



2025 AI+ Development  
Digital Summit

# AI+ 研发数字峰会

拥抱AI 重塑研发

05/23-24 | 上海站



# 2025 AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

下一站预告

08/08-09 | 北京站

11/14-15 | 深圳站



查看会议详情

## 北京站论坛设置

大模型和 AI 应用评测

智能存储与检索技术

下一代知识工程

AI+ 金融业务创新

智能需求工程

智能体与研发效率工具

AI 产品运营与出海策略

大模型安全与对齐

大模型应用开发框架与实践

智能体经济 (Agentic Economy)

智能测试工具的开发与应用

具身智能与机器人

代码生成及其改进

AI+ 新能源汽车

AI 前沿技术探索与实践

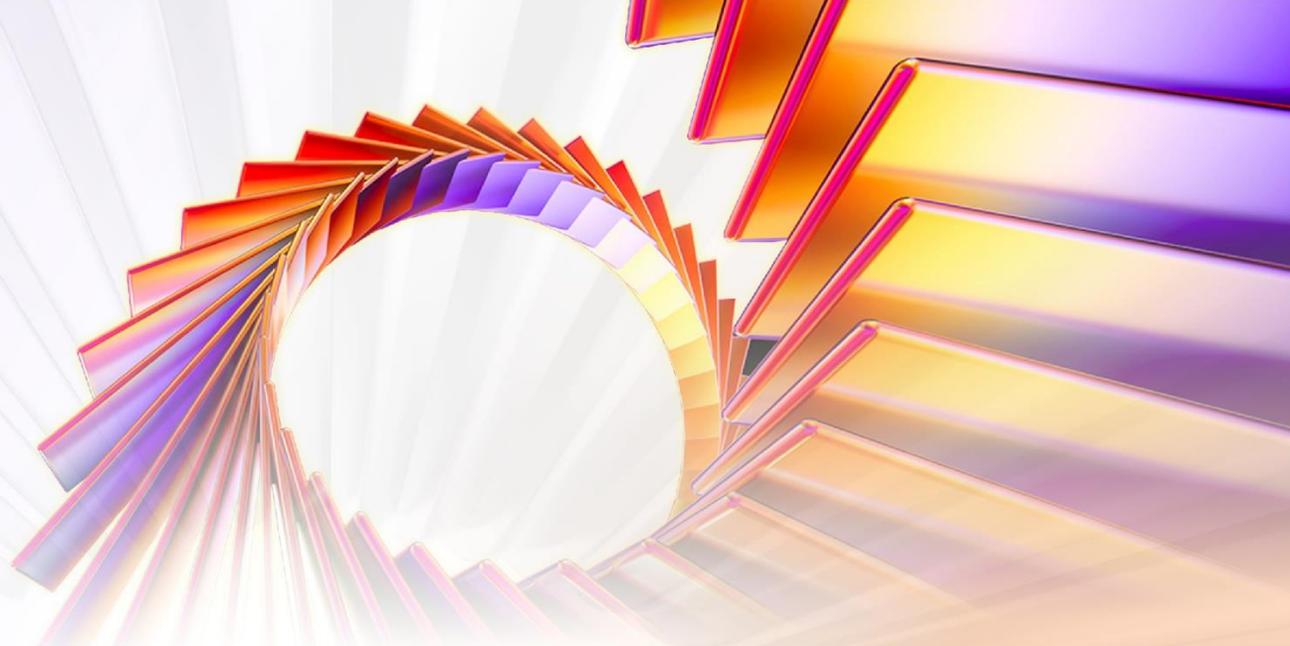


| 05/23-24 | 上海站

**2025** AI+ Development  
Digital Summit

**AI+研发数字峰会**

拥抱AI 重塑研发



# DeepSeek R1-MoE-NAS 融合赋能领域大模型高效训练体系

张延 | 联通数字政府产品部



## 张延

联通数字政府产品部 交付总监

---

现任联通数字科技有限公司数字政府产品部交付总监，拥有高级技术职称。加入中国计算机学会、中国人工智能协会、ACM 协会、中国电子学会，并在部分协会中担任职务。深耕人工智能与大数据领域多年，凭借算法优化与技术创新推动企业数据资源整合标准化，为数字化转型提供核心支撑；主导智能应用领域关键系统研发与解决方案落地，助力机构提升运营效率，加速数字化进程并推动行业智能化升级。

