型的现实需求，传统通信业务增长放缓，5G、物联网等新兴领域对智能化运营提出更高要求。例如，中国移动提出从“+人工智能”转向“人工智能+”，构建人工智能全要素数字基础设施，依托网络、算力优势赋能业务变革，高质量数据集成为运营商实现网络智能化、业务创新的关键支撑。

### 1. 现有主要数据集基础

随着通信运营商业务发展及 IT 系统演进，运营商沉淀了较多结构化与非结构化数据，并开展了基本的数据管理，其中包括非结构化数据如文本、图片、音频、视频、深度图像等大模型密切相关的多种类型的数据管理。同时，部分运营商聚焦人工智能赋能的业务运营、网络运维、管理决策、客户服务及行业应用等价值领域，构建了部分基础数据集，如：

- **业务运营数据集：**主要由经营分析系统、客户关系系统、客户服务系统、渠道运营系统等系统构成，应用以报表输出、业务办理为主，主要包括报表、人脸认证、入网视频、客服录音、语音文本等数据集。
- **网络运维数据集：**由网络资源管理、网络工单、网络信令等

系统构成，如用户位置、通话记录、流量使用模式，用于网络规划与故障诊断；设备日志如基站、核心网设备的运行状态与告警日志，支持运维分析。

- **管理决策数据集：**由计划建设、采购系统、报账系统、工单系统等系统构成，应用以业务流程、电子工单、合同管理为主，包括电子合同、流程附件、操作日志等数据集。
- **客户服务数据集：**包括客户消费行为如套餐选择、流量使用、增值业务订购记录，用于客户精准营销；交互行为如客服对话记录、APP使用习惯优化服务体验，主要包括客户时空轨迹、音频、视频、图片、电子邮件等数据集。
- **外部行业数据集：**5G行业应用数据如工业物联网设备数据、车联网通信记录，各类行业信息及数据，支持垂直行业领域解决方案建设。

上述数据集是通信运营商各领域价值数据汇聚、治理、使用及应用的重要战略基础，也是推进运营商数据能力从分析支撑向主动注智转变的核心内容，但面临行业竞争及企业转型要求，通信运营商当前仍面临数据孤岛、数据标注成本高、数据标准缺乏、数据内容不全面等挑战，需进一步丰富数据管理体系、加强高质量数据集建设，将企业业务、数据与专业知识整合，形成差异化竞争优势。

## 2. 高质量数据集建设要求

高质量数据集是指经过系统化设计与处理的、满足人工智能模型

训练、验证及测试需求、质量有保障测试、可度量的数据集合。高质量数据集具有主题明确性，需要面向明确的应用场景和目标，覆盖特定领域知识；同时高质量数据集成要有多维度质量属性，在完整性、规范性、准确性、均衡性、及时性、一致性、相关性等方面达到高标准，并且遵从“伸缩法则”，具有足够数据量，覆盖多样化的场景与边缘案例，避免模型欠拟合、过拟合。

- 在技术要求方面：高质量数据集成要求通过去噪、格式规范化、人工/自动化标注等手段提升数据纯净度与标注精度；同时高质量数据可能需要利用数据合成及增强技术扩展数据多样性，尤其在行业垂直领域填补数据空白，高质量数据集应客采用自动化检测与模型验证确保数据对模型效果的正向作用，可以对数据质量进行评测评估。

- 在管理要求方面：高质量数据集需有明确的管理主体，如需设置原始数据 Owner 或人工智能数据集责任人，确保全生命周期可追溯，在数据安全及隐私方面，遵循数据边界保护原则，禁止泄露公司内部信息或违反隐私法规；在标准化管理流程方面，高质量数据集的管理需涵盖数据规划、采集、解析、清洗、标注、合成、质检及发布等环节的标准化流程。

### 3. 高质量数据集建设难点

通信运营商在数据集建设方面仍然面临多重挑战，主要体现在数据来源多样性不足、标注能力薄弱、数据孤岛问题及治理机制不完善

等方面。首先，运营商数据来源以基础数据为主，依赖大模型回收数据和合成数据，但普遍缺乏高效的数据治理能力，导致可利用的高质量数据较少。其次，数据标注能力不足，尤其是人工智能辅助标注和场景化标注能力尚未完全建立，而运营商参与的数据标注基地建设仍需进一步扩展。

数据质量控制也是关键短板。大规模数据集建设中存在重复和冗余问题，去重技术不足影响模型训练效果，且缺乏统一的数据清洗、评估标准，运营商内部及跨部门数据分散，难以高效整合与传输，限制了数据的人工智能利用效率；同时，数据共享机制不畅，内部部门间数据隔离严重，外部流通则面临可信流通方案缺失，影响数据要素价值释放。具体体现如下：

- 在连接方面：部分业务领域数据是离散、孤立的，游离于各个角落中，被烟囱林立的系统割裂。
- 在服务方面：数据仍然存在找不到、不流转，用户找不到文件，不能获取完整的文件，业务化后的数据没有回到业务中，不能高效服务于业务。
- 在质量方面：数据质量可能是混乱的、困惑的，非结构化内容标准不一样，同类内容存于不同路径；没有统一元数据模型，没有元数据，数据质量较低。
- 在能力方面：高质量数据集服务是被普遍渴望的但不能有效满足，每个岗位都希望得到数据支撑，但很多应用场景得不

到智能赋能；

- 在体系方面：高质量数据集管理体系尚未建立，尤其是对非结构化数据意识薄弱，无战略规划，数据资产看不到，未实施有效的资产管理。
- 在安全方面：尚未体系性建立数据集安全管理机制，敏感数据文件可能容易被泄漏，且难以追溯；难以应对各种安全、合规审查。

### **（三）通信高质量数据集建设实践**

当前，部分领先运营商正通过管理与技术的协同创新，积极开展高质量数据集建设，推动数据要素价值释放，助力人工智能时代的数据资产化。

在管理方面，运营商正在构建企业数据治理体系，并把非结构化数据管理作为其中重要内容，通过建立专门的数据管理组织、制度，强化数据治理流程与标准化建设，优化数据采集、传输、存储、计算和应用环节。在技术方面，运营商积极构建高质量数据管理及供给平台，拓展数据来源并提升数据标注能力，数据来源涵盖运营商基础数据集、行业外购数据及合成数据，并通过人工智能辅助标注和场景化标注技术提升质量。同时，借鉴石油炼化流程，实施数据设计、采集、治理、标注、质检和运营的全流程管控，实现高质量数据集的“源、采、存、管、算、供、用”端到端的闭环管理。

