CAICT 中国信通院　中国中车 CRRC

集智 专题报告

# 人工智能模数共振体系研究报告（2026 年）

中国信息通信研究院人工智能研究所

中车工业研究院有限公司

2026年5月

# 版权声明

# 前　言

当前，全球新一轮科技革命与产业变革加速演进，人工智能技术正加速从单点突破向系统化赋能演进，"模型能力"与"数据要素"的深度融合与共振协同已成为驱动产业智能化转型的核心动能。党的二十大报告明确提出要构建新一代信息技术、人工智能等新的增长引擎。人工智能模数共振体系在推动数据要素价值释放、加速模型技术迭代升级、赋能产业智能化转型方面发挥着日益重要的战略作用，是支撑人工智能高质量发展的核心要素。2025 年 9 月 23 日，在 2025 人工智能产业及赋能新型工业化大会上，南京、济南、青岛、武汉、深圳等先导区和部属单位代表共同启动了“模数共振”行动。2025 年 10 月 27 日，“人工智能赋能新型工业化模数共振专题研讨会”在北京召开，会议探讨了模数共振的概念内涵与实践路径。2026 年 1 月，工业和信息化部等八部门联合发布《“人工智能+制造”专项行动实施意见》，其中提到模数共振行动相关要求，对构建数据驱动、模型赋能、应用牵引的模数共振协同发展格局具有重要意义。

人工智能模数共振体系是人工智能技术与产业应用深度融合的核心载体，其本质在于通过高质量数据集与高效能模型的双向共振，实现“以模引数、用数赋模”的良性循环。该体系以分层分类、精准赋能为原则，通过构建行业通识与专识数据集，培育行业大模型与特色智能体，并探索建立跨主体的“模数共振空间”与生态协同机制，打通数据流通壁垒，完善算力供给、标准规范与安全治理体系，为“人

工智能+”应用落地和各行业数字化转型提供全方位支撑。

本研究报告首先阐述了人工智能模数共振体系的具体定义和内涵，全面总结了模数共振的三大核心要素、五大核心基础能力支撑和三大协同运行机制，并提出模数共振下一步落地发展的具体建议，可为政策制定者、行业从业者及企业投资者等提供全面的行业洞察、策略建议与决策依据。面向未来，人工智能模数共振体系仍存在诸多问题与挑战，还需要产学研各界紧密合作，共同推进模数共振技术创新与产业发展，为“人工智能+”全面落地提供有力支撑。

# 目 录

# 图 目 录

# 一、模数共振定义与内涵

人工智能模数共振体系是推动人工智能与实体经济深度融合发展的系统性工程，其核心要义在于实现高质量数据集、高效能模型、高价值应用三大要素的协同共振与价值倍增。该体系以数据要素为根基、以模型能力为枢纽、以场景赋能为导向，通过构建数据驱动模型进化、模型赋能应用创新、应用反哺数据积累的良性循环机制，打通从数据资源到智能服务的全链条价值通路。模数共振体系是连接数据治理、算法创新产业数字化转型的关键纽带，是释放人工智能乘数效应的核心载体，是培育新质生产力的重要引擎。在当今智能化浪潮加速演进的时代，构建完善模数共振体系已成为抢占人工智能发展制高点、赋能千行百业智能化升级的战略支撑。

## （一）模数共振具体内涵

**“模数共振”体系具体是指建立数据质量提升、模型优化与应用反馈的协同联动及闭环迭代机制，实现数据动态适配模型需求、模型输出反哺数据质量提升，**旨在通过数据汇集、标注、合成、治理与管理全方位提升数据质量，并以高质量数据为底座夯实大模型训练、生成、推理能力，激活数据场景应用价值，有效破解 AI 模型训练中数据量不足、质量参差、场景适配性差等瓶颈，为大模型研发、智能装备升级、生产流程优化提供关键支撑。

传统人工智能数据集构建模式呈现出典型的“线性断裂”特征，往往止步于简单的预处理后即投入训练，导致训练过程中产生的反馈

信号被截断，缺乏对原始数据的有效检验与修正机制。这种“一次性交付”的粗放模式，使得数据集无法根据模型表现进行动态调优，进而引发场景覆盖存在盲区、特征提取能力孱弱、质量管控失效等系统性问题。在此割裂的体系下，模型难以习得稳健的泛化特性，导致其在复杂的业务实践中落地效果大打折扣，无法形成从数据到应用的价值闭环。

随着人工智能从通用技术向垂直场景深化，高质量数据集的建设已超越了单纯的“规模堆砌”，演进为一种动态的“闭环迭代生态系统”。这一系统构建了“原始数据—训练微调—测试评估—反向优化”的全链路流转机制，将数据的全生命周期管理与模型的进化周期深度耦合，形成了“数据滋养模型、模型反哺数据”的共生共荣格局。其实质是利用科学化、结构化的流程设计，精准破解数据质量与模型需求错配、训练效果与实际应用脱节、应用需求与技术迭代断层等核心痛点。在以数据为中心的新一代人工智能范式中，核心目标在于推动数据集从“被动供给”向“主动适配”转型，构建起具备自我进化能力的 AI 数据闭环，具体如图 1 所示。

人工智能模数共振体系通过循环反馈机制通过“数据处理—模型训练—性能检测”三位一体构成完整的循环圈，把数据和模型绑定在一起，再把数据利用的效能嵌入到闭环体系中，不断改进提高训练效果，从根本上提高整个闭环体系的数据集训练效率和准确度。

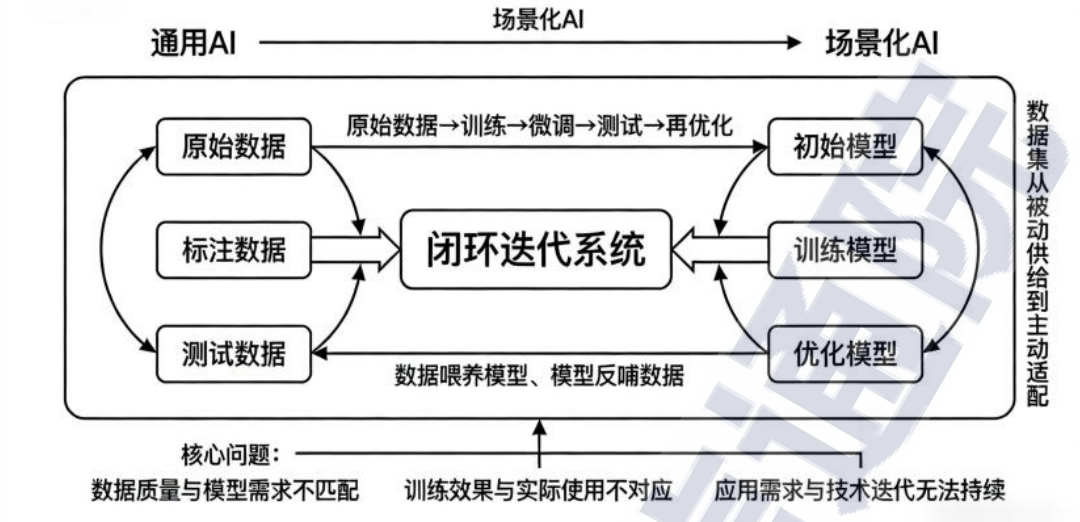


图 1　AI 数据闭环迭代系统流程

## （二）模数共振必要性分析

模数共振体系是对人工智能“数据-模型”关系的一种再认识，由“数据供给决定模型能力”的单向逻辑转变为“模型需要引导数据进化、数据进化反哺模型升级”的双向主动机制，解决好数据质量与模型需求精准对接的问题，实现“好数育好模，好模引好数”，为人工智能技术由“实验室”迈向“产业界”提供坚强保证。