# 目录

## CONTENTS

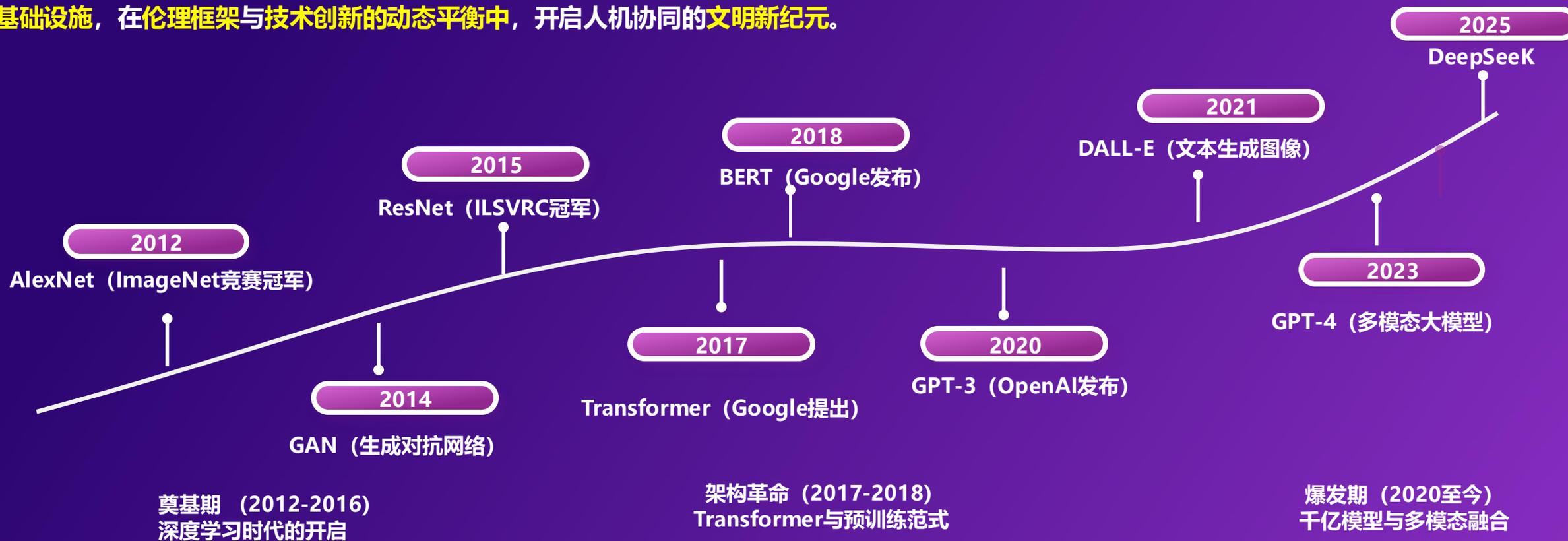
- I. 领域大模型发展瓶颈剖析：数据、模型与应用挑战
- II. 技术赋能破局：DeepSeek R1 与 MoE/NAS 协同的领域大模型优化策略与方案
- III. 渐进优化：领域大模型全流程优化框架
- IV. 实践验证：政务领域大模型应用成效
- V. 技术融合与展望：领域大模型优化策略的总结与未来探索

## PART 01

# 领域大模型发展瓶颈剖析：数据、 模型与应用的挑战

# 大模型技术发展浪潮：从奠基到爆发

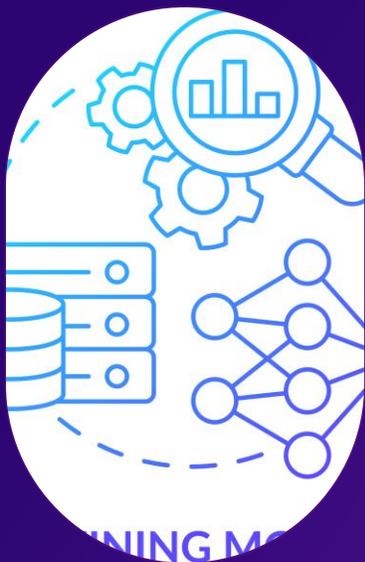
模型技术历经三次浪潮：从 AlexNet/ResNet 突破视觉瓶颈**开启深度学习时代**，到 Transformer/BERT 重构 NLP 范式，最终在 GPT-3/DALL-E 等千亿参数模型推动下**实现多模态融合与通用智能演进**，持续突破算力效率与模型泛化能力边界。这场技术革命正加速重塑人类社会。随着 GPT-4 等万亿参数模型涌现，AI **从工具属性向协作主体演进**，通过开源生态共建共享，或将成为驱动科学发现、产业升级与社会治理的**通用基础设施**，在**伦理框架与技术创新的动态平衡中**，开启人机协同的**文明新纪元**。



# ▶ 领域模型的必要性：从通用性到精准性的跨越

当今人工智能发展，从**通用性**迈向**精准性**是关键趋势。大模型通用性强，但在垂直领域存挑战。领域模型以独特优化策略，破解数据治理、模型适配及应用落地难题，成为实现跨越的关键，推动AI进入“**领域深耕**”精准赋能新阶段。

通用模型



## 领域模型的必要性：从通用性到精准性的跨越

数据分布差异：领域数据具有独特模式，通用模型难以适配；  
 知识深度不足：通用模型在医疗、金融等领域的专业知识准确性受限；  
 计算成本高昂：千亿参数模型在边缘设备部署成本高。

领域模型



### 大模型核心特点与优势

- 参数规模驱动：千亿级参数突破符号逻辑处理能力，实现复杂知识推理；
- 泛化能力跃升：通过零样本/少样本学习适应跨领域任务；
- 多模态融合：支持文本、图像、语音统一建模；
- 涌现能力：在参数临界点后展现出超越训练数据的能力。

### 领域模型应运而生，通过三大优化实现突破

- 数据预处理：针对领域数据特点设计特征提取模块；
- 专业模型：融合领域知识（MoE）；
- 轻量应用化部署：通过神经架构搜索（NAS）生成专用模型。

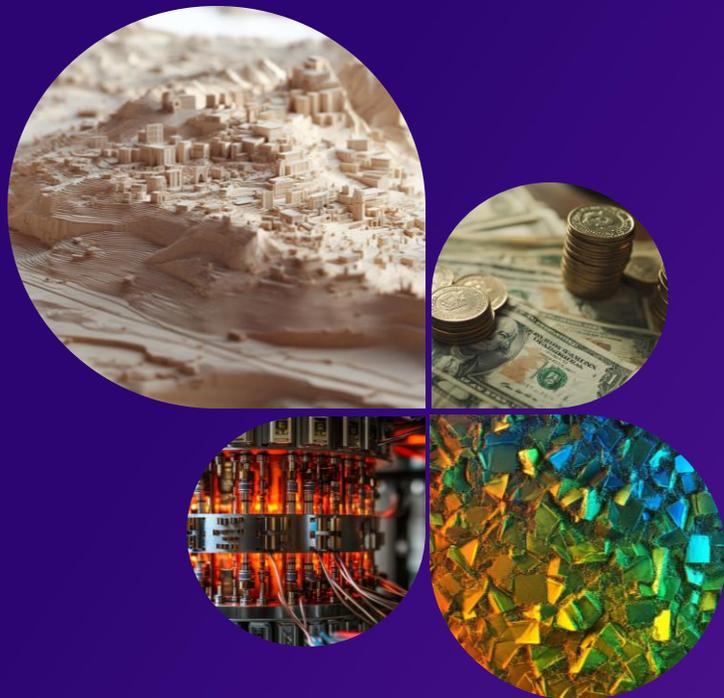
领域模型构建中，数据治理难题突出。领域数据分散，受隐私法规限制，传统采集与标注方式效率低。