#### **1. 高质量数据集管理**

高质量数据集的建设离不开科学的数据管理体系支撑，运营商在实践中逐步建立涵盖高质量数据集数据管理框架、组织、制度、流程、合规、共享、六位一体的管理机制：

- **管理体系：**以发挥多模态数据资产化价值为核心，以赋能人工智能+为目标，通过构建高质量数据集管理体系，明确多模态数据管理范围，制定多模态数据管理流程，建立统一数据标准与分类体系，形成数据生命周期管理策略，使得企业高质量数据集可采、可管、可视、可供、可用、可演进。
- **组织保障：**设立专门的“数据治理办公室”或“人工智能数据管理中心”，统筹非结构数据管理及高质量数据集数据资产规划，实现数据标准制定与跨部门协同，明确数据责任人制度，落实数据质量归口。
- **制度规范：**参考国家、行业标准及集团要求，制定高质量数据集管理办法，建设数据标注操作方法，统一标签体系、标注工具与审核流程；同时在原企业数据管理制度体系基础上，完善非结构数据元数据管理、数据采集管理、数据标准管理、数据质量管理、数据生命周期管理、数据安全等方面的多模态数据管理标准、规范建设，提升领先性。
- **安全与合规：**遵循个人信息保护法及相关法规，实施数据脱敏、匿名化处理，采用区块链技术记录数据访问日志，确保审计可追溯。

- **共享生态：**通过数据集管理与采集供给平台，实现行业外部集成及内部数据采集汇聚生态，实现跨企业、跨领域数据协作，为企业各智能体供数。

## 2. 高质量数据集技术

基于数据采集、数据处理、数据标注、数据评测、数据供给等环节，通过数据工程工具链端到端完成数据的加工和处理，为智能应用准备高质量数据。



图 4：运营商数据知识化基本过程

### ➤ 数据采集

部署数据集边缘计算节点实时采集网络与用户数据，利用 Flume、Kafka、Flink 等实现多源异构数据实时采集与流式接入；基于企业统一数据湖，整合通信运营业务领域、网络领域、管理领域等多域数据。

### ➤ 数据处理

结合人工智能算法自动清洗噪声数据。应用规则引擎与机器学习

结合的方式识别并修复异常值、重复记录、格式错误；使用 Spark 进行大规模数据去重、归一化、空值填充等操作。

➤ 数据标注

采用半自动化标注工具（如 NLP 文本分类）提升效率，结合数据增强技术（如合成攻击样本）扩充数据量。对于文本类数据，采用预训练模型（如 BERT）进行弱监督标注，再辅以人工校验；对于语音数据，结合 ASR 输出与专家规则生成初步标签，提升标注效率；探索主动学习（ActiveLearning）策略，优先标注对模型提升最有价值的样本。

➤ 数据评测

根据高质量数据集质量评测规范，结合高质量数据集业务实践情况，发布数据集质量标准，针对数据集相关内容有效性、文本干净度、格式规范性、语料毒性等维度，开展数据评测，快速发现质量问题，持续提升数据质量。

| 分类    | 评估                  |
|-------|---------------------|
| 内容有效性 | 语句不通顺               |
|       | 语义不完整，缺乏必要的上下文，背景信息 |
|       | 内容不完整               |
|       | 无效内容                |
|       | 重复内容                |
|       | 语义冲突                |
|       | 事实错误                |
|       | 错别字                 |
|       | 公式错误                |
|       |                     |
| 文本干净度 | 冗余引用：如图，如表，章节交叉引用等  |
|       | html 冗余头部、尾部        |
|       | 页眉页脚信息              |
|       | 目录信息                |

|       |                           |
|-------|---------------------------|
|       | Copyright 说明              |
|       | 作者简介                      |
|       | 参考文献                      |
|       | url                       |
|       | 特殊文本内容                    |
|       | 图片名称                      |
|       | 包含乱码，如表情符号，base64 编码等     |
|       | 广告                        |
|       | 水印                        |
| 格式规范性 | 多余标点符号，如多余空格，多余的换行等       |
|       | 缺少标点符号                    |
|       | 序号格式不正确                   |
|       | 局部格式杂乱（段落顺序，无关信息插入等）      |
|       | 错误表格格式                    |
|       | 全文排版杂乱（如 PDF 单双栏拼接错误）     |
|       | 表格缺表头                     |
| 数据毒性  | 政治相关                      |
|       | 色情、暴力、恐怖等内容               |
|       | 个人信息：电话，邮箱，QQ，IP，个人账号，密码等 |
|       | 负面价值观                     |

## ➤ 数据集管理

对各类未标注和已标注的数据进行管理，通过提供数据上传、版本控制和权限设置等手段，确保数据的有序存储和安全。涵盖大模型数据管理、数据版本管理、数据上传、数据存储、数据订阅、数据下载、数据共享及对应的审核服务。

## 3. 高质量数据集评测

高质量数据集评测是对已使用的数据集相关数据质量、应用效果、安全合规、成本效益、数据可用性及人工智能应用导向等方向开展数据评估评测，使数据集可迭代演进发展。

## ➤ 数据质量评测

通过完整性（覆盖率，如字段非空比例）、准确性（标注误差率，如经抽样核验正确比例）、时效性（更新频率）、一致性（跨系统数据对齐，如跨系统匹配度）等指标对数据集质量进行评估评测。

#### ➤ 数据可用性评估

针对数据集文档完整性（含元数据说明、字段定义、使用示例）；数据访问响应时间与 API 调用成功率；用户满意度调查（来自算法工程师、数据科学家的反馈）；通过定期开展数据集“健康度评估”，形成“建设—使用—反馈—优化”的闭环机制，持续提升数据资产价值。