**从技术层面来讲，**闭环迭代体系的核心价值的是实现人工智能从“静态训练范式”向“动态持续进化范式”的根本性转变，打破了传统 AI 模型“训练-部署-停滞”的静态瓶颈。依托闭环架构，人工智能系统能够构建起“环境交互-信息感知-自主优化”的持续循环机制，持续接入真实场景中的动态数据、环境变量及反馈信号，在预设的算法边界与安全规则框架内，自主完成参数调优、模型迭代与策略适配，实现从“被动响应”到“主动进化”的跨越。同时，智能算法将深度渗透于数据

处理与模型优化全流程，构建“问题自动识别—策略自动生成—优化自动执行—效果实时反馈”的全链路自动化闭环，通过自动化工具覆盖数据采集、清洗、标注、模型训练、测试验证、部署运维等各个环节，大幅降低人工干预成本，提升系统运行效率、迭代速度与优化精度，推动 AI 技术从“可用”向“好用、耐用”持续升级。

**从产业层面来讲，**随着基础算法的趋同化与算力资源的普惠化，高质量数据集的规模与迭代体系的敏捷度已取代单一的技术指标，成为决定企业核心竞争力的关键变量。自适应闭环迭代体系正演变为重塑人工智能产业格局的决定性力量，具备成熟闭环能力的企业，能够依托“数据飞轮”效应，针对垂直场景的细粒度需求做出毫秒级反馈，在高频的实战演练中持续打磨模型精度。同时，在反复的迭代循环中，不断沉淀高质量、高价值的场景化数据资产，形成“数据积累-模型优化-效果提升-数据增量”的正向循环，持续深化自身技术优势与场景适配能力，逐步构建起难以被复制的技术壁垒与竞争优势，在同质化竞争中实现差异化突围，推动人工智能技术与产业场景的深度融合落地。

未来，人工智能闭环迭代体系将朝着“智能化、自动化、一体化”发展，数据和模型的深度融合，最终实现“数据即模型、模型即数据”的共生状态，实现人工智能的自我进化成长。这样才能不断激活高质量数据集的价值，带动人工智能技术进入“数据驱动、自主进化”的新阶段。

## 二、模数共振三大核心要素

**人工智能模数共振体系以高质量数据集、高效能模型、高价值应用为核心要素，构建“数据驱动模型进化、模型赋能场景落地、场景反哺数据迭代”的闭环生态。**高质量数据集作为基石，通过汇聚行业核心知识与多模态信息，为模型提供高价值、高密度的“燃料”；高效能模型作为引擎，融合通用能力与行业机理，实现从云端训练到边缘推理的精准适配；高价值应用作为出口，锚定产业刚需场景，推动 AI 从辅助工具升级为生产主体，最终形成“模数共生、价值倍增”的智能化发展新范式。

### （一）高质量数据集

**高质量数据集是指用于训练、验证和优化人工智能大模型而收集、整理、标注形成的覆盖行业核心专业知识和生产经营活动信息的数据资源集合。**高质量数据集覆盖制造、金融、医疗、交通、公共安全、自然资源、地理信息、人力资源、社会治理、科学研究等重点行业的公域数据和私域数据，具有高技术含量、高知识密度、高效益场景的“三高”特征。

**一是高技术含量。**当前，高质量数据集的建设已突破传统劳动密集型的“人工标注”模式，全面迈入自动化与智能化的新阶段。这一变革的核心在于技术链路的深度重构：一方面，利用大模型辅助生成、扩散模型增强及主动学习技术，实现高价值样本的智能筛选与合成数据的自动化生成，显著提升了数据清洗与标注的效率与精度；另一方

面，引入隐私计算与差分隐私技术，构建起全链路的数据安全防护网，确保敏感数据在脱敏状态下的高保真与高合规性。这种“人机协同、智能驱动”的技术体系，不仅大幅降低了数据处理成本，更通过算法的自我迭代与优化，解决了复杂场景下数据获取难、标注难的问题，为模型训练提供了坚实的技术底座。例如，百度开发的人工智能数据智能标注平台内嵌百种智能算法，助力数据标注效率提升 50%；英伟达运用生成对抗网络创建的自动驾驶合成数据集，可模拟暴雨、暴雪等极端天气场景，有效缓解了恶劣驾驶环境数据不足的问题。

**二是高知识密度。**高知识密度是高质量数据集的核心价值所在，其本质在于单位数据中蕴含了高度浓缩、结构严谨的专业知识与跨模态信息。这要求数据集建设必须超越简单的信息堆砌，通过深度清洗、去重、知识抽取与结构化标注，剔除冗余噪声，将文本、图像、语音、点云等多维数据与行业机理、专家经验及科学规律深度融合。从通用语料向医疗、工业、自动驾驶等垂直领域延伸，高质量数据集正在形成“数据+知识”的复合型知识体系。这种高密度的知识注入，赋予了 AI 模型更强的逻辑推理能力与泛化能力，使其能够理解复杂的业务逻辑与物理世界规律。例如，自动驾驶数据集 Waymo Open Dataset 整合了激光雷达点云、交通标志语义解析等多模态数据，涉及计算机视觉、交通工程等学科知识；生物医药领域，AlphaFold 使用了 20 多万个蛋白质结构构成的数据集，集成了 X 射线晶体学、冷冻电镜成像等多学科实验数据，构建出覆盖 98.5%人类蛋白质的三维结构图谱。