## 数据分散与整合难题

在医疗、金融等垂直领域，**数据分散现象严重**。医疗数据分布于医院信息系统、影像存储系统及第三方平台；金融数据存在于银行核心系统、征信与监管机构数据库。**数据格式多样**，依赖不同技术架构与数据库。欧盟GDPR、我国**数据安全法与个人信息保护法**等法规，严格规范数据处理，加大跨机构共享难度。加之加密技术复杂，传统ETL工具与接口调用难以应对，严重阻碍数据整合进程。

## 数据标注效率与质量问题

领域模型训练高度依赖高质量标注数据，然而目前主流的人工标注方式弊端丛生。在自然语言处理的文本分类任务中，标注员逐字细读文本进行分类，如新闻分类，**效率极为低下**，熟练者每天最多标注几百条，大规模训练集标注常需数月。因标注员专业背景与理解能力不同，标注结果易出现分歧，**疲劳**等因素还会**引发人为错误**，降低数据准确性，最终导致模型在实际应用中**准确率、召回率**等性能难以达到高精度要求。



## 01 通用架构难以适配领域特征

通用的Transformer架构在自然语言处理和计算机视觉等通用领域成绩斐然，但面对特定领域数据却存在明显短板。像医疗病历、金融市场、法律文本、工业生产等领域数据，分别具有复杂因果逻辑、动态交织关系、专业术语和独特数据模式等特点。通用架构难以捕捉这些领域数据的特性，在语义理解、市场预测、文本分析和故障预测等方面表现欠佳，无法充分满足各领域的特定需求。

## 02 算力需求与成本问题

大模型因庞大参数量，在训练和推理时对算力要求极高。训练需海量矩阵与复杂神经网络计算。普通硬件资源，像个人电脑和小型服务器集群，因计算核心、内存及存储有限，无法满足需求。而购置专业计算设备，如GPU集群，价格高昂，且运行维护成本也高，包括电力、散热、维护等费用，这极大限制了大模型在各领域的广泛应用与发展。

01

## 过拟合与泛化能力受限

领域模型训练因数据局限易过拟合，泛化能力受限。简化模型虽能避免过拟合，但会牺牲准确性。此困境影响模型在医疗、金融等实际场景的应用效果与价值。

02

## 静态训练难以适应动态场景，资源消耗大

传统领域模型采用静态训练策略，难适应动态变化的业务场景，致使性能下降。虽在线学习可解此困，却面临高算力消耗，两者问题限制领域模型在动态场景的应用。

03

## 准确性与幻觉问题

领域模型存在准确性与幻觉问题。准确性不足在医疗、金融、智能客服等领域影响重大。幻觉问题源于训练数据质量欠佳、过拟合及对上下文理解能力缺乏，这些问题严重制约领域模型在实际场景中的应用。



## PART 02

# 技术赋能破局：DeepSeek R1 与 MoE/NAS 协同的领域大模 型优化策略与方案

# ▶ DeepSeek R1、MoE与NAS的协同赋能机制

三位一体架构，以DeepSeek R1为核，MoE与NAS协同破难题。

## 全流程协同机制

- 数据处理阶段
- 模型训练阶段
- 模型优化阶段

## 协同赋能

在政务热线，DeepSeek R1三位一体架构

1. 数据治理：DeepSeek R1多模态数据汇聚、MoE智能路由；
2. 模型适配：DeepSeek R1动态注意力+MoE分发
3. 架构优化：NAS搜索最优架构加速推理。