#### ➤ 人工智能应用导向评价

模型性能提升度：使用新数据集训练的模型在测试集上的准确率、F1-score 等指标相比旧数据集提升幅度；

训练收敛速度：高质量数据是否显著缩短模型收敛所需迭代次数；

模型泛化能力：模型在跨区域、跨时段数据上的表现稳定性。

#### ➤ 应用效果评估

通过对数据集支撑大模型、智能体相关业务模型性能（如准确率、召回率）、业务指标（如故障响应时间、用户满意度提升）进行评测。

#### ➤ 成本效益分析

数据集建设成本与人工智能应用带来的收益（如成本节约、效率提升、收入增长）对比；为持续优化数据集质量，确保数据集真正服务于人工智能应用目标。

➤ 合规性与安全性

针对高质量数据集相关数据脱敏合规性检查、访问权限审计、攻击检测模型的误报率与漏报率进行评估评测。

（四）河南移动人工智能驱动智慧业务运营实践

河南移动通过多年数据治理、数智化转型工作实践，积极落实中国移动通信集团公司数智化转型“五定”“五一”数据治理专项部署。通过深刻把握数据作为新型生产要素的战略定位，从五大数据问题出发，以“五要素”转型方法论为指引，以“高价值业务场景”为牵引，建设1+4+1+X 河南特色数据治理体系，治用结合，全过程强数赋能，锻造数智新质生产力。



图 5：河南移动企业数据治理体系

1. 持续开展数据治理，实现全过程强数

河南移动从 2023 年开始锻造企业数据治理中枢，以“推进数智化转型，实现高质量发展”为发展主线，以“数据治理委员会”为引领开启企业级数据治理新征程，发布公司数据管理“基本法”，完成 150+ 公司系统数据纳管及汇聚，建设了数据资源、数据资产、数据开放三层目录，并通过数据资源管理与开放能力面向全公司提供数据订阅能力，满足业务侧智慧用数需求。

河南移动围绕数据产生、汇聚、开发、使用、退役等全生命周期进行强化管理，提升数据质量以及端到端流转效率。识别影响数据发挥价值的关键节点，针对数据产生、数据汇聚、数据治理、数据使用四大过程开展流程重塑，建设一套智能化数据治理能力，打通数据流通断点，实现全过程强数。

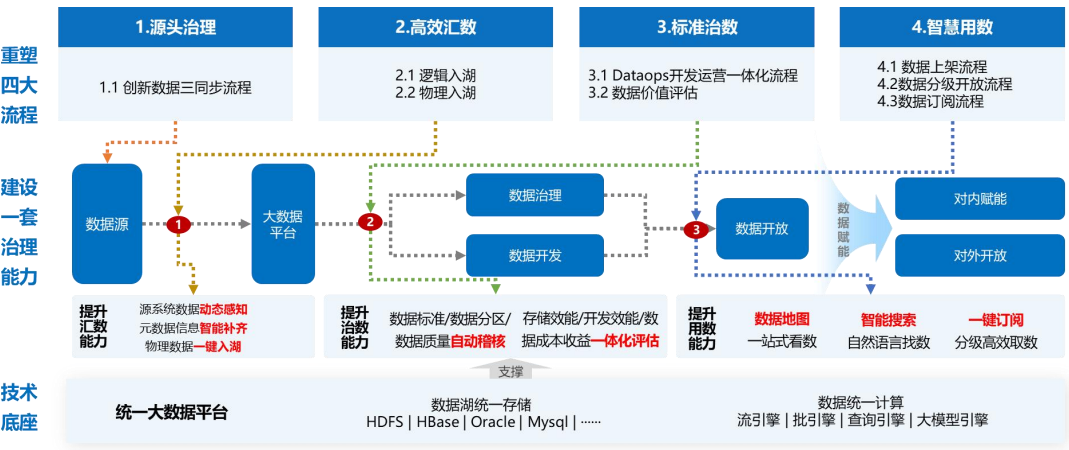


图 6: 河南移动全过程“强数”方案

## 2. 人工智能赋能智能治数，提升数据运转效率

在智能治数方面，河南移动通过打造数据自动感知能力，自动发现内外部源端数据变化，支撑存量系统元数据统一纳管；通过打造“数

据治理助手”大模型应用，实现公司数据目录中 7+万张数据资源表，200 万个字段的字典信息智能生成，单表字典信息整理平均耗时由天缩短至秒级，数据目录的字典信息完整率从 40%提升至 100%，字典信息准确率超 95%，节省数据字典整理人工成本 3000 人天。



图 7：AI 数据治理助手提升数据运转效率

在数据研运一体化方面，河南移动借鉴 DataOps 构建高效跨域协同机制、制定敏捷开发流程、打造研发治理一体化流水线、精细化数据运营等核心理念，统筹数据开发及治理全过程，并结合大模型能力提升智能化能力，构建数据开发治理全流程管理体系，建设开发质量稽核、数据质量管控、数据生命周期管控、数据价值评估等能力，并制定相应管理规范，将数据治理相关工作融于数据开发流程，实现“开发即治理”，促进数据汇聚、数据开发、数据供给、数据使用、数据退役全流程管理提效、质量提升。通过上述举措，实现可信源认证 3000 余个，数据质量问题定位的效率提升 50%，低价值数据识别 500T，开发维护效率方面也提升明显，数据模型上线实现“建管一体”。



为全面落实集团平台集约化与两级协同战略部署，河南公司以数据治理为核心基础，建成集约化“九合一”智慧运营平台，全面贯通“察、策、打、评”四大业务流程环节，实现了客户全生命周期精准洞察、全链路智能化运营以及执行过程的全程可视化。平台使用率达到 99%，存量中台派单综合成功率为 31%，渠道管理效率提升 80%，有效增强了一线运营的协同能力和响应速度。

