

大模型赋能的企业智能化运维：落地难点与解决方案

朱凌梅

中国国家博物馆，北京，100016

摘要：随着数字化转型的深入和扩展，传统 IT 运维模式面临严峻挑战。大模型技术凭借其超强的学习、识别和生成能力为企业智能化运维提供了新的解决方案。然而，大模型在运维领域的落地存在人才稀缺、数据质量与治理难题、业务场景拆解复杂性及成本收益平衡等核心挑战。本文系统分析了这些难点，并提出针对性的解决方案。通过高价值场景试点、渐进式推进和持续迭代优化，企业可逐步实现运维智能化转型，最终提升 IT 运维效能与系统稳定性。

关键词：企业智能化；数据复制；智能化运维

作者简介：朱凌梅，（1982-），女，汉族，北京，硕士研究生，中级工程师，研究方向：云计算、数据复制和数据迁移。

一、背景与概述

在数字化转型的浪潮下，企业对 IT 运维的要求日益提高，传统运维模式依赖人工监控和规则脚本，局限性逐渐凸显。运维效率低下、数据爆炸导致分析滞后、故障被动响应、知识管理困难等问题严重制约了企业的发展^{[1][2]}。大模型技术的兴起为智能化运维提供了新的解决路径，其强大的自然语言处理、数据分析和预测能力，帮助企业实现故障主动预防、自动化任务执行和知识系统化沉淀^{[3][4]}。然而，大模型在运维领域的落地仍面临诸多挑战。

本文从落地难点、团队能力要求、核心问题解决、工具整合策略四个维度，系统梳理了大模型赋能智能化运维的实践路径，旨在为企业可提供可参考的实施方案。

二、企业落地大模型智能运维的核心难点

企业落地大模型智能运维的核心难点分布在人才、技术和成本三个方面。其中技术方面的困难来源于训练数据和业务场景拆解两个方向。

1. 新型运维人才稀缺

传统运维团队 AI 知识储备薄弱。统运维人员多为 Linux 操作系统、网络或硬件领域专家，对人工智能、机器学习及大模型的能力边界、应用方式等理解有限，缺乏推动智能化落地的基础。

技术领域分歧导致复合型才稀缺。大模型和机器学习作为新兴技术领域，其知识体系与传统运维的实践需求存在显著鸿沟。运维领域核心技能集中于

Linux 等操作系统、网络协议及自动化工具等，其知识体系与掌握 AI 技术、大模型训练所需的数据科学、深度学习等知识和能力几乎无交集。这种分野导致两类人才的教育背景和职业发展路径截然不同，使得企业难以通过常规招聘或内部培训填补人才缺口。

2. 运维数据质量与治理难题

数据是大模型训练的核心要素，其规模和质量直接决定了模型的性能上限。只有当数据覆盖足够广泛的运维场景时，模型才能建立准确的表征能力，避免在特定任务上出现系统性偏差^[5]。数据的标注质量同样关键，噪声数据会导致模型学习到错误的关联关系，提升模型可靠性必须依赖经过严格校验的数据。大模型的性能提升，很大程度上依赖于数据规模的扩展和数据清洗技术的改进。当下，拟用于模型训练的运维数据主要面临数据质量和数据样本不足两个问题。

数据孤岛与格式多样化

运维历史数据分散于各类操作系统和软件应用的日志、监控、告警平台、运维记录等不同位置，存在数据孤岛问题。此外，其记录格式、记录风格、存储方式、保存时限都不统一，数据源中噪声多，标准各异。这些现象影响了数据的可用性与模型训练的效率^[6]。

运维数据样本不足

目前多数运维操作仍以命令式、脚本化为主，历史故障解决方案和日志数据缺乏系统的归纳总结，导致可供建模的数据样本不足，导致大模型训练缺乏高质量语料，难以支撑有效的大模型训练。

3. 业务场景拆解与落地复杂性

运维面临的环境复杂多变，问题定位过程复杂多变，甚至需要跨设备定位问题。故障引发因素不同，修复的解决方案也有所不同。如何将运维场景拆解为大模型可理解的子任务，需要技术人员对业务场景和大模型功能都有深度理解。不同运维任务对模型能力的需求差异较大，如何进行模型选择与部署，确保资源利用率和效果之间的平衡，是一项挑战。