MoE

NAS

Transformer

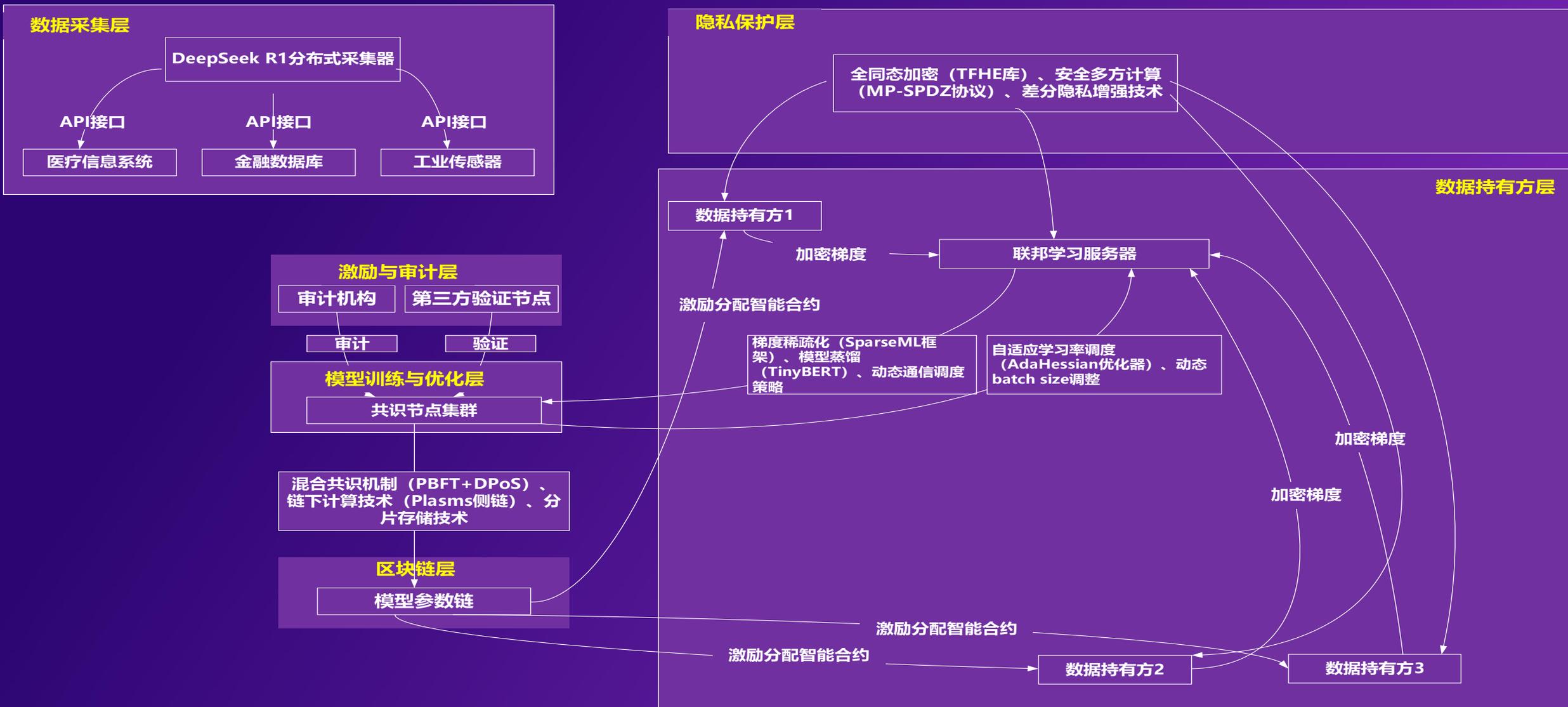
三位一体架构以DeepSeek R1为核，融合MoE与NAS技术，破解数据治理、模型适配及应用落地难题。数据治理实现多模态整合与智能路由；模型训练提升适配性；架构优化阶段提升性能降低成本，为领域模型落地提供创新方案。

# 数据分散与合规解决办法—数据融合与安全共享

依托DeepSeek R1分布式采集链，利用API接口实现多模态数据高效聚合，可无缝对接各类异构数据源并完成格式转换与安全传输。同时，融合区块链技术搭建安全共享平台，结合联邦学习实现数据本地训练、联合建模，辅以同态加密保障数据全流程加密交互，确保数据安全合规。

关键问题	解决方案
计算与通信开销	借助 <b>SparseML</b> 框架运用 <b>Top-k</b> 稀疏化算法筛选 <b>梯度</b> ； <b>BERT</b> 蒸馏， <b>知识迁移</b> 训练轻量版本；动态通信调度策略 <b>自适应调度</b> 数据传输频率和量，降低通信开销。
计算资源异构	利用 <b>Hessian</b> 矩阵 <b>计算</b> 学习率分配计算设备；通过 <b>batch size</b> 动态调整计算资源分配；轻量化计算下沉至 <b>边缘计算</b> 资源。
区块链性能瓶颈	融合 <b>PBFT</b> （ <b>拜占庭容错</b> ）强一致性快速确认与 <b>DPoS</b> （ <b>委托权益证明</b> ）选举验证节点降低共识复杂度；中间结果、数据预处理通过 <b>侧链处理</b> ；不同业务领域通过 <b>分片存储</b> 技术；综上突破区块链性能瓶颈
隐私与安全保护	基于 <b>TFHE</b> 全 <b>同态加密</b> ，在加密状态下完成 <b>联邦学习</b> ； <b>MP-SPD</b> <b>多方安全协议</b> ，在不曝露原始数据完成 <b>统计</b> ；辅以差分隐私技术，去除数据中 <b>干扰噪声</b> ，并通过添加 <b>可控噪声</b> 抵御逆向攻击，防止原始数据泄露。