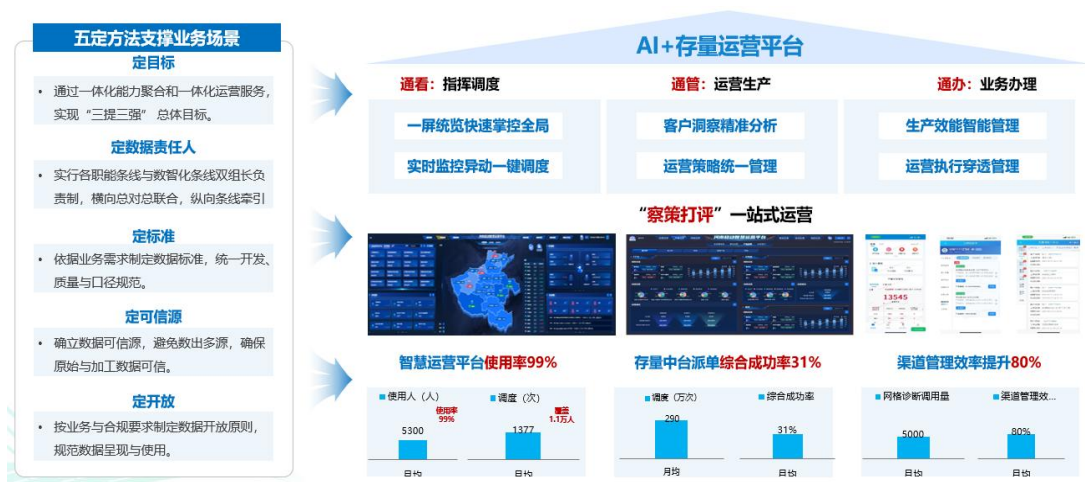


图 10：AI 数据治理驱动业务运营全流程升级

#### 4. 业务全链条注“智”，赋能业务效率提升

围绕数智化转型战略要求，河南移动通过强化数据质量管理，充分引入大模型技术，推动人工智能技术与业务发展深度融合。在业务分析、意图识别、智能策划、产品推荐、渠道质检及网格报告等关键场景实现全面赋能，月均系统调用超 2200 次，识别有效商机 420 万+，智能话术准确率达到 95%，自动质检覆盖 10 类异常问题，月均自动派发报告 4000+次，网格覆盖率达到 100%，全面提升数据洞察能力与一体化运营水平。



图 11: AI 数据治理赋能业务效率提升

### 三、 通信智能化数据治理实践

数据规模的爆炸式增长使得运营商面临着传统数据管理手段成本高、效率低、准确率不足等难题。随着人工智能技术发展，通过自动化、智能化手段辅助开展数据标准、数据质量、元数据、数据资产管理等工作，可以显著降低数据治理成本、提升数据运转效率，提升企业数据管理智能化水平。深入开展人工智能使能企业数据治理实践，已不仅是数据治理的技术上的进步，更是运营商未来向信息科技公司转型的核心驱动力。

#### （一）人工智能驱动数据标准智能化管理

引入人工智能后，通信运营商在数据治理活动中的数据标准管理将经历深刻变革，其变化主要体现在自动化、智能化、实时监控与动态调整等方面，而其价值则体现在效率提升、成本降低、数据质量增强以及战略决策支持等多个维度。

##### 1. 人工智能驱动的数据标准管理带来的变化

###### ■ 数据标准管理的自动化与智能化

传统数据标准管理依赖人工制定、执行和监控，流程繁琐且效率低下，人工智能的引入将彻底改变这一模式。例如，可借助人工智能大模型自动识别数据字段的业务含义，并根据历史数据分布生成符合标准的定义，提升数据标准建设效率。此外，人工智能能够通过自然语言处理技术解析业务需求，自动生成数据标准文档，减少人工干预。例如，某企业通过 Data Copilot 工具通过人工智能算

法在 10 万+数据标准中快速匹配所需标准，解决了数据管家长期面临的“标准建设难题”。

#### ■ 数据标准执行的实时监控与动态调整

人工智能技术赋予数据标准管理实时监控能力。例如，河南移动通过人工智能引擎可基于无监督学习检测数据流中的异常波动，如某地市公司销售出现异常波动时，系统会实时告警并推荐可能原因，将问题从“事后发现”转变为“事中拦截”。此外，人工智能能够根据业务变化动态调整标准，例如在数据安全领域，当新的安全要求出台时，人工智能可自动更新数据标准并通知相关系统，确保合规性。

#### ■ 数据标准的跨部门协作与统一化

人工智能可以打破传统数据标准管理中“部门墙”的困境。通过人工智能驱动的协作工具，如河南移动通过数据治理助手，不同部门可基于统一的人工智能模型进行数据标准的制定与执行，避免“同名不同义”或“同义不同名”的问题。例如，河南移动 33 个数据治理业务技术融合的团队，通过数据治理工具支持跨部门的元数据管理、数据标准管理、数据质量管理等模块，确保重点系统的核心数据标准在全公司范围内的统一。

#### ■ 数据标准管理的效率与准确性提升

人工智能大幅降低了数据标准管理的人力成本。例如，传统人工标注日均仅能处理 50 张实体表，而人工智能驱动的工具可将效

率提升数倍。同时，人工智能通过机器学习优化数据标准的清洗流程，例如通过数据治理工具自动去除重复数据、填补缺失值，确保数据标准的执行质量。

## **2. 人工智能驱动的数据标准智能管理价值**

### **■ 降低数据治理成本，提升效率**

人工智能的引入直接减少了数据标准管理的人力与时间成本。例如，某运营商通过人工智能智能治理方案，将年数据治理成本从千万级降至百万级，年累计节省人天 1500 人天，同时数据错误率从 5% 降至 1% 以下；此外，利用人工智能数据标准自动化生成及监控工具使数据标准管理的效率提升 60% 以上。

### **■ 提升数据标准与质量的一致性**

人工智能通过数据清洗、增强等技术使得数据标准准确率及效率提升，从而显著提升了数据质量，人工智能驱动的数据质量检测可识别逻辑冲突、标签一致性问题，并自动修复或告警，在运营商数据分析领域，人工智能模型通过无监督学习建立数据质量基线，确保数据字段符合业务逻辑，从而将数据质量错误率从 5% 降至 2%。

### **■ 促进数据共享与价值释放**

统一的数据标准是数据共享的前提。人工智能通过数据标准智能化管理，确保数据在不同系统、部门间的语义一致性和格式统一性，从而打破“数据孤岛”。例如，某企业数据治理工具支持跨系统的元数据采集与分析，使数据共享效率提升 40%。此外，高质量的

数据标准为人工智能模型训练提供了可靠输入，间接推动了数据价值的释放。

### ■ 增强业务决策支持能力

数据标准的智能化管理确保了数据的准确性和完整性，为业务决策提供了可靠依据。某企业在数智营销领域，人工智能通过数据标准管理优化客户画像数据，使 5G-A 用户精准营销的转化率提升 20%。同时，人工智能生成的数据质量报告可为管理层提供实时业务洞察，缩短管理决策周期。