**三是高效益场景。**面向工业制造、医疗诊断、智慧教育、自动驾驶等重点行业领域，高质量数据集建设正从“通用基础化”向“深度定制化”转型，重点覆盖极端、稀缺及长尾场景，有效解决了数据偏见与样本失衡难题。通过构建领域专属数据集并进行微调与知识增强，能够快速适配行业特定任务，驱动各行业实现智能化升级与价值倍增。这种“场景牵引、效益导向”的建设模式，确保了数据要素能够直接转化为生产力，在提升诊断准确率、降低缺陷过检率等方面展现出巨大的经济与社会价值。例如，美国国立卫生研究院发布的胸部 X 光数据集 ChestX-ray14，汇集了 11 万张标注精准的医学影像，支撑 AI 辅助诊断系统使肺炎识别速度达到人类的 160 倍；苏州阿丘科技构建了 500 多万张 PCB 图片数据集，研发出识别上百种缺陷的模板，实际应用将缺陷过检率降低了 90%。

## （二）高效能模型

**高效能模型是指具备深度行业适配能力、高推理效率与强泛化能力的 AI 大模型体系，涵盖基础大模型、行业大模型及场景小模型的协同架构，能够实现通用能力与专业场景的精准匹配。**高效能模型以“懂行业、通机理、能落地”为核心特征，通过云-边-端三级部署体系满足各个重点行业领域场景的实时性、可靠性、安全性的严苛要求。

**一是高算效比。**高算效比强调在有限的算力预算或参数规模下，实现接近甚至超越超大模型的性能水平。这一特征的核心技术路径包括轻量化网络架构设计（如深度可分离卷积、稀疏注意力机制）、知

识蒸馏（以大模型指导小模型训练）、模型量化与剪枝（降低数值精度或移除冗余连接）以及混合精度训练等。通过这些技术，模型在训练阶段可大幅减少浮点运算次数和显存占用，在推理阶段能够部署于边缘设备、移动终端或嵌入式系统中，实现实时响应。高算效比还体现在“小参数、大智慧”的设计理念上，即通过更高质量的数据和更优的训练策略，让中小模型在特定任务上超越盲目扩规模的巨量模型，从而降低部署成本和能耗，推动人工智能从云端走向万物互联。

**二是高泛化性。**高泛化性是指模型能够从有限训练样本中提取出具有普适性的规律，并在未见过的新数据、新任务或新领域上保持良好表现。实现高泛化性的关键技术包括自监督预训练（在海量无标签数据上学习通用表征）、元学习（学会如何学习，快速适应新任务）、域自适应与域泛化（消除训练域与目标域之间的分布差异）以及正则化方法（如 Dropout、权重衰减等）。高泛化性使得模型无需大量人工标注即可迁移到下游场景，显著降低了针对每个新任务重复收集和标注数据的成本。在实际应用中，具备高泛化能力的模型能够应对数据分布的自然漂移（如季节变化导致的路面外观改变），并且仅需少量微调甚至零样本即可完成跨场景、跨模态的任务适配，极大提升了人工智能模型的复用效率和规模化落地能力。

**三是高鲁棒性。**高鲁棒性是指模型在面对输入噪声、数据分布偏移、对抗性攻击、传感器故障或环境突变等干扰因素时，仍然能够维持稳定、可靠的预测性能。高鲁棒性的构建依赖于多种技术手段：数

据增强与对抗训练（在训练过程中主动加入扰动，让模型学会抵抗干扰）、不确定性估计（输出预测置信度，在不可信时主动拒绝或求助）、模型集成（融合多个模型的输出降低偏差）以及因果推断（避免学习虚假相关性）。高鲁棒性是人工智能从实验室走向真实复杂场景的关键保障，尤其是在自动驾驶、医疗诊断、工业控制等高可靠性要求的领域，模型必须能够处理极端天气、设备老化、恶意攻击等长尾风险。一个高鲁棒性的模型不仅要求平均准确率高，更要求在最恶劣情况下的性能下降可控，从而确保系统整体的安全冗余和容错能力。

### （三）高价值应用

**高价值应用是指依托人工智能高质量数据集与高效能模型两大核心要素，深度融入各行业生产经营全流程，能够精准解决行业核心痛点、显著提升生产效率、创造明确且可观的经济与社会价值的人工智能落地场景集合。**覆盖制造、金融、医疗、交通、公共服务等多个重点领域，贯穿行业研发、生产、管理、服务、售后的全产业链，是人工智能技术从理论创新、技术研发走向实际应用、实现价值转化的最终体现。高价值应用具有场景刚需化、价值可量化、产业深度化的“三化”核心特征，三者共同构成了高价值应用的核心内涵，确保应用场景贴合产业实际、应用价值可落地、应用范围可拓展，推动人工智能与产业深度融合。

**一是场景刚需化。**高价值应用的核心导向是解决实际问题，精准锚定各行业发展中的核心痛点与刚需场景，摒弃脱离产业实际的“无

用应用”，聚焦传统模式下难以突破的瓶颈问题，为行业高质量发展提供核心支撑。其核心逻辑是围绕产业生产经营中的关键环节、薄弱环节，结合高质量数据集与高效能模型的技术优势，打造针对性的落地场景，填补传统模式的短板。无论是基层医疗资源短缺、工业质检效率低下，还是交通管控压力大、政务服务繁琐等痛点，高价值应用都能提供切实可行的解决方案，成为推动行业转型升级的重要动力。

**二是价值可量化。**高价值应用与普通应用的核心区别在于，其能够产生明确、可衡量的经济与社会价值，而非模糊的“赋能”概念，价值可通过效率提升、成本降低、风险管控、服务优化等具体指标进行量化衡量，形成可落地、可推广、可复制的价值闭环。从经济效益来看，可体现为生产成本降低、生产效率提升、营收增长等；从社会效益来看，可体现为服务质量提升、公共资源利用效率提高、社会治理水平优化等。这种可量化的价值，不仅能够让企业、机构清晰看到人工智能应用的实际成效，也为应用场景的迭代优化、规模化推广提供了重要依据。

**三是产业深度化。**高价值应用不再局限于单一环节、单一场景的浅层赋能，而是深度融入行业全产业链、全流程，推动产业模式重构、业态创新，助力产业实现全方位、深层次的数字化、智能化转型。其核心是通过高质量数据集与高效能模型的深度融合，打通产业各环节的数据壁垒，实现生产、管理、服务等各环节的协同优化，推动产业从“传统模式”向“智能模式”转型，从“大规模生产”向“个性化

定制”转型。这种深度赋能，不仅能够提升单个环节的效率与质量，还能推动整个产业的转型升级，培育新的产业业态、新的增长动能，实现产业高质量发展。

## 三、模数共振五大能力支撑

模数共振体系中核心功能的实现是借助“数据集设计和构建、数据集质量评估、模型微调测试、模型基准测试验证、数据增强与优化”五个核心环节相互循环来完成的，即形成一个完整“反馈—优化”的闭环。数据集设计和构建是闭环反馈体系的起点，数据集质量评估是高质量数据集构建的技术保障，微调测试环节实现数据与任务的初步适配，模型基准测试验证完成模型性能的全面检测，数据增强与优化则根据检测结果对闭环反馈体系进行全面迭代升级，这几个环节形成完整的“反馈—优化”闭环，并且各环节自成体系，同时又互相联系互相结合。