# 数据分散与合规解决办法—数据融合与安全共享



# ▶▶ 标注提效：三策协同发力

半监督与主动学习降本提效，知识蒸馏赋能标注加速，GAN/VAE集成缓解稀缺，共促智能化标注与数据增强。

## 半监督学习与主动学习

通过结合少量标注样本和大量未标注样本，筛选低置信度样本由专家标注后优化模型，降低标注成本并提升模型对数据的利用效率。

## 知识蒸馏

借助将专家标注样本训练的教师模型知识迁移至学生模型，使轻量级模型快速准确标注新数据，提升标注效率与模型对数据的处理能力。

## GAN/VAE 网络

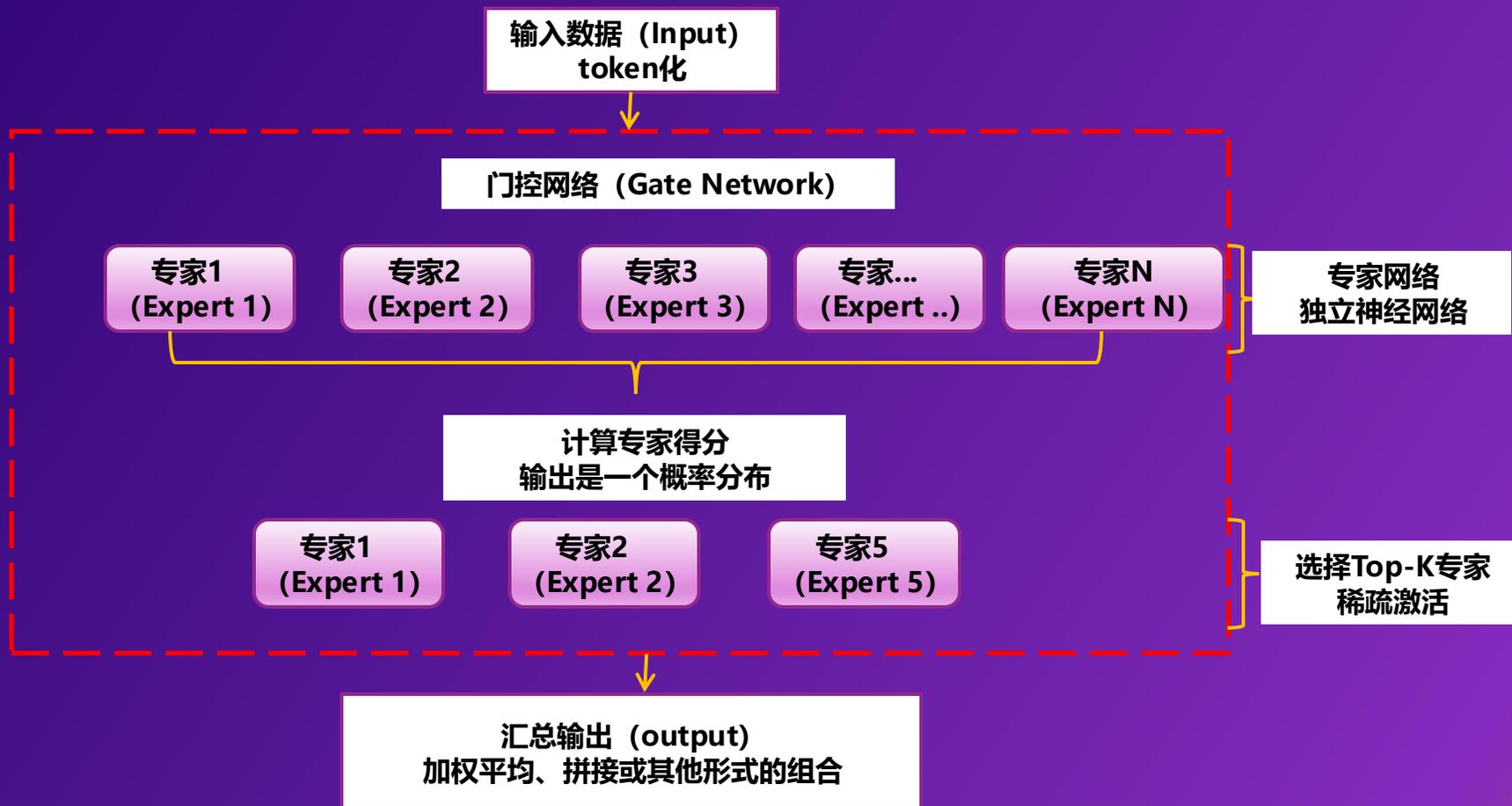
通过相互迭代生成多样特征样本，缓解数据稀缺问题，增强数据集规模与多样性，提升模型泛化能力和智能化标注准确性。

# MoE 架构：突破通用架构局限，领域适配方案

**1. 核心技术：** MoE架构以稀疏激活、动态路由、负载均衡为核心，从计算、匹配、调度三方面优化，应对领域数据复杂性。

**2. 架构流程：** 数据经token化后，门控网络完成专家选择，专家处理数据时负载均衡调节，最终汇总输出，展现特性实际应用。

**3. 技术赋能：** 结合MLA与MTP技术，强化数据关系捕捉与多token处理，提升MoE架构在特定领域的任务处理能力。



# 模型适配与算力优化—NAS破解硬件成本

大模型因参数量巨大，训练推理需海量计算，普通硬件难胜任，专业设备成本高昂。**NAS（神经架构搜索）**以自动化算法替代人工设计，探寻最优架构，**破解算力瓶颈**，显著降低人力与试错成本。

## 核心策略：

1. 自动搜索：自动组合、优化模型
2. 参数共享：ENAS权重复用减少资源消耗
3. 稀疏量化：剪枝降精度省算力耗时
4. 动态计算：依负载调节优化算力分配

指标	传统架构	NAS	幅度
运算量	4.1 GFLOPs	1.2 GFLOPs	-70.7%
推理延迟	120ms	45ms	-62.5%
训练成本	10,000	3,500	-65%
准确率	76.1%	75.8%	-0.3%

## 迭代优化流程：自动化架构进化



# 应用部署优化策略—多模态蒸馏在线学习抗过拟合提泛化

领域模型常因**数据局限陷入过拟合**，致使泛化能力不足。传统简化模型结构虽能缓解过拟合，却会削弱表达力，影响复杂任务精度。为此，提出应用部署优化策略，借助多模态知识蒸馏与在线学习技术，通过多阶段优化，突破**过拟合与泛化能力的矛盾**。



# ▶ 破局在线高耗能：领域模型动态适应与资源优化方案

1. 基于PyTorch Federated整合跨区政务数据，利用Apache Kafka与SPC捕捉数据动态，借助ILOD及AdamW+warmup加速模型适配，通过联邦学习与在线学习协同架构适配政务热线，以SVF联合量化剪枝加速推理、降低算力消耗。

3. 基于CLIP和3D CNN构建多模态特征提取器融合文本、语音、图像等多源数据，借助GNN实现多模态动态交互并运用自适应KL散度蒸馏策略优化模型训练，以适配政务热线多渠道诉求。

传统政务热线静态模型难适配动态诉求，在线学习算力消耗高。构建AI闭环优化体系，提升服务效率、控制成本，推动政务服务从“被动”转“主动”。

2. 通过强化学习挖掘Transformer奇异值生成专家向量调优模型，借助TensorRT量化与TorchSparse剪枝加速推理并压缩模型，助力政务热线高效运行。

4. 借助Kubernetes与寒武纪动态调配算力，通过Dask优化调度能力，可快速调度Kubernetes分配的寒武纪算力资源，实现政务热线场景下数据的并行处理与任务加速执行，有效提升服务响应速度与处理效率，保障政务热线服务稳定运行。