### ■ 推动数据治理可持续发展

人工智能通过自动化与智能化降低了数据标准管理的技术门槛，使非技术人员也能参与数据治理。某企业通过自然语言交互实现低代码数据处理，使数据工程师的工作效率提升 30%。此外，人工智能的持续学习能力使数据标准管理能够适应快速变化的业务环境，确保数据治理的长期有效性。

## （二）人工智能协助提升数据质量管理效率

数据质量是数据价值的基石。在人工智能技术的驱动下，数据质量管理工作正经历着从“事后抽查”到“事中感知、事前预警”的根本性变化。

### 1. 人工智能对数据质量管理带来的变化

#### ■ 从规则驱动到“规则+智能”双驱动

传统方式的数据质量管理主要依赖专家经验预设数据质量规

则，如字段长度、数值范围、枚举值检查等，这种方式对于未知的数据异常模式显得捉襟见肘；在人工智能驱动的数据质量管理智能化实践中，通过构建人工智能模型（如通过无监督学习、时间序列预测等）可以学习数据在历史周期内的分布、关联和变化规律，自动识别出海量数据中偏离模式的“异常点”，例如，运营商如果出现某个基站的流量数据在凌晨时段突然异常飙升，即便在系统中没有提前预设“流量上限”规则，人工智能也能敏锐地捕捉到这一反常现象并告警。

#### ■ 从抽样检查到全量监控

传统数据质量管理受限于计算资源，通常对海量数据（如运营商通话业务详单、业务稽核数据、CDR 话单、网络信令数据、日志数据等）进行抽样检查，存在“漏网之鱼”；在智能化实践中，利用分布式计算和人工智能算法，可以对流入数据湖、数据仓库的全量数据进行实时或准实时的数据质量扫描与评估，提升数据质量稽核的全面性。

#### ■ 智能数据血缘与影响分析

当利用人工智能智能检测到某个核心数据源出现质量问题时，可以自动触发影响性分析。通过已构建的数据血缘图谱，人工智能能快速定位到下游有哪些数据模型、BI 报表和人工智能应用会受到影响，并自动通知相关责任人，实现精准、快速的故障定界与溯源。

#### ■ 智能数据清洗与修复

对于识别出的脏数据，人工智能不仅能告警，还能尝试进行智能修复。例如，福建移动在市场智慧运营过程中，对于业务运营中缺失的用户行为信息，可以根据其前后时间段的行为数据，通过轨迹预测模型进行合理插补。对于格式不一致的数据，可以通过自然语言处理（NLP）技术进行智能归一化。

## 2. 人工智能数据质量智能化管理价值

通过引入人工智能辅助数据质量管理，将显著提升数据质量管理效率、降低人工成本、增强数据信任，保障业务高质量发展。

- 提升效率：利用人工智能，将数据工程师从繁琐、重复的数据质量稽查工作中解放出来，从而专注于更高价值的任务。
- 降低成本：通过人工智能早期发现并修复数据质量问题，避免了“垃圾数据进，垃圾应用出”所带来的巨大业务决策风险和返工成本。
- 增强信任：提供更全面、更精准的数据质量报告，增强了业务方、人工智能模型对数据采集供给方的信任度，推动人工智能驱动质量管理文化落地。
- 保障业务：通过对数据质量的实时感知与快速响应，保障了实时营销、在线故障诊断、客户服务、管理决策等关键业务的稳定运行。

## （三）人工智能赋能数据安全智慧化

在数据成为核心生产要素的今天，数据安全和隐私保护是通信运

营商的生命线，也是法规遵从的硬性要求。人工智能技术为应对日益复杂和隐蔽的安全威胁提供了新的利器。

## 1. 人工智能赋能数据安全管理的变化

### ■ 从静态权限控制到动态智能风险识别

传统数据安全主要依赖于基于角色的访问控制（RBAC），权限一旦授予，在回收前长期有效，存在内部滥用和数据泄露的风险；在智能化数据安全实践中，利用用户行为分析技术，人工智能可以持续学习每个员工和数据开发者的正常访问模式（如访问时间、频率、数据量、操作类型）。一旦检测到异常行为（如非工作时段批量下载敏感客户资料），系统会实时进行风险评分并触发二次认证、操作拦截或告警，实现动态的、自适应的安全防护。

### ■ 从关键字匹配到智能敏感识别与分类

传统数据安全通过正则表达式匹配身份证、手机号等具有固定格式的敏感信息。但对于无法用简单规则描述的敏感信息（如一份包含客户隐私的投诉工单文本）则难以识别；在智能化实践中应用自然语言处理（NLP）和深度学习模型，可以智能地扫描非结构化数据（如工单日志、客服录音文本、邮件），准确识别出其中的敏感内容、商业机密或个人隐私，并自动打上分类和分级标签，为后续的加密、脱敏或访问控制提供依据。

### ■ 敏感数据智能脱敏、模糊化与匿名化

在数据开发、测试和共享场景中，人工智能可以支持更复杂的脱

敏策略。例如，在保持数据统计分布和关联关系的前提下，生成高度仿真的合成数据，既满足了数据利用的需求，又彻底消除了隐私泄露的风险。

## **2. 人工智能赋能智能化数据安全管理价值**

通过引入人工智能辅助数据安全管理工作，将帮助数据主动防御能力提升、促进合规增效、推动数据共享、保护品牌及信任。

- **主动防御：**智能化数据安全管理工作将数据安全防线从“边界防御”和“事后追溯”前置到“实时感知”和“事中阻断”，有效防范内部威胁和高级持续性威胁。
- **合规增效：**引入人工智能后，可以借助人工智能模型，自动化地完成数据资产盘点、敏感数据发现与分类分级，帮助运营商快速满足数据安全相关法律法规、个人信息保护等相关法规的合规要求，大幅降低审计成本。
- **促进共享：**在智能化的安全技术保障下，企业可以更放心地在内部乃至生态伙伴之间进行安全的数据共享与流通，释放数据的更大价值。
- **保护品牌与信任：**有效避免因数据泄露事件导致的巨额罚款、客户流失和品牌声誉受损，维护企业与亿万用户之间的信任关系。

## **（四）人工智能使能智能元数据管理转型**

引入人工智能技术后，通信运营商在数据治理活动中的元数据管

理将经历深刻变革，其变化主要体现在自动化、智能化、实时监控、标准化以及协作与可追溯性等方面。这些变化不仅提升了数据治理的效率和准确性，还为企业带来了显著的价值，包括成本降低、合规性增强、决策支持优化以及数据资产化等。