### （一）数据集设计与构建

**一是锚定模型的基础数据诉求。**数据设计首先要能将抽象的模型需求拆解为具体的数据指标，并能够实现“需求可量化、指标可落地”。对于需求拆解应当从“模型类型、能力基线、场景范围”三方面开展工作。做好需求拆解就必须做到“精准匹配”，既不能为了追求数量多而不考虑数据本身的质量，又不能因为数据量少而无法支撑模型的基本能力，。如要开展通用语言模型的预训练数据设计，应当把焦点放在对“语义覆盖度、语法多样性、语域全面性”的拆解上，保障数

据具有支撑模型学习通用语言规律的能力；当开展专业领域语言模型的设计时，要在通用需求的基础上增加“行业术语密度、专业场景占比”的专项指标，以促进后续模型的应用。

**二是构建多维度数据供给体系。**提取原始数据的渠道是影响数据集质量的重要因素，因此对于数据来源的规划要充分考虑“多样性、可靠性、合规性”三个方面的因素，从多渠道、多角度入手建立多元化的供给体系，避免出现数据单一来源导致的过拟合问题。数据源主要来源于“公开数据集、自有业务数据、第三方采购数据、主动采集数据”四种，不同数据源所对应的运用场景及适配范围不同。根据需求拆解需求后，需要确定各类数据来源，并依据所需权重分别调用不同来源数据，并组合使用。构建通用预训练集时可采用“公开数据集为主、自有数据为辅”的方式，利用公开数据提供量级和多样化保障，同时增加自有数据的占比来满足质量和相关性；构建专业领域预训练集时，则应采用“自有数据为主、第三方数据与主动采集为辅”的方法，在自有数据确保对应业务场景适用性的基础上，增加其他数据补充对应场景覆盖度及样本多样性。

**三是构建标准化的数据标注流程。**一方面，数据处理标准化主要实现数据格式化、数据特征统一，如对所有文本都采用相同的编码方式及长度、所有图像具有统一的分辨率、色彩空间以及所有音频都具有统一的采样率和时长等，确保模型训练能够接收标准格式的输入数据。另一方面，针对不同任务和模态数据，处理要求标准也不尽相同。

举个例子，数据清洗的主要目的就是剔除掉“无效数据”和“错误数据”，比如剔除掉文本里的非文本信息，修正图像中的错别字像素，填平表单中缺失值等。清洗需要根据不同的数据模态制定对应规则，即针对文本型的数据需重点解决语法错误、语义重复、语义不清等问题，针对图片型的数据需重点解决分辨率不足、照明差异大、角度不正确等问题，针对音频型的数据需重点去除背景噪声、音量不均、时长太短等问题。

**四是建立全流程的数据质量保障机制。**人工智能数据集质量管理是数据设计和构建过程中的“最后防线”。通过建立“过程检查+结果验收”全链条机制，将模型训练基础要求落实到具体数据集上，以验收标准倒逼数据预处理各阶段完成数据质检任务。过程检查侧重于各个数据处理关键点，在清洗、去噪、标准化步骤结束后分别开展质量抽检，并将存在的问题反馈回前序处理工作；结果验收采取对整体的数据集进行全面的质量验收方式，按照制定最终验收质量指标进行分项打分和问题筛查。

## （二）数据集质量评估

大模型时代，人工智能高质量数据集已成为决定人工智能系统性能上限与可信水平的关键基础，高质量数据集不仅能够提升模型的泛化能力，还能减少过拟合的风险，使得人工智能系统更加稳健。人工智能数据集质量评估体系旨在对数据集的完整性、准确性、一致性、多样性、时效性及合规性等核心数据质量维度进行科学量化与综合评

测，为人工智能模型研发与产业落地提供坚实、可信、可持续的数据支撑，助力人工智能向安全、可控、高质量方向发展。

在《关于促进数据标注产业高质量发展的实施意见》《关于促进数据产业高质量发展的指导意见》《关于加快公共数据资源开发利用的意见》《“数据要素×”三年行动计划(2024-2026 年)》等政策的指导下，2025 年 12 月，中国信通院迭代发布“动静结合”的可信 AI 人工智能数据集质量评估体系 2.0，覆盖人工智能数据集质量评估评估标准、评估指标、评估工具以及评估方案等四大核心维度。**评估标准方面，**具体包括国家标准《高质量数据集 质量评测规范》和正式发布的工信部人工智能行业标准《面向人工智能的数据集质量通用评估方法 总体要求》（YD/T 6486-2025）。**评估指标方面，**具体包括通用基础质量指标体系和不同重点行业专属质量指标体系，涵盖说明文档、前置数据质量、模型应用质量三大核心评估维度。**评估工具方面，**采用分层随机抽样+自动化评估+人工辅助校核的评估方式，迭代开发人工智能数据集质量评估工具平台 2.0 版本。**评估方案方面，**通过全量指标迭代、专属指标筛选、侧重权重设计、算子规则匹配以及安全方案对齐实现不同行业、不同类型数据集定制化测试服务方案，覆盖文本、图像、音频、视频、多模态、结构化数据、传感器数据、时间序列等多种数据模态。具体体系参照图 2 所示。


图 2 可信 AI 人工智能数据集质量评估体系 2.0

## （三）模型微调与优化

微调测试是“通用预训练”和“场景化应用”之间的连接点。通过“指令微调数据集构建—模型参数调整—适配效果测试”方式把预训练模型由“拥有通用能力”转变为“掌握执行某项任务的能力”。所以，微调测试的过程是“在数据驱动下适配模型”，使用结构化指令信息数据来指引模型学习的方向，然后用测试的方式验证是否达到了针对该任务所进行的模型适配，得到可以用于评估基准测试的优化方案。

**一是指令微调数据集的结构化设计。**任务解构是构成指令的前提条件，要将复杂而繁杂的场景任务拆分为“输入格式、处理逻辑、输

出标准”这三大要素。以“智能客服的问题回答”这一任务为例，首先确定其输入格式为“表述用户提出的问题”，处理逻辑为“区分用户的问题种类、理解用户的提问意图、形成智能机器人的对应答复”，最后确定输出标准为“保证答案的准确度、完整性以及符合公司及产品风格的话术表达”，基于此要素，每一条指令都需要使用“结构化指令——指令对应的响应”的方式进行设计，且每一条指令的数据信息均由“指令描述”“输入内容”“期望输出”三个部分组成。