# ► 多维度协同治理：大模型幻觉问题破解方案

大模型幻觉源于**数据缺陷、过拟合及上下文理解不足**，影响智能服务可靠性。政务热线通过实施幻觉治理方案，在政策解读、工单分类等方面成效显著，幻觉率大幅下降，处理时效提升，市民信任度增强，推动智能客服向可信化升级。

## 多模态融合与交叉验证

提取融合文本语音图像特征，计算模态相似度，低阈值触发人工复核。

## 实时监测与闭环优化

图谱验证实体关系，人工反馈调优，动态更新模型应对变化保准确。



## 数据治理与知识增强

清洗多模态政务数据，剔除冗余错误；  
构建知识图谱，强化知识关联支撑。



## 模型优化与训练强化

增强动态注意力，对抗训练结合正则化，  
自适应蒸馏优化模型训练。



## PART 03

# 渐进优化： 领域大模型全流程优化框架

# 领域大模型全流程优化架构——四步十七法

“四步十七法”为核心，实现数据驱动、技术协同、持续进化及安全合规保障。

## 数据筑基（5法）

- 数据采集 → 跨域聚合
- 安全共享 → 隐私保护
- 标注优化 → 质量提升
- 知识迁移 → 成本降低
- 数据增强 → 特征多样

## 架构革新（3法）

- 架构搜索 → 自动搜索最优模型
- 效率成本平衡 → 体系调优
- 多维优化 → 普惠应用推进

## 群智重塑（4法）

- 门控分配 → 垂直领域训练
- 知识迁移 → 模型轻量化
- 动态激活 → 领域微调
- 动态路由 → 多模态融合决策

## 闭环进化（5法）

- 多维度监控 → 性能全览
- 自动诊断 → 问题溯源
- 策略优化 → 参数架构调优
- 测试验证 → 验证策略
- 持续进化 → 持续适配业务变化

# 数据筑基五法：安全合规的数据底座构建

围绕“数据筑基”，采用5法，借助数据存储治理技术，构建安全合规的数据底座支撑融合架构。



湖仓融合、智能缓存、存储压缩，跨模检索、治理升级降本

# 群智重塑四法：MoE驱动领域适配

基于DeepSeek R1的MoE与多模态融合技术，构建领域大模型从初始化到持续进化的完整技术闭环。通过动态知识缓存、路由策略优化、增量学习框架三大创新，显著提升多领域模型效能。

## 门控分配

门控网络借自注意力机制精准提特征，经余弦相似度匹配部署2048个专家模块，对比学习构建知识图谱实现垂直预训练。

## 稀疏激活

输入数据经哈希映射至专家模块索引空间，LoRA冻结大量参数缩短模型微调耗时，知识缓存区实现专家模块和知识图谱自动更新完成微调。



## 知识迁移

依领域数据选预训练模型，构建蒸馏数据集，借自适应温控与注意力对齐，实现领域模型压缩与性能优化。

## 动态路由

基于强化学习的Q-Learning、蒙特卡洛树，在多模态数据处理中选取最优路径。借GraphSAGE跨模态特征融合；门控机制对专家模块结果进行加权整合。

# 架构革新三法：NAS架构探索 (1/3)

NAS架构搜索通过自动化探索混合搜索空间，突破大模型千亿级参数的算力瓶颈，形成“动态适配-参数压缩-模式创新-持续进化”的完整优化链路：

## 动态适配

基于强化学习，分析数据特征，构建100+维度复杂度矩阵，通过“快速验证-精度验证-硬件适配”三级验证机制，自动生成混合架构。

01

## 参数压缩

融合量化与剪枝技术，采用压缩损失函数动态平衡精度与效率，关键层采用FP16混合精度、非关键层INT8量化，并通过泰勒展开排序实现神经元智能剪枝。

02

## 模式创新

引入局部卷积、因果注意力等扩展搜索空间，针对时序数据构建模型，创新组合降复杂度，实现工业物联网实时决策。

03

## 持续进化

构建在线搜索与硬件协同闭环，动态更新、蒸馏迁移知识，开发专用单元与库，提升模型资源利用效率。

04

# 架构革新三法：效率与成本平衡 (2/3)

依托万卡集群并行计算能力，DeepSeek R1与NAS技术深度协同，形成三重核心优化策略，实现大模型训练的成本与效率双突破。

解决痛点	解决方案	技术创新点	解决问题	实测效果
GPU负载不均衡 资源利用率不均	动态负载均衡	基于梯度稀疏性的微批次调度算法	基于梯度稀疏性微批次调度算法，结合NAS轻量化架构提升任务分片精度，实时监测调整任务分配，使GPU利用率稳定超85%，训练周期缩短40%，解决资源分配不均问题。	利用率提升至85%+
内存不足	混合精度训练	FP16/FP32 自适应精度控制	采用FP16/FP32自适应策略，关键层FP32保收敛、非关键层FP16降显存，结合NAS量化压缩，千亿模型内存需求从1.2TB降至680GB，存储成本减少43%。	内存占用减少43%
稳定性与精度停滞	学习率智能调度	warmup+余弦退火联合策略	集成warmup+余弦退火策略，结合NAS搜索的最优架构动态调学习率，模型训练300 epoch后仍保持0.8%周精度提升，降低重复训练成本。	后期精度提升0.8%/周

# 架构革新三法：多维优化 (3/3)

NAS与DeepSeek R1的协同优化实现“**架构降本+工程增效**”双重突破，通过神经架构搜索、模型量化、半监督学习等技术术语的深度应用，为大模型在金融、医疗等领域的普惠化应用铺就可行路径，推动技术落地与场景应用的深度融合。