## **1. 人工智能使能智能元数据管理变化**

### **■ 元数据自动化与智能化能力提升**

传统元数据管理高度依赖人工操作，例如手动编写元数据、手工检查元数据质量等，效率低下且容易出错。通过人工智能的引入通过自动化和智能化技术显著改变了这一状况。例如，河南移动通过智能化元数据管理，使得人工智能大模型能够根据字段名和样例数据自动生成技术元数据，减少了人工标注的工作量。此外，人工智能还能自动发现数据中的模式和关系，识别敏感数据并进行分类、脱敏处理，从而提升元数据管理的效率和准确性。

### **■ 元数据实时监控与动态管理**

人工智能技术使元数据管理从被动响应转向主动监控。通过实时分析数据流，人工智能能够动态追踪数据的变更和流动，及时更新元数据信息。例如，基于人工智能的元数据管理平台可以实时监控数据源的变化，自动更新元数据目录，并识别潜在的数据质量问题。这种动态管理能力特别适用于处理海量、异构数据的场景，如数据湖和湖仓一体架构中的数据治理。

### **■ 驱动数据标准化与统一化推进**

人工智能技术推动了元数据管理的标准化进程。通过机器学习模型，人工智能能够自动识别和纠正不符合标准的元数据，确保数据资产的统一性和一致性。例如，某企业在数据治理中通过人工智能工具实现了元数据的标准化注册和管理，覆盖了从数据采集到数据消费的全生命周期。此外，人工智能还能整合不同来源的元数据，构建统一的数据目录，便于跨部门协作和数据共享。

### ■ 增强跨部门协作与可追溯性

人工智能技术提升了元数据管理的协作效率和可追溯性。例如，通过自然语言处理（NLP）技术，人工智能能够将业务术语与技术元数据关联起来，帮助业务人员和技术人员更高效地沟通。同时，人工智能生成的数据血缘关系图谱能够清晰展示数据从源头到目标的完整链条，便于审计和问题追溯。

## 2. 人工智能使能智能元数据管理价值

### ■ 极大提升元数据管理效率

人工智能的自动化能力显著降低了数据治理的人力成本。例如，河南移动通过人工智能技术将元数据管理效率提升了数倍，人工标注日均处理量从 50 张表提升至数百张表，同时数据错误率从 2% 降至更低。此外，人工智能能够实时监控元数据质量，减少元数据人工检查的频率，从而降低整体数据治理成本。

### ■ 增强数据合规性与安全性

人工智能技术通过自动化发现和分类敏感数据，帮助企业更好地

遵守隐私法规。例如，人工智能能够自动识别包含 PII（个人身份信息）的数据字段，并实施相应的脱敏策略，降低数据泄露风险。同时，人工智能生成的数据血缘关系图谱能够快速定位数据来源和流向，支持合规审计和风险评估。

#### ■ 支持数据驱动决策的优化

人工智能生成的元数据能够为决策提供更精准的支持。通过分析元数据中的业务元数据和技术元数据，人工智能能够自动识别数据质量问题、推断数据关系，并检测潜在的数据安全风险。例如，在数据集成过程中，人工智能能够根据元数据自动匹配数据格式和标准，减少人为错误，提升数据项目的成功率。

#### ■ 促进数据资产化与价值释放

人工智能技术通过元数据管理加速了数据资产的变现。例如，人工智能能够将数据封装成标准化的 API 服务，支持数据的灵活消费和价值变现。此外，人工智能生成的元数据目录能够帮助企业发现潜在的数据资产，并通过数据质量分析和价值评估，优化数据资源的分配和利用。

### **（五）人工智能帮助提升数据资产管理能力**

通信运营商数据资产管理领域正经历深刻变革，数据资产管理从传统的依赖人工操作转向智能化、自动化的高效模式。人工智能不仅改变了数据资产管理的执行方式，更重新定义了数据资产的价值释放路径。

## 1. 人工智能给企业数据资产管理带来的变化

人工智能技术通过自动化和智能化手段，全面优化了数据资产管理的各个环节，显著提升了效率与准确性。

### ■ 数据发现与分类的自动化

人工智能能够通过无监督学习和语义理解，自动识别非结构化数据中的敏感信息（如 PII/SPI）并打标签。例如，某企业内部采用人工智能模型扫描全集团数据，不仅基于正则表达式检测身份证号，还能通过上下文识别合同中的隐私条款，将传统人工分类的 3—5 天工作量缩短至 1~2 天。这种自动化减少了人为错误，确保数据分类的准确性提升 5% 以上。

### ■ 数据资产质量实时监测

人工智能引擎通过分析数据字段的历史分布，建立质量基线，并在异常波动时实时告警。例如，当某地区销售额突然归零时，系统能推荐可能原因，将问题从“事后发现”变为“事中拦截”。这种主动监控机制显著降低了数据异常对业务决策的影响。

### ■ 数据资产管理流程优化

人工智能驱动的工具（如自然语言处理开发助手）支持开发者用自然语言描述需求，自动生成 SQL 代码并优化。这不仅简化了数据开发流程，还减少了代码错误，提升了开发效率。

## 2. 人工智能帮助提升数据资产管理价值

人工智能与数据资产的深度融合，推动数据从“企业副产品”转变

为“核心生产要素”，其价值释放呈现以下特征：

- 数据驱动决策的智能化：人工智能通过分析海量数据，提供实时洞察支持决策。例如，“管理驾驶舱”整合战略大屏与业务分析报表，将数据转化为可执行的管理建议。这种“数据－洞察－行动”的闭环模式，使决策周期缩短 40%。
- 数据与人工智能的协同生态：企业通过统一数据平台整合数据与人工智能流程，实现数据底座与人工智能应用的无缝对接。Databricks 提出的“湖仓一体”架构进一步打破数据孤岛，使事务型数据、分析数据与人工智能数据统一存储与访问，提升了数据利用率。
- 数据资产的货币化：人工智能加速了数据资产的商业化进程。例如，某企业通过数据共享与交易机制，将数据资产转化为收入来源。同时，人工智能生成的数据集（如向量组、Token）被纳入资产管理，确保其可追溯与合规
- 数据资产管理平台的智能化升级：未来平台将集成更多人工智能能力，如自动数据清洗、智能标签推荐等，如某企业 DMAP 数小二已通过元数据管理整合企业架构、流程与数据资产，形成统一的资产管理视图。
- 数据湖与人工智能的深度融合：数据湖作为人工智能训练的数据源，将通过人工智能技术实现自动特征工程与数据预处理。“人工智能数据湖”解决方案通过打破数据孤岛，提升数据

