

# 基于大语言模型的微电网负荷最优分配方案的研究综述

毛志忠<sup>†</sup>, 刘文程

(东北大学 信息科学与工程学院, 沈阳 110819)

**摘要:** 微电网负荷优化分配对于提高分布式新能源消纳水平和系统运行经济性具有重要意义, 但受可再生能源出力波动、负荷随机性及源荷储耦合特性的影响, 相关优化问题具有较强的复杂性与不确定性. 大语言模型 (LLM) 在语义理解、知识提取和推理生成等方面展现出较强能力, 为微电网负荷优化分配研究提供了新的方法支撑. 针对 LLM 在微电网负荷优化分配中的应用研究, 首先介绍 LLM 的基本架构、训练范式及微电网负荷分配研究基础; 然后结合微电网能量管理系统的典型功能划分, 从电源调度、负荷预测、储能管理和需求响应 4 个方面, 梳理 LLM 的应用方式、研究进展及适用特点; 在此基础上, 分析 LLM 在微电网场景中面临的解释性、多模态融合、实时性与隐私安全等关键问题, 并总结相应改进思路; 最后对能源专用大模型构建、物理机理与数据驱动融合、标准化评测以及拓展应用等研究方向进行展望.

**关键词:** 微电网; 负荷优化分配; 人工智能; 大语言模型; 能源管理系统; 优化调度

**中图分类号:** TP391.8 **文献标志码:** A

**DOI:** 10.13195/j.kzyj.2026.0001

**引用格式:** 毛志忠, 刘文程. 基于大语言模型的微电网负荷最优分配方案的研究综述 [J]. 控制与决策, xxxx, x(x): xxxx-xxxx.

## Research review on optimal allocation schemes for microgrid load based on large language models

MAO Zhi-zhong<sup>†</sup>, LIU Wen-cheng

(College of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

**Abstract:** Microgrid load optimal allocation is of great significance for improving the consumption level of distributed renewable energy and enhancing system operating economy. However, due to the fluctuations in renewable energy output, load uncertainty, and the coupling characteristics among source, load, and storage, the related optimization problems are highly complex and uncertain. Large language models (LLM) have demonstrated strong capabilities in semantic understanding, knowledge extraction, and reasoning generation, thus providing new methodological support for research on microgrid load optimal allocation. Focusing on the application of LLM in microgrid load optimal allocation, this paper first introduces the basic architecture and training paradigms of LLM, as well as the research foundation of microgrid load allocation. Then, in combination with the typical functional division of microgrid energy management systems, it reviews the application approaches, research progress, and applicable characteristics of LLM in four aspects: power source dispatch, load forecasting, energy storage management, and demand response. On this basis, it further analyzes the key issues faced by LLM in microgrid scenarios, including interpretability, multimodal fusion, real-time performance, privacy, and security, and summarizes corresponding improvement strategies. Finally, future research directions are discussed, including the development of domain-specific large models for energy systems, the integration of physical mechanisms with data-driven methods, standardized evaluation, and the expansion of application scenarios.

**Keywords:** microgrid; optimal load distribution; artificial intelligence; large language model; energy management system; optimal dispatch

## 0 引言

在全球能源转型和“双碳”目标驱动下, 微电网 (Microgrid) 因其在分布式新能源消纳和局域负荷分

配中的优势, 成为电力与能源领域的研究热点<sup>[1]</sup>. 微电网通常由本地分布式电源、储能装置和多样化负荷构成, 可以在并网或离网模式下运行, 实现局部区

域的能量自给和优化<sup>[2]</sup>.然而,微电网的运行优化也面临诸多挑战.大规模可再生能源接入带来的出力随机性以及负荷侧用电行为的多样性,增加了供需平衡和调度决策的难度<sup>[3]</sup>.微电网内部多种能源之间的相互作用和动态交互,使得采用精确物理模型进行实时优化求解变得困难<sup>[4]</sup>.现有的微电网负荷优化分配主要依赖于模型驱动的方法,但在复杂不确定环境下往往难以兼顾优化性与鲁棒性<sup>[5]</sup>.近年来,数据驱动的方法开始应用于负荷预测和能量管理,取得了一定效果,却仍存在泛化性不足和可解释性差等问题<sup>[6]</sup>.