同时，企业 IT 系统存在自研与采购并存、IaaS 标准不一等问题，使数据对接与整合难度上升，形成从数据源到模型的多重障碍。

4. 成本与收益平衡困境

多数企业对数据安全要求严格，禁止将数据传出企业网络，导致公有云大模型难以应用，私有化部署又面临算力与成本压力，如大模型训练所需 GPU 等硬件投入、数据清洗和治理费用，中小企业难以承担^[7]。此外，对于模型训练所需数据集，仅靠中小企业自身的运维数据，很难支撑其模型训练，很可能还需要从第三方数据平台购买数据或购买预训练模型。

但是，智能化运维的收益和效果难以直观衡量。运维优化的产出，如系统稳定性提升、MTTR 降低等，需通过下游业务指标间接体现，且受多重变量干扰，导致归因分析困难。此外，对提升运维质量和效率的投入类似质量成本，是一种“预防成本”，其收益隐性化，并且需要从长期周期回报中进行测量，存在显著的隐形价值和滞后性效应，其回报往往难以通过传统财务指标直接量化。因此，企业缺乏对其进行投入的动力^[8]。

三、大模型智能化运维的关键问题与解决方案

1. 数据收集与清洗

全面、准确、多样、平衡、干净的数据可大幅提升大模型解析与推理能力。

首先，对于数据收集，应借助专业监控系统、埋点技术，实现多维度指标的全面采集。注意数据格式统一，需要制定好元数据标准，如采用 OpenTelemetry 等标准协议，实现不同系统间的数据兼容与统一输入；或将日志、指标、告警数据转换为标准 Schema（如 JSON），便于模型训练。之后需做好数据清洗和数据治理，结合 NLP 技术，去除冗余与

无效信息，如过滤控制字符、降低低价值日志比重。

2. 设计好自动化运维任务的执行机制

做好 Agent 机制。将复杂任务分解并构建对应的 Agent，进而构建可执行的函数调用，将大模型与自动化脚本通过 Agent 函数调用方式打通。依据历史日志与预设意图，实现特定场景下的自动触发与快速响应。当日志中发现一些之前设定好的场景或者意图即可触发对应的修复 Agent 函数调用完成任务，实现从监控分析到任务执行的全链路闭环。

与此同时，应做好人机协同工作，大模型在提出每一个修复措施前应由人工审核并决策。对于执行结果，应引入自动化实时监控机制和人工监控机制，对任务效果进行评估和反馈，优化策略和调用链路。

3. 增强模型可解释性，缓解“黑盒”焦虑

通过优化提示词设计，将过程透明化。可通过 prompt 引导模型输出思考路径、引用依据，并列出多种解决方案供人工对比参考。为了兼顾性能和可解释性，可以将简单结构模型和深度学习黑盒模型结合起来形成混合可解释模型。通过这种方式在保持模型高性能的同时，获得模型的可解释性，解释模型内部的决策过程和预测结果^[9]。另外，应做好决策机制内嵌，将运维红线、决策审批逻辑写入模型提示中，降低模型“幻觉”引发误判的风险。最后，可考虑多方案生成，供人工决策。