序号	优化方面	工具	效果
1	硬件优化	NAS (神经架构搜索)、模型压缩工、边缘计算平台	边缘设备部署成本下降90%；智能客服系统单节点推理成本从0.05元/次降至0.005元/次，实现云端到边缘的低成本迁移。
2	能耗降低	模型量化工具、能耗监测工具	模型单次推理能耗从120J降至15J，日均百万次推理年节省电费超200万元，达成能耗与性能的精准平衡。
3	效率提升	半监督学习框架 (PyTorch Lightning+自定义半监督库)、标注工具	标注效率提升4倍，单样例标注成本从8元降至2元，通过主动学习策略显著减少人工标注工作量与周期。

# ▶ 闭环进化五法：智能反馈迭代机制

为推动模型持续进化，构建“评估 - 诊断 - 优化”智能闭环体系，通过五大核心模块协同运作，实现模型性能的动态提升与领域适应性增强



1. 多维度评估矩阵：开启监控与知识交互
2. 自动诊断引擎：解析异常，联动专业支持
3. 策略引擎：生成方案，应对优化挑战
4. A/B 测试验证：验证效果，闭环循环
5. 持续进化：持续迭代，增强领域适配性与泛化能力。

# PART 04

## 实践验证： 政务领域大模型应用成效

# ▶ 聚焦群众、企业和政府痛点、难点，提升服务满意度

## 诉求办理渠道分散

存在多个民生诉求办理平台，群众在多平台反映问题后需要在多个渠道查询办理情况，部门和基层需要在多个渠道反馈办理结果。

## 反映诉求不够便利

反映诉求等待时间长；群众填写诉求常常遇到问题，比如手工点选问题发生时，不清楚问题发生地的群众无法准确填写，增加流转办理时间等。

## 业务 痛点

## 协同治理能力不强

有的问题涉及多个责任部门，经常出现责任主体不明、工作边界不清的情况，导致相关问题在部门间反复转办，相互推诿扯皮，迟迟得不到解决。

## 分析预警能力不强

对民情民意数据的加工整理、分析研判、风险预警方面还有较大差距，通过大数据分析赋能社会治理方面的能力还需提升。

聚焦群众、企业和政府痛点、堵点、难点，提升获得感

# 政务便民热线智能化升级路径探索

政策引领+技术赋能+深挖重点场景

热线服务端到端全流程全场景升级

一站式使能热线数智化建设



应用5G新通信

线上线下多渠道丰富与协同



大模型+小模型

实现AI深度应用与广泛赋能



大数据+AI

提升数据治理能力



# ▶ 数据问题破解—构建高质量语料库，加速应用场景落地

基于中国联通数据中台和数据合作生态，沉淀**大规模**数据资源  
 构建规模央企领先的语料库，强化元景大模型**多模态**数据能力  
 通过行业最高数据管理能力认证，**高质量**创新数据质量管理体系

依托集约数据中台  
 和良好数据服务生态

一站式智能化数据集  
 生产流水线

通过DCMM 5级认证  
 数据管理能力进入国家第一梯队

**大规模**

结构化数据 600PB+  
 非结构化数据 5PB+  
 客服、网络、法律等多知识中心  
 18个军团深度融入垂直行业

**多模态**

日均汇聚多模态数据 480TB  
 数据自动化处理与增强能力  
 多模态数据智能化标注能力

**高质量**

打造完备数据标准体系  
 发布共享数据集目录  
 创新场景化数据质量管理体系

# 模型问题破解—元景MaaS-DeepSeek产品构建

中国联通元景MaaS-DeepSeek平台，是一款支持国产化硬件的自主创新大模型产品，其集成了元景大模型、DeepSeek系列大模型、具备RAG、智能体，是国内领先的引入DeepSeek大模型的创新人工智能产品。



# 行业模型问题破解—联通元景构建政务行业大模型

- 源自多个等省级数字政府实践和数据，以及联通自有数据等在联通基座大模型上进行训练构建政务行业大模型
- 基于超大模型训练平台，提供提示词、插件以及模型微调等大模型扩展工具，为快速开发智能化政务应用场景提供支持
- 以多个省级数字政府实践，基于上述自研技术体系，打造公共服务、社会管理、政务运行等智能化场景应用



# 场景落地——打造六类政务热线智能体

**热线“智服”智能体**  
咨询诉求自动应答

**诉求“智报”智能体**  
诉求风险智能上报

**民意“智办”智能体**  
接诉即办未诉先办

**职场“智管”智能体**  
质检排班监控培训

**关切“智访”智能体**  
民众关切主动回应

**社情“智判”智能体**  
智能报告智能问数

一号接入、多端协同

人机协同、高效辅助

决策支撑、业务联动

分析预警、问题洞察



? 来电人是谁? (启动来画像) -->以前问过那些问题, 问题是否解决?

 来电人诉求尚未满足, 预估是来咨询处置进度, 建议直接转交熟悉事件的人工坐席。

? 根据画像, 来电人可能咨询办件进度, 推荐给客服响应话术。

 同步转写市民诉求, 发现重要信息, 生成风险研判报告, 将事件转交给对应部门。

? 提醒处置人员情况变化, 通过趋势研判发现事件严重程度升级。

 为分析人员分析事件热点, 发现近期多类事件爆发, 需要特别关注。

MAAS平台:

语料加工

选模型

改模型

用模型

基础模型:

元景语言大模型 (7B、13B、34B、72B)