**二是模型参数的适应性调整逻辑。**微调过程就是模型参数的“局部优化”，模型在指令微调数据集上做小范围训练，从而修改一些预训练模型的参数，使得模型更加适应所要完成的任务。与预训练阶段的“全局参数更新”不同，微调阶段的参数更新需要遵循“精准触达、最小干预”的准则，在保持模型通用能力的基础上提升模型完成任务的能力。参数调节需要实时监视参数调优过程，当在验证集上发现损失值变小并且持续下降，或者准确度提升时，保持原设置，适当地将学习率降下来，并减少训练轮次。如果在验证集上发现准确率不再上升，应停止训练。这样，在降低模型过拟合风险的同时，也不丢失原有的普适能力，并使模型根据具体的任务得到较好的适配度。

**三是适配效果的初步测试与问题诊断。**将微调后的模型拿来运行初步测试，主要是检验模型适配情况，及时发现“适配不足”或者“适配偏差”等问题，给后面的基准测试和优化提出方向指导。初步测试主要侧重于“快速定位问题”，而不像后面所说的基准测试那样，“全

面评估性能”，一般采用“小规模测试集+核心指标”方式。问题诊断旨在进行模型性能“问题归因”，更好地为模型迭代提供改进方向，例如增加一些相应的场景数据，修改指令表述，或者调整参数。

### （四）模型性能基准测试

模型基准测试验证是闭环体系中的“全面质检环节”，涵盖大模型的测评指标、方法、数据集等多项关键要素，是指导大模型落地实践的规范，旨在通过对微调后模型进行全面、系统性的性能检验，保证模型在选定场景下符合业务要求，主要检验模型对特定任务的执行能力及评价模型性能的数据质量问题，并据此修正进一步优化数据。大模型基准测试验证核心价值不是“评估性能”，而是“定位根源”，通过建立“指标拆解—问题列举—归因定位”的循环机制，准确找出因什么数据问题造成什么模型性能瓶颈，再根据两者之间的关联性找到影响最大的数据质量源头问题。

为形成一个全面、公正、科学、规范和客观的大模型测试体系，精准掌握我国大模型产品技术情况、长短版和面临的挑战，为企业研发改进方向提供建议，中国信通院联合多加头部大模型企业、用户单位和科研机构共同发布“方升”大模型基准测试体系 3.0，如图 3 所示。“方升”大模型基准测试体系 3.0 涵盖大模型基准测试的四大关键要素，即指标体系、测试方法、测试数据集和测试工具。其中，**测试能力方面，**主要规定了测试维度与指标，覆盖大模型的基础属性测试（Basic-Oriented Test，BOT）、通用能力测试（General- Oriented Testing，

GOT）、应用能力测试（Application- Oriented Testing，AOT）、行业能力测试（Industry-Oriented Testing，IOT）、未来高级智能测试（Advanced Intelligence-Oriented Test，AIOT）。**测试工具方面，**按照“一个体系+一套方法+N 个评测数据集”的工作思路，搭建一个大模型基准测试工具平台，集成评测数据、评测管理、大模型对战及评测排行等多方面综合功能，为业界提供科学、客观、全面的大模型评估服务。


图 3 “方升”（FactTesting）大模型基准测试体系 3.0

面向未来，大模型基准测试体系发展主要呈现出以下几个发展趋势：**一是高质量评测数据集持续增加，**围绕复杂推理、多模态、代码及智能体应用等重点领域和重点行业方向，行业场景专属高质量评测数据集将会持续新增和完善，满足多语言、多任务、多场景下的模型评测与优化需求。**二是构建体系化研究和先进测试方法，**聚焦大模型评测流程中的关键技术卡点，未来将突破高质量测试数据合成与质量

评估、数据污染检测及人机对齐裁判模型构建等核心技术；同时围绕通用人工智能演进趋势，将率先构建高级智能能力的评测范式，实现对未来智能水平的前瞻性度量与引导。**三是开发新一代大模型基准智能评测基座，**围绕智能体应用场景，未来将新增多智能体交互与环境感知的仿真测试环境，满足复杂真实场景下智能体协同交互、动态环境适应能力的系统性测试与评估需求；同时构建一体化基准评测系统，集成动态自适应测试工具、高级智能能力评估工具及评测数据全生命周期管理工具，实现评测能力的自动化、可扩展与前瞻性统一。

### （五）数据增强与优化

数据增强与优化构成了模数共振体系的“价值升华闭环”，其本质是基于模型基准测试与真实场景反馈，对数据集构建全流程进行的逆向重构与系统性迭代。这一过程绝非简单的“数据修补”或“增量补充”，而是通过建立“数据质量感知—模型性能评估—应用效果反馈”的双向联动机制，实现数据要素价值的动态跃升。它要求我们在接收到反馈信号后，打破原有的数据静态边界，针对数据处理规则、数据集拓扑结构及核心内容要素进行重新定义与深度重塑，从而确保数据供给始终与模型进化及业务需求保持高频共振。

**一是实施基于性能反馈的“精准靶向”优化策略。**“精准靶向”是数据优化的核心法则，旨在杜绝“无的放矢”的盲目迭代。依据基准测试暴露的模型短板与归因分析，我们将优化路径细分为三个维度：首先是“数据质量修复”，针对模型出现的幻觉或逻辑错误，回溯至

源头清洗噪声数据、修正错误标注，提升数据的信噪比与准确性；其次是“数据结构调整”，针对模型在特定类别上的表现失衡，通过重采样或长尾场景挖掘，优化数据的类别分布与特征空间覆盖度；最后是“数据内容补充”，针对模型泛化能力不足的问题，利用合成数据生成或跨域数据融合，填补关键场景的知识盲区。这种分层分级的优化策略，确保了每一次数据迭代都能精准击中模型性能的痛点。

**二是推动全流程数据处理规则的“迭代升级”。**数据优化不应止步于解决当前数据集的存量缺陷，更应着眼于构建长期的质量保障体系，即对数据全生命周期的处理规则进行系统性升级。遵循“问题追溯—规则分析—标准迭代”的逻辑闭环，我们将单次优化中发现的共性问题转化为通用的处理标准。例如，当发现特定领域的术语识别率低时，不仅要补充该领域数据，更要更新分词规则、实体抽取模板及标注规范。通过这种从“治标”到“治本”的规则演进，我们能够在源头上提升数据生产的标准化水平，形成一套随技术演进而自我进化的数据处理规范体系，确保持续产出高质量的数据资产。

**三是构建闭环优化的“持续运行与长效保障”机制。**数据优化是一项贯穿数据集全生命周期的动态工程，而非一劳永逸的静态建设。随着模型应用场景的拓展、业务逻辑的变更以及底层算法技术的迭代，数据与模型之间必然会产生新的“适配摩擦”。因此，必须建立一套具备自我进化能力的长效保障机制。这要求我们部署自动化的监控与评估流水线，实时捕捉模型在生产环境中的性能衰减或异常表现，并