4. 运维团队能力体系构建

要真正发挥大模型的价值，团队需构建多维度的能力体系，除熟悉运维领域各类知识技术等，还应熟练运用数据科学知识、大模型技术，具备工具链及业务场景落地能力。

1) 熟悉数据科学、数据管理体系

运维团队需建立完善的数据科学知识体系，能够识别出有价值的运维数据，并且具备日志采集、清洗、标注能力，确保训练数据的高质量，以支撑大模型的输入与训练。

2) 具备运维领域知识与大模型结合能力

运维团队需能够将故障排查、性能优化等运维场景拆解、转化为模型可理解的任务、抽象为可调用的函数，并能够验证其输出的准确性与合理性。

此外，运维团队应该具备编程能力和大模型基础

认知，熟悉主流大模型的开源 / 商用规则，掌握微调、编写 Prompt 工程等技术，并能通过 API 或本地化部署（如 vLLM）快速验证。

3) 掌握常用技术，具备工程化能力

首先，在运维工具方面，要能够熟练使用如 Prometheus、ELK、Zabbix 等监控与日志分析工具；在 AI 工具栈方面，要能掌握 DeepSpeed、Hugging Face 等微调框架，以及 Docker、CUDA 等部署与排错技能。在掌握工具的基础上，还需要深入分析业务需求，根据需求和大模型的特点选择适配大模型，并具备本地化部署与优化能力；并且，通过实验调优模型输出，结合真实运维反馈不断改进。

4) 注重团队能力培养

应注重团队整体能力培养，做好明确分工之和确保各类人员的技术专长外，还应关注团队成员知识体系的完整性，从而减少沟通成本，提高团队协作能力。可通过系统培训、组织技术分享、增加业务专长以外的实践机会等方式来提升团队整体的 AI 技术能力。

实施路径

考虑到成本投入与收益的平衡，可考虑按照以下路径进行实施：

首先，企业需明确目标与场景，优先选择高价值、低风险的领域（如日志分析）作为试点，以验证技术可行性并控制初期风险。其次，应采取分阶段推进策略，从基础的数据治理入手，逐步过渡到模型验证，最终实现全流程自动化，避免盲目追求一步到位而导致资源浪费。在人才方面，需通过外部合作与内部培训相结合的方式，构建既懂传统运维又掌握 AI 技术的复合型团队，弥补跨领域人才缺口。最后，建立持续迭代机制，根据实际业务反馈不断优化模型与流程，例如利用检索增强生成（RAG）技术动态更新知识库，确保系统长期适应运维需求的变化。这一渐进式路径可帮助企业平衡技术投入与实效，稳步实现运维智能化转型。

同时，由于利用大模型来增强运维能力投入属于预防性成本，难以直接衡量收益，影响持续投入。可考虑参照质量管理工作的收益衡量方法。如：通过建立预防性投入的基线确保可持续的建设投入；用风险热力图展示已规避或者快速解决掉的问题，展示建设

成果；通过长期追踪和数据统计，将运维稳定性提升对资本成本的影响纳入计算。

四、结语

大模型智能化运维是提升企业 IT 能力的重要手段，但其落地也面临着技术融合、训练数据收集、人才储备建设、成本和收益规划等多方面困难，企业应做好数据管理、数据治理、技术储备、技术团队建设，并提升成本控制等方面的能力。考虑到落地可行性和增强管理层及投资者信心，在实施路径方面，企业应理性评估自身条件，以“小步快跑”的方式实现从传统运维到智能运维的转型升级。

参考文献

- [1] 郭东旭. 数据中心智能运维系统的研究 [J]. 电子技术与软件工程, 2023(6): 255-258.
- [2] 潘哲. 数据中心运维管理现状及策略探讨 [J]. 电子技术与软件工程, 2014(04): 210.
- [3] 欧阳金福. 人工智能视角下数据中心的网络运维与安全防护策略 [J]. 中国宽带, 2025, (06): 46-48.
- [4] 曾小明, 李伟, 王江颖, 等. 数据中心智能运维技术探索与应用 [J]. 数字通信世界, 2025(03): 97-99.
- [5] 张欣. 生成式人工智能的数据风险与治理路径 [J]. 法律科学 (西北政法大学学报), 2023, 41(05): 42-54.
- [6] 赵朝阳, 朱贵波, 王金桥. ChatGPT 给语言大模型带来的启示和多模态大模型新的发展思路 [J]. 数据分析与知识发现, 2023, 7(3): 26-35.
- [7] 潘哲. 数据中心运维管理现状及策略探讨 [J]. 电子技术与软件工程, 2014(04): 210.
- [8] 罗兰贝格管理咨询公司. 预测性维护：工业数字化领域潜在爆发点 [J]. 中国工业评论, 2017, (11): 72-78.
- [9] 廖勇, 韩小金, 刘金林, 等. 可解释性人工智能研究进展 [J/OL]. 计算机工程. (2024-10-30)