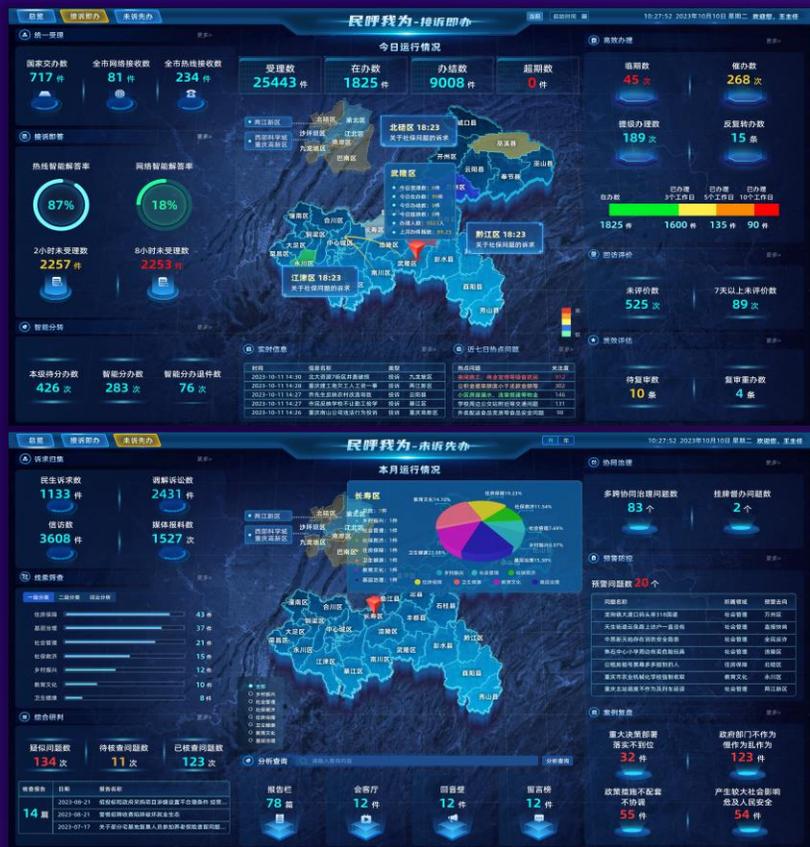
元景语音大模型

元景多模态大模型

DeepSeek-V3/R1

# 场景落地——“民呼我为”打造“民生诉求办理”一件事

民呼我为重大应用在重庆市“互联网+督查”平台、“12345”热线的基础上迭代升级，打造民呼端、办理端、治理端和决策端，建成民生诉求“一个库”、诉求办理“一张网”、问题线索“一本账”、指挥调度“一屏览”，旨在搭建政民互通“连心桥”、打造民生诉求办理新范式，形成民办事长效机制，有效改善人民群众民生诉求服务体验，切实减轻基层单位办理负担，提升党政部门治理能力。



## 项目亮点:

整合“12345”热线和“互联网+督查”平台，聚焦民生诉求办理的“接诉即办”“未诉先办”场景，实现“民有所呼，我有所为”，承接领导驾驶舱、数据治理和智能分析等工作，打造民生诉求办理新范式。

## 建设成效:

实现了国家、市、区县、镇街、村社、网格六级贯通。运用“治未病”思维，把个性问题和共性问题、局部问题和面上问题结合起来，提前发现、预防、化解一批卡点难点问题，推动民生诉求从“事后管理、被动反应”向“事前治理、主动回应”转变。

## PART 05

# 技术融合与展望： 领域大模型优化策略的总结与未来探索

**DeepSeek R1 - MoE - NAS** 融合体系通过**技术创新**，系统性解决领域大模型在数据、模型及应用方面的难题，为行业智能化转型提供了切实可行的高效路径，在政务热线、医疗、金融等多领域展现出显著的应用价值与优化成效。提出了一套**解决方案**和一个**框架**。

## DeepSeek R1、MoE与NAS协同解决方案

1. DeepSeek R1、MoE、NAS**协同赋能**
2. DeepSeek R1等多技术**促数据融合与安全共享**
3. MoE**领域模型适配**，NAS**优化算力瓶颈突破**
4. 行业模型**过拟合、高耗能、幻觉应用痛点**解题

## 全流程框架

1. **数据筑基**：借DeepSeek R1构建安全优质数据底座
2. **群智重塑**：MoE助力领域适配，加速模型构建
3. **架构革新**：DeepSeek R1、MoE、NAS协同降本
4. **闭环进化**：闭环机制提升模型适应性与性能

以DeepSeek R1为核心的“MoE+NAS”等多技术协同促进行业领域模型，加速各产业智能化转型。



## 智图强推

知识图谱与MoE相结合，增强模型对复杂逻辑的推理能力，助力处理复杂场景下的问题。



## 实参优调

借助强化学习与在线学习技术，实现对模型参数的实时优化，使模型能够快速适应动态变化的环境。



## 轻模精构

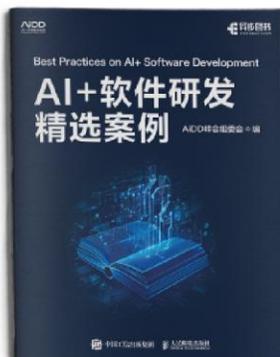
通过神经架构搜索（NAS）搭配知识蒸馏技术，生成小于10亿参数的边缘模型。



## 跨模同构

跨模态协同工作，构建统一的特征空间，打破不同模态数据之间的壁垒，促进多模态信息的融合与交互。

## 参与调研您将优先获得



AiDD定制版  
《AI+软件研发精选案例》



专属学习顾问  
1对1需求对接

# AiDD会后小调研

AiDD峰会致力于协助企业利用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和数字化的新时代。作为峰会的重要共建者，您的真知灼见对我们至关重要。衷心感谢您的参与支持！

# 2025 AI+研发数字峰会

## 拥抱 AI 重塑研发



扫码参与调研

# 科技生态圈峰会 + 深度研习

—1000+ 技术团队的选择



**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: 2025.09.26-27

**K+峰会** **香港站**  
**K+ 思考周®研习社**  
时间: 2025.11.17-18



K+峰会详情



**AiDD峰会** **上海站**  
**AI+研发数字峰会**  
时间: 2025.05.23-24

**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: 2025.11.14-15



AiDD峰会详情



# 2025 AI+研发数字峰会

AI+ Development Digital Summit

# 感谢聆听!

扫码领取会议PPT资料