触发相应的数据增强流程。通过建立“监测—评估—优化—验证”的自动化闭环，确保数据集能够像生物体一样，随着外部环境的变化而持续新陈代谢，始终保持对模型训练与应用的高效支撑能力。

## 四、模数共振三大协同机制

模数共振体系有效运转的前提是模型数据映射关系达到“数据供给与模型需求”匹配，拥有自适应测试系统作为“性能验证与问题诊断”的工具，以及能产生持续优化迭代能力来实现“自我进化与持续升级”，三者互相促进、互相提高，在技术上共同构筑起一套有效的高质量数据集建设与模型动态优化迭代的方法论。

### （一）建立模型-数据关联映射关系

模型-数据映射关系是人工智能模数共振体系的“导航系统”，是将具体的“输入数据特征—模型能力需求—输出性能目标”实现一一匹配，所有的数据才能变成模型的“营养品”，而不是简单的以数据量来衡量是否达到要求，而需要实现“特征维度适配”，所以要从模型类型、任务场景、性能指标三个维度考虑。

**一是基于模型类型的映射设计。**不同技术架构的人工智能模型对数据分布与特征的提取逻辑存在本质差异，因此必须锚定模型的底层技术特性，构建差异化的数据映射策略。针对自然语言处理模型，映射设计应聚焦于“语义连贯性、句法规范性、语境多维性”，重点构建能够支撑模型捕捉长距离依赖与深层语言规律的语料空间；针对计算机视觉模型，则需确立以“像素级清晰度、特征显著性、场景覆盖度”

为核心的映射准则，确保数据能够充分支撑模型对视觉信息的抽象与重构；而面对多模态大模型，核心在于建立“跨模态语义对齐”的映射机制，通过高维特征空间的投影与匹配，实现文本、图像、音频等异构数据在语义层面的深度融合与互补，从而为模型提供全维度的感知能力。例如，OpenAI 在开发 GPT-4o 模型时，通过引入海量的网络文本与代码数据，强化了模型对长距离语义依赖的理解，从而实现了极高的语义完整性和语境丰富性；而在开发 DALL-E 3 时，则构建了高质量的“文本-图像”配对数据集，确保视觉特征与语言描述精准对齐，完美体现了不同类型模型对数据映射的不同侧重。

**二是基于任务场景的映射优化。**当模型部署指向特定垂直领域时，数据映射关系需完成从“通用泛化适配”向“场景深度定制”的范式转变。通用数据集往往因包含大量非任务相关的冗余噪声，导致模型在特定任务中出现算力空耗与特征聚焦偏差。因此，场景化映射必须通过“高价值特征筛选—领域权重动态分配—精细化场景标注”的链路，实现数据分布与任务目标的精准耦合。这一过程旨在剔除通用背景噪声，强化领域特异性特征的表达，使数据流能够精确响应业务逻辑，从而在降低训练成本的同时，最大化模型在特定场景下的执行效能与决策精度。例如，百度在构建其“文心”系列行业大模型时，针对搜索与推荐场景，专门筛选并注入了大量的长尾搜索词与专业领域文档数据，通过“场景标注”强化了模型对特定意图的识别能力；特斯拉的自动驾驶系统则通过筛选极端天气和罕见路况的视频数据进行加权训

练，大幅提升了模型在复杂驾驶场景下的决策准确性。

**三是基于模型性能的映射校准。**模型-数据的映射关系的有效性需由模型的各项性能指标来评判并进行纠正和校准，在形成“指标反馈—映射调整—性能提升”的良性循环后，不同的性能指标要对应不同的数据优化方向，才能不断趋近理想的状态。如果模型的“准确率”不够好，则应改善数据的“特征－标注值”映射关系，降低错误的数据对模型的判断造成的影响；如果模型的“召回率”不够高，则应该优化数据的“场景－特征”映射关系，补充遗漏的场景关键信息；如果模型的“泛化能力”不够强，则应该对数据的“多样性－分布平衡性”映射进行改善，并且加入更多没见过的场景样本。例如，Meta 在优化 Llama 系列模型时，通过建立大规模的人工审核数据集（如 Llama-2-Chat 的 SFT 数据），修正了大量错误的标注映射，显著提升了模型的准确率；谷歌 DeepMind 在开发用于蛋白质结构预测的 AlphaFold3 时，通过引入全球蛋白质数据库中更多未见过的物种样本，改善了数据的多样性分布，从而使其在泛化能力上达到了前所未有的高度。

## （二）创新模数闭环迭代能力机制

持续优化迭代是模数共振闭环体系“生命力源泉”，“规则迭代、技术迭代、机制迭代”共同支撑“反馈—分析—优化—验证”这一个闭环，让数据集构建由“一次性项目”变为“持续性工程”，并能不断地吸纳技术更新迭代、场景变革以及模型反哺带来的新的需求点，呈螺旋式提升数据集的质量。

**一是基于规则迭代的数据处理标准动态升级。**数据处理规则是保证数据集质量的重要基础，包括清洗规则、标注规则和标准化规则等，但是规则本身的有效性会因为场景的变化和技术的发展出现衰减。规则迭代需要沿着“问题收集－效果评价－规则修订－落地验证”的过程展开，首先通过监控模型的测试结果及实际使用数据情况获得“数据质量问题反馈”，然后通过量化评估的方式来确定已存在的规则是不是合理的，并把规则问题对数据质量的影响予以量化的评估；据此再次调整和修正规则，如修正规则清洗算法参数、修改规则清洗标注手册和更新规则清洗标准标准化模板；最后将经过调整后的规则再一次放到样本数据上批量清洗，并根据清洗优化前后整体数据的质量及模型效果，验证本次规则迭代是否达到预期的效果，从而使后续的数据处理紧跟技术进步和业务需求变化。例如，优必选在打造 Thinker 通用基座模型时，建立了“弱监督+自监督+少量人工校验”的动态迭代优化体系，将模型训练后的误差反馈至标注流水线，持续优化标注算法参数，实现了标注成本降低 99%，效率提升超百倍。

**二是基于技术迭代的工具方法持续革新。**随着技术工具的持续迭代创新，高质量数据集建设已突破传统“人工辅助”的效能瓶颈，全面迈向“智能自动化”新阶段，这种技术革新为闭环体系注入了强劲动力，实现了数据处理、标注及分析全流程的自动化流转与智能化升级，显著提升了数据供给的效率与质量。数据清洗阶段，应用智能清洗技术可以将“预训练语言模型+规则引擎”用于识别和修复数据中的语法