人工智能领域兴起的大语言模型 (large language model, LLM) 为复杂系统的优化提供了新的研究思路<sup>[7]</sup>.大语言模型基于 Transformer 架构,并通过在海量文本数据上的预训练,展现出较强的上下文理解与推理能力<sup>[8]</sup>.自 2022 年以来,生成式预训练 Transformer (generative pre-trained transformer, GPT)、路径规模化语言模型 (pathways language model, PaLM) 等一系列超大规模模型相继出现,在多种任务中表现出接近甚至超过人类的能力<sup>[9]</sup>.这类模型不仅在自然语言处理方面具有优势,还表现出一定的数学推导和知识迁移能力,这为其在电力系统数据密集、结构复杂的领域中的应用提供了新的契机<sup>[10]</sup>.

将大语言模型引入微电网负荷优化与分配的研究近两年逐渐兴起<sup>[11]</sup>.LLM 能够解析和利用微电网运行过程中产生的多源异构数据,包括历史负荷记录、气象数据、调度日志等,从而为负荷预测和调度决策提供更全面的知识支撑<sup>[12]</sup>.LLM 具备人机对话与解释生成的能力,可作为智能助手帮助调度员分析复杂场景、制定优化方案,并以自然语言形式提供原因与建议<sup>[13]</sup>.美国国家可再生能源实验室 (NREL) 提出的 eGridGPT 愿景,设想利用 LLM 在调控室中协助操作员读取监控与数据采集系统 (supervisory control and data acquisition, SCADA) 实时数据,并通过问答式交互分析故障情景、推荐调度措施<sup>[14]</sup>.Long 等<sup>[15]</sup>也总结了 LLM 在电力系统运行优化中的应用潜力,认为其在发电调度、故障诊断和决策支持等方面均具有前景.尽管如此,目前 LLM 在微电网方向的研究仍处于起步阶段,相关成果多集中于初步探索与概念验证,如用于短期负荷预测、辅助能量管理决策等<sup>[16]</sup>.但针对该领域的系统性综述和前瞻性展望仍相对有限,这在一定程度上影响了后续研究的整体推进.

基于上述背景,本文聚焦于大语言模型与微电

网负荷优化分配的结合,对相关研究进行系统性综述.首先,阐述大语言模型的核心原理,包括 Transformer 架构及其主要训练范式;其次,回顾微电网负荷优化分配的研究进展,为引入 LLM 的应用奠定背景;系统梳理 LLM 在电源调度、负荷预测、储能管理与需求响应中的典型应用,并比较不同方法的性能特点与适用场景;随后,讨论 LLM 在微电网落地过程中面临的主要挑战,包括模型可解释性、多模态信息融合、实时性保障以及隐私安全问题,同时简要介绍联邦学习和数字孪生等潜在解决途径;最后,展望未来研究方向,重点关注电力系统专用大模型的构建、物理约束嵌入推理过程的可行路径与标准化评测体系的建立.

## 1 大语言模型基础

### 1.1 模型架构与原理

大语言模型通常采用 Transformer 架构,即以自注意力为核心的深层神经网络<sup>[17]</sup>.经典 Transformer 以编码器-解码器范式提出,在长序列关系建模方面表现突出,并已成为自然语言处理与多模态学习的重要基础<sup>[18]</sup>.其核心由多头自注意力与前馈网络交替堆叠构成,依托并行计算缓解循环神经网络 (recurrent neural network, RNN) 的长程依赖与串行瓶颈<sup>[19]</sup>.图 1 为标准的编码器与解码器构成<sup>[20]</sup>,左为编码器层,右为解码器层.其中,编码器由多层自注意力与前馈子层叠加,每层含残差连接与层归一化;解码器在此基础上引入掩蔽自注意力以保证因果性,并增加编码器-解码器注意力以融合上游信息<sup>[21]</sup>.该架构在捕捉长程关联与并行训练方面较 RNN 与长