供应效率。

## （六）福建移动人工智能数据治理业务实践

在过去一年推进“AI+应用”过程中，福建移动意识到高质量的数据成为企业人工智能表现的关键差异化因素，通过利用人工智能技术，深入开展数据治理，逐步形成高质量数据集和良好的数据伦理，积极开展市场领域大模型的推理、增训工作，促进存增一体“AI+应用”快速发展，赋能市场数智运营取得良好成效。



图 12: 福建移动 AI 数据治理赋能存增一体运营

### 1. 深入开展数据&AI 现状问题分析

在福建移动数据&AI 赋能市场业务场景，涉及较多内部业务知识、行业专业知识或敏感信息，需通过私域部署大模型并结合领域规则来实现，实际过程遇到的问题主要体现在以下几方面：

#### ■ 业务规则、业务知识尚未结构化

通用大模型缺乏对通信行业复杂业务逻辑（如客户洞察、产品互斥规则、指标口径、营销策略等）的理解，需通过深度对接公司 IT

系统（如大数据平台、经分系统、客户管理系统）提取结构化规则，如果对前端系统业务逻辑的解析覆盖不足，可能导致私域模型无法精准适配场景需求。

### ■ 业务流程的数据知识化程度不足

各生产环节（如营销方案、网格晨会、客户服务）的关键数据（如销售达成率、任务进度）还分散在不同 IT 系统中，缺乏统一的数据接口与标准化定义，需强化业务数据治理能力，构建全域数据视图以支持模型训练。

### ■ 大模型对角色意图识别能力不足

各条线对人工智能的需求都非常急迫，以大干快上的方式铺开建设，但对于人工智能背后的场景与规则梳理的重视度不足，对于各生产环节、各作业角色的关键业务诉求梳理不充分，导致询问专业领域问题时，大模型无法准确把握提问者的意图。

### ■ 深度思考和联网搜索能力不足

福建移动大模型算力资源相对有限，深度思考功能和互联网搜索对算力消耗较大，资源限制无法为所有场景提供能力，导致热点事件、行业动向的更新不如公域大模型及时，思考过程不如公域大模型有深度。

## 2. 加强规划，建立多模态数据管理体系

面向业务领域在人工智能画客、人工智能问策、人工智能智选、人工智能挖需方面强劲的人工智能市场数智运营诉求，市场部门与数

智化部门一起，重点开展以下三方面的数据治理工作。

首先，建立面向 **AI+应用的多模态数据管理体系**。数据治理适配 **AI+应用**发展，为公司 **AI+应用**做好数据采供服务，要强化非结构化的数据治理，建立结构化、非结构化数据相融合的多模态数据管理体系，为公司领域大模型发展提供高质量数据集，提升福建移动数据服务及业务智能化水平。

其次，强化面向 **AI+数据工程及知识工程能力建设**。通过了解 **AI+应用**对数据的需求，清晰公司非结构化数据现状，提升非结构化采集、存储及处理能力，开展包括数据目录、元数据、数据标准、数据流转、数据质量、数据安全及生命周期等非结构化数据管理，沉淀高质量的知识，畅通 **AI+应用**数据供给。

最后，推进**数据治理与 AI+应用更加紧密结合**。将数据治理工作融入 **AI+应用**业务流程，建立数据驱动的决策机制，确保数据价值真正服务于业务创新。一方面在业务场景选择遵循先易后难，优先开展办公、营销、客服场景；另一方面是在业务规划中同步开展数据规划，在数同步开展生产环节的数据采集和自身领域的的数据资源盘点，推动数据治理与人工智能应用的协同，实现数据飞轮效应。

### **3. 数据治理驱动 AI+应用工作举措**

福建移动在面向 **AI+应用**的数据治理不同环节，围绕在人工智能应用的设计阶段、开发阶段及运营阶段，将数据&AI 深入业务流程，工作举措包括：

## ■ 在设计阶段，数据治理驱动知识自动提取

针对长期以来业务生产主要聚焦生产任务的完成，对规则的体系化梳理和生产过程的数据化相对不足问题，福建移动通过深化开展数据治理，首先，在系统对接与规则自动化抽取方面，由数智化部门主导对接前端 IT 系统，通过 API 接口与日志分析提取隐含的业务规则，构建规则引擎供给模型调用。其次，在数据治理与场景化建模方面，建立统一数据标准，清洗历史业务数据，按场景建训练数据集，提升模型对业务术语与流程的理解能力；最后，在自动化工具方面，加强知识管理技术能力，提升效率，通过培养专业的技术团队，利用大语言模型等自动化工具高效提取和整理知识，提升知识图谱构建、模型微调和场景化知识标注能力，降低知识管理的技术门槛和人工标注成本。

## ■ 在开发阶段，数据驱动流程优化及体验提升

AI+应用的关键在于能否为用户提供流畅的交互体验、准确理解用户意图、精准定位领域知识，并最终提供高质量的答复。首先，福建移动通过建立 AI+的小军团支撑机制，推动简单场景“短频快”建设支撑，确保 3-4 月一线人员有实际人工智能+应用场景体验；其次，通过流程拆解、专家访谈输出标准化模板，通过多轮交互明确用户意图，提升用户体验，通过流程拆解，针对关键场景拆解业务环节如数据收集、分析、汇报，明确各环节所需业务知识。并通过设计标准化模板（如产品知识卡、话术清单），确保大模型知识格式统一，容易

被人工智能模型理解；最后，通过数据治理，从多角度、多维度进行分析，精准识别用户意图。在意图不明确的情况下，通过多轮交互式问答确认用户需求。回答用户问题后，可增加预设的相关问题推荐，尽可能覆盖用户潜在需求点，提升交互深度和用户体验，市场业务知识问答准确率指标达到 70%。