错误、语义冲突等，还可检测图像中的模糊及失真等情况，并以较高质量较高效率完成清洗工作。数据标注阶段，利用“半监督学习+主动学习”的方式，只需要少量的人工标注作为样本来训练标注模型，之后进行大规模数据标注，再通过主动选出“高价值样本”交给人工来复审确认，既确保了标注精度又大幅降低了成本。数据分析阶段，利用实时分析的方法实时对模型运行过程中产生的数据反馈信息进行加工处理，及时发现其中存在的问题，并对其原因进行快速分析定位，为其优化迭代提供实时支持。应用技术迭代构建的闭环体系，能够更好地完成闭环体系内的任务以及模型调优等操作，是对技术迭代机理的高效、充分应用。例如，百度智能云的多模态 AI 数据智能生产平台，通过集成自研的智能预标注模型，可通过动态模型优化机制和项目数据特征自动调整预标注模型参数，在语音、图像等场景实现了 90%以上的自动标注覆盖率。

**三是基于机制迭代的组织流程的协同适配。**持续的技术迭代必须依托协同适配的组织流程才能发挥最大效能。传统的“部门割裂”式作业模式—即业务部门采集、技术部门处理、算法团队训练的线性流程，往往导致需求传导阻滞、方案落地困难等“孤岛效应”。为打破这一僵局，必须构建“跨部门协同+全流程贯通”的保障机制。通过设立聚合多方资源的 AI 资源协调机构，建立“周度反馈、月度评估、季度优化”的常态化闭环流程：业务端实时反馈数据适配痛点，技术端动态评估工具适用性，算法端精准测算模型与数据的契合度。同时，配套“优化

效果与绩效挂钩”的激励机制，倒逼各技术协同部门主动融入迭代循环，从而形成“全员参与、协同优化”的良性生态，确保技术工具与规则标准能够持续演进。例如，微软 Azure AI 构建“业务-技术-算法”三位一体的跨部门协同体系，打破传统数据生产的部门壁垒，通过统一数据中台与共享工作流，实现从需求提出、数据采集、处理标注到模型训练、部署应用的全流程打通，显著缩短人工智能模型训练优化迭代周期并提升落地效率。

### （三）构建模型自适应性能测试系统

模型自适应测试系统作为闭环体系中的“质检中枢”，功能在于突破“固定测试集+单一指标”局限性，在测试时通过不同模型类型、任务场景和训练阶段等参数自适应生成测试集，并结合不同模型特点灵活选择多元测试指标与测试方法，测试完毕后根据模型以及任务的不同特点，自适应采集测试结果与日志信息，以精准测试结果反哺模型。相较于传统测试方式，自适应测试系统具有“场景自适应、指标自适应、反馈自适应”的特点，能够全面暴露数据和模型的问题，具体可参照图 4 所示。

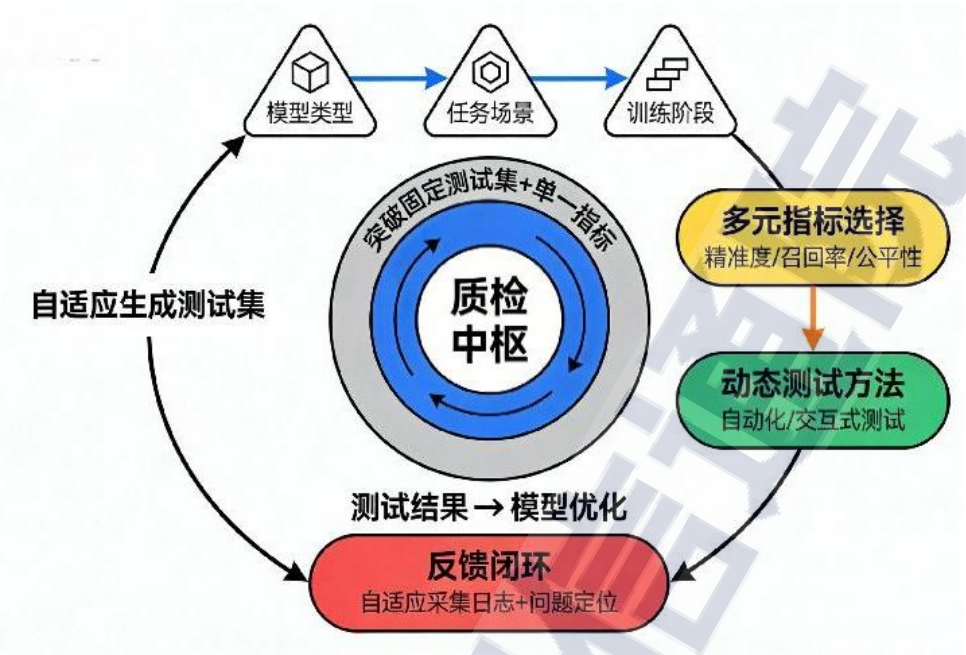


图 4　模型自适应测试闭环体系

**一是通过场景自适应覆盖全维度测试空间。**不同于传统测试集的“静态划分”模式，传统方式从训练数据中抽取部分样本作为测试集，易出现“测试场景与实际应用场景脱节”的痛点——测试集涵盖的场景与真实业务场景脱节，大量高频真实使用场景未被纳入测试范围，导致测试结果无法有效反映模型实际部署后的性能表现。自适应测试系统依托“场景分层分类+动态样本补充”的核心逻辑，先对全量实际应用场景进行分层拆解，再根据场景出现频率、复杂度动态调整测试样本分布，实现对所有潜在应用场景的全面覆盖，确保测试过程与真实应用场景高度一致。例如，谷歌 DeepMind 在 AlphaFold 系列模型测试中，采用场景自适应测试方案，先将蛋白质预测场景按物种、结构复杂度分层，再动态补充罕见蛋白质结构样本，避免传统静态测试的场景遗漏问题，大幅提升模型在真实科研场景中的适配性。

**二是通过指标自适应构建多维度模型评价体系。**模型性能的全面评估离不开多维度指标的支撑，单一评价指标存在局限性，无法真实、全面反映数据质量优劣与模型核心能力水平，易导致模型评价偏差，影响后续优化方向的判断。自适应测试系统可基于模型类型、任务场景的差异，自动生成“基础指标+核心指标+特色指标”的三层评价体系，基础指标保障模型基本性能达标，核心指标聚焦任务场景核心需求，特色指标适配具体行业个性化要求，实现对模型与数据的全维度、精准化评测。例如，商汤科技 SenseTime 自适应测试平台，在自动驾驶场景测试中，构建“基础指标（响应速度）+核心指标（目标识别准确率）+特色指标（极端天气适配性）”的三层体系，全面评测模型在复杂路况下的性能，为模型优化提供精准指引。

**三是通过反馈自适应实现模型-数据精准问题定位。**对自适应测试系统而言，发现模型性能不达标仅完成问题解决的第一步，精准定位问题根源才是推动模型与数据优化的关键。传统测试方式仅能判定模型性能“不达标”，无法有效区分问题根源是数据质量缺陷（如样本标注错误、特征缺失）还是模型算法漏洞（如架构不合理、参数设置偏差），导致后续优化工作盲目无序、效率低下。自适应测试系统通过“指标分层分解-问题归因分析-路径反向追溯”的闭环逻辑，对每一项不达标指标进行拆解，精准锁定问题成因，明确区分数据层面与模型层面的问题，为后续针对性优化提供清晰指引。例如，Meta 在 LlaMA 系列大模型测试中，依托反馈自适应机制，当模型生成准确率

不达标时，通过指标拆解与路径追溯，快速区分是训练数据标注偏差还是模型注意力机制设置问题，大幅提升优化效率。

## 五、模数共振落地发展建议

### （一）统筹推进行业数据集建设与模型优化

面向制造、金融、医疗、教育、交通、能源、政务等重点行业领域，整合公域数据与私域数据，建设覆盖行业核心专业知识和通用业务场景的高质量通识数据集，以此为基础训练具备广泛行业理解能力的行业大模型，形成支撑多场景应用的通用智能底座；另外，聚焦特定细分领域和专业化场景，构建高知识密度、高标注精度的专识数据集，研发面向垂直场景的特色智能体，实现从通用能力到专业技能的精准适配。通过通识与专识数据集协同布局、行业大模型与特色智能体联动发展，构建分层递进、有机融合的行业智能化赋能体系，全面提升人工智能服务实体经济的专业化水平与实用化效能。

### （二）持续完善模型性能评测能力机制

面向行业大模型、特色智能体等多样化能力评估需求，构建特色化、定制化的评测数据集体系，充分发挥评测数据集在模型能力诊断中的基准标尺作用。针对不同行业应用场景的专业特性，建立覆盖知识理解、逻辑推理、任务执行、安全可控等多维度的分级分类评测标准，形成科学权威的模型能力评估框架。强化评测结果的应用导向，将模型评测反馈作为行业高质量数据集建设和优化的重要依据，精准识别数据短板与能力缺口，指导数据集定向扩充与质量提升。通过构

建"评测诊断—数据集定向优化—模型能力提升"的良性循环机制，推动评测体系从单一评分向持续改进赋能演进，实现模型性能迭代与数据质量升级的双向促进，为人工智能产业高质量发展提供可靠的评价基准和优化路径。

### （三）探索建立模数共振生态协同机制

推动各地区选择第三方中立机构、各央企选择集团公司或专业子公司作为建设运营主体，打造"模数共振空间"创新载体。重点研发能够承载跨主体数据汇聚和模型训练的软硬件基础设施，构建高性能算力支撑、数据安全流通、模型协同开发的技术底座；同步制定跨主体数据协同、模型共建、收益分配、责任划分、安全保障的全链条管理机制，明确各方权责边界与利益共享规则，破解数据孤岛与信任难题。通过技术架构与制度设计的双轮驱动，构建开放协同、安全可控的模数共振生态，形成数据供给、模型研发、场景应用的有机联动，为人工智能大模型训练和行业应用提供可持续的要素保障与协作平台。

### （四）加强模数共振关键要素保障

着力夯实技术、标准、人才和生态等核心要素基础，为模数共振体系长效发展提供坚实保障。技术创新层面，实施模数共振“专项行动”或者"揭榜挂帅"等机制，聚焦人工智能算法模型、数据集质量评估、数据标注与增强、数据工程工具等关键领域开展攻关，突破核心技术瓶颈。标准引领层面，积极参与国际标准、国家标准及行业标准

的研究制定，并通过人工智能模数共振"专项标准行"等宣贯活动，推动人工智能数据工程等重点标准的广泛落地应用，提升我国在全球人工智能治理中的话语权和影响力。生态培育层面，创新开展重点行业人工智能赋能新型工业化"深度行"系列活动，搭建模型企业、数据企业、应用单位等多元主体的常态化交流合作平台，通过建设实训基地、组织项目对接、推广标杆成果等方式，促进产业链上下游深度协同与价值共创。人才建设层面，构建跨学科培养体系，重点培育既懂行业业务场景、又精通数据科学和模型技术机理的复合型高端人才队伍。

# 编制说明

本研究报告自 2026 年 1 月启动编制，分为前期研究、框架设计、文稿起草、征求意见和修改完善五个阶段。本报告整体分析了模数共振具体定义与内涵，全面梳理总结了模数共振发展的三大核心要素、五大基础能力支撑以及三大协同运行机制，并进一步提出了落地发展建议，为政策制定者、行业从业者及企业投资者等提供全面的行业洞察、策略建议与决策依据。本报告由中国信息通信研究院人工智能研究所撰写，撰写过程中得到了人工智能大模型及软硬件评测工业和信息化部重点实验室的大力支持。

参编单位: 中车工业研究院、沈阳市数据局、沈阳数字经济协会、北京市政数局、北京市经信局、东莞市政数局、东莞市万江街道办事处、砺英数智、海天瑞声、中电信人工智能公司、国家呼吸医学中心、广州国家实验室、招商局集团、中国建筑、中国物流、中国东方航空、中国商飞、中国矿产资源集团大数据有限公司、中汽信息科技( 天津 ) 有限公司、国家电投集团能源科学技术研究院、上海航空工业( 集团 ) 有限公司、公安部大数据中心、交通部科学研究院、中国科学院成都文献情报中心、煤炭科学研究总院有限公司、中央财经大学、中央民族大学、广东外语外贸大学、甘肃省甘谷第一中学、都市圈大数据(北京)科技有限公司、兰州新区商贸物流投资集团有限公司、兰州新区大数据投资建设管理有限公司、地衣子集（福建）数字科技有限公司、数据磐石科技（北京）有限公司、中国信通院河北科技创新研究院、

国网四川省电力公司电力科学研究院、中标政联咨询(北京)有限公司、河南多鲸信息技术有限公司、陕西淘丁数据科技有限公司、南方医科大学珠江医院、红星新闻、之江实验室、北京信息基础设施建设股份有限公司、沈阳东软智能医疗科技研究院有限公司、中科斯欧(合肥)科技股份有限公司、华为技术有限公司、希尔贝壳、陕西省大数据集团、甘肃鑫亿隆汽车服务有限公司、厦门弘信电子科技集团股份有限公司、陕西金隽人才科技有限公司、勤为科技有限公司、河北天泰信息技术有限公司、中关村数智人工智能产业联盟、扬州人工智能集团有限公司、上海金润联汇数字科技有限公司、上海市工商联数字经济商会。

中国信息通信研究院 人工智能研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62301618

传真：010-62301618

网址：www.caict.ac.cn