### ■ 在运营阶段，数据驱动大模型 AI+应用持续调优

在“AI+时代”，数据驱动持续运营的重要性被进一步凸显。人工智能推荐、问答的准确性和实用性，以及不准确回答后的反馈机制与智能体对新知识的迭代优化，都更加强调运营的作用。为应对这些挑战，福建移动通过建立数据驱动人工智能的反馈机制，不断迭代优化，通过用户反馈和系统自我监测，自动触发规则库更新或模型重训练，持续提高其准确性和实用性；同时，持续贯通业务生产系统与人工智能智能体平台的数据流，确保模型输出与业务实际状态一致；同时，通过建立人工智能应用持续运营的流程机制。实现开发运营一体化，建立跨部门的项目团队，共同参与人工智能应用的运营和优化工作。以一线反馈为衡量标准，确保一线问题真正得到解决，提升一线获得感。

## 4. 人工智能原生数据治理使能数智运营实践

福建移动以“DT”为基石，通过深入开展数据治理，构建高质量数据治理与供给体系，打通内部系统平台，实现数据融通；以“IT”为桥梁，整合 B/O/M/S 四大领域系统与共性能力平台，支撑前端应用

敏捷响应；以人工智能为引擎，打造“人工智能智脑”智能决策与运营体系，实现从数据到洞察、从洞察到行动的闭环管理。



图 13：福建移动 AI 数据治理协助构建核心运营能力

在人工智能原生的福建移动围绕“洞察－施策－生产－调度”全流程，构建并落地六大人工智能能力，形成智能化运营闭环，取得良好的业务成效。

- **在 AI 画客方面：**基于多维数据构建用户“四度三维”画像，精准识别用户行为特征与潜在需求，AI 用户画像精准度达 90%，画像查询响应时间小于 1 分钟。
- **在 AI 挖需方面：**通过自然语言理解客户产品需求，自动生成客群标签与组合，打破传统人工配置模式，转化率提升 1.5 个百分点，实现“从产品找客户”的智能化跃迁。
- **在 AI 问策方面：**基于用户画像与六大策略引擎，实现“一客一策”精准推荐，各类营销剧本转化率达 8.4%，人工智能策略查询率超 35%，显著提升营销精准度与用户体验。

- **在 AI 优客方面：**通过人工智能辅助人工客群调优，优化标签配置与客群组合，转化率提升 1.3 个百分点，形成“模型赋能人工、人工反哺模型”的协同闭环。
- **在 AI 智选方面：**实现多策略智能优选与产品推荐，结合用户画像与业务规则，输出优先级排序与情境化话术，提升推荐成功率与客户满意度。
- **在 AI 调度方面：**构建“指标问答—智能诊断—现场管理”三大能力，实现网格运营 5 分钟一键诊断，提升调度效率与执行精准度。

基于数据&AI 驱动的业务实践，福建移动在 AI+运营领域紧密结合生产实践探索提升路径，AI+应用场景研发模型从传统的“瀑布型”优化为“六阶十四步”敏捷开发，有效改善人工智能项目可复制性和规模化落地能力；规范交付流程制定人工智能+应用“四阶三级”品控标准，对智能体分层分级管理，确保达成成熟度标准才能进入下一交付阶段，提升整体人工智能应用质量与实施效率。



## 四、 通信数据智能发展趋势

在数据智能浪潮之下，通信行业数据治理工作在数据智能应用一体化开发、“人机协同”、内生安全方面进行着进一步的探索。

### （一）“数智一体化”能力体系

“数智一体化（Data intelligence development and operations，DIOps）”能力体系，是针对数据与智能融合应用的核心需求，面向场景，构建系统性能力框架，通过一体化、工程化的方式，打通数据到数据智能应用的全链路流水线，更高效、高质量地解决问题。

数智一体化能力体系旨在构建“1 个核心+ 3 类关键能力+ 3 大保障”，确保数据智能应用全链路流水线高效运转。体系以**工程化能力为核心**，贯穿数据采集、处理、模型开发、应用构建、运维监控的全生命周期，实现各环节的技术协同与流程贯通。**三类关键能力**包括一体化基础能力（基础设施、算子、数据开发、数智应用、运维监控）、持续集成机制（数据集成、模型集成）、持续优化机制（研发态优化、生产态优化），构成体系的核心。**三大保障能力**以治理能力（数据标准、质量、安全等）、运营能力（资源调度、数据运营、模型运营等）、资源管理（数据、模型、计算、存储）。

这一体系在业内已有相关实践，在通信行业的落地，将帮助解决数据智能相关开发过程中流程割裂、资源浪费、规范缺乏等问题。通过持续测试提升数据智能应用效果，通过持续监控优化总体成本和效率，通过工程化改造强化研发过程标准化水平，最终实现数据智能价

值释放。

## （二）“人机协同”技术体系

“人机协同”技术体系是以“智能体为基础、数字员工为场景化落地、人机协同为核心范式”，构建高效协同模式，推动通信运营商在数智化转型过程中，实现数据治理工作从“劳动密集”转向“智能密集”，提升运营效率、服务质量与创新能力。

**智能体**是体系基础，作为具备自主感知、分析、决策与执行能力的核心载体，整合多源数据与专业算法，通过 MCP、A2A 等方式，打通跨系统接口，实现“分析、决策、执行”闭环，构建可以支撑各类场景的智能体或多智能体协同系统。**数字员工**是场景化的智能体，聚焦标准化业务流程，可以将例如网络运维、客户服务等场景的重复任务封装为自动化流程，以模块化专业能力适配不同岗位需求，是智能体在具体场景的落地形态，核心解决流程自动化问题。**人机协同**是核心工作范式，智能体与数字员工承接结构化、重复性工作，人类聚焦复杂决策、创意策划与异常处理，通过双向反馈机制，人类经验优化智能模型，智能系统提升人类工作效率，形成可持续迭代的协同生态。

## （三）内生安全

内生安全是通信行业数据智能可持续发展的底线保障，核心是将安全理念融入数据智能全生命周期，针对人工智能应用特有的伦理风险、隐私泄露、模型安全等问题，构建“事前预防、事中管控、事后追溯”的全链条安全体系，通过建立伦理审查机制、对用户敏感数据

进行全流程脱敏处理、构建模型全生命周期安全体系等手段，确保数据智能在合规、公平、安全的框架内释放价值。

## 联系方式:

大数据技术标准推进委员会

地 址:北京市海淀区花园北路52号

邮 编:100191

邮 箱:cuiyiyang@caict.ac.cn

官网:www.tc601.com