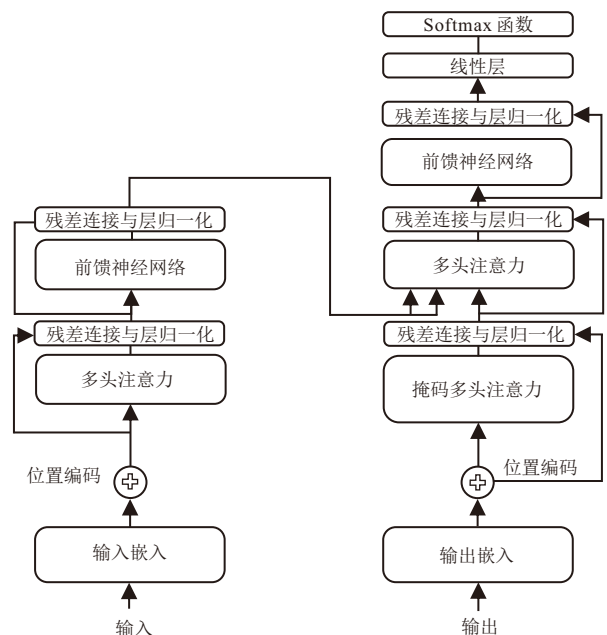


图1 Transformer 编码器-解码器架构

短期记忆网络 (long short-term memory, LSTM) 具有明显优势, 从而在效率与效果上更具竞争力<sup>[22]</sup>. 自2017年该范式提出后, Transformer 迅速成为大规模语言模型的主干, BERT (bidirectional encoder representations from transformers) 与 GPT 系列均以此为基础<sup>[23]</sup>.

表1对比了传统循环神经网络 RNN、LSTM 与 Transformer 大模型架构的关键特性差异. 可以看出, Transformer 架构在建模长程依赖关系和并行计算

等方面表现优异, 具有更好的训练稳定性和可扩展性, 因此成为大语言模型的主流架构基础. 同时需要指出, Transformer 在长序列场景下的计算与内存开销增长较快, 工程实践中通常借助高效注意力与推理优化以满足应用需求. 相较之下, RNN、LSTM 在实时数据处理与快速响应方面仍具有一定优势, 但在超长序列建模与大规模并行训练方面存在先天限制.

表1 模型架构对比

维度	RNN/LSTM	Transformer (LLM)
长程依赖建模 <sup>[24]</sup>	对长距离依赖更为敏感, 信息随距离增大易衰减	自注意力可直接建立任意位置间的关联, 长距离依赖刻画更充分
训练、推理并行性 <sup>[20]</sup>	时间步递推, 串行特性明显, 并行度受限	无循环结构, 矩阵运算并行度高, 训练与推理吞吐更好
训练稳定性 <sup>[19]</sup>	易出现梯度消失、爆炸, 稳定性依赖特殊设计与超参调节	残差连接与归一化等机制有助于稳定收敛
序列长度与资源代价 <sup>[17]</sup>	计算、存储随序列长度增加较为平稳	长序列下计算与内存压力上升, 需配合工程化优化与高效注意力
多模态与工具生态 <sup>[25]</sup>	对多模态与外部工具集成支持有限	多模态与工具链生态成熟, 便于与检索、求解器、仿真器等协同
模型扩展性 <sup>[26]</sup>	在大规模预训练与跨任务迁移中受限	便于随参数与数据规模扩展, 已成为大模型主流架构

LLM 通常在 Transformer 上通过超大规模预训练获得强语言建模能力<sup>[27]</sup>. 典型范式是在海量语料上进行下一个词预测, 以学习语言概率分布与语义结构<sup>[28]</sup>. 这一过程可视为无监督或自监督语言建模, 使模型逐步掌握词汇、句法、语义乃至一定的世界知识. 预训练后, LLM 的隐藏向量往往蕴含与多类自然语言处理 (natural language processing, NLP) 任务相关的表示与特征<sup>[29]</sup>. 当前大型模型在预训练后已能生成较长、连贯的文本, 并表现出一定的推理与知识整合能力.

在完成预训练后, 大语言模型通常还需要进一步的对齐过程, 以确保其输出更符合人类使用意图. 常见的做法包括指令微调与基于人类反馈的强化学习 (reinforcement learning from human feedback, RLHF)<sup>[30]</sup>. 指令微调通常在包含指令与响应的标注数据上继续训练, 以提升模型对任务说明的理解与合规作答能力<sup>[31]</sup>. RLHF 通过收集人工偏好训练奖励模型, 并据此进行策略优化, 从而引导 LLM 生成更符合用户期望的答案<sup>[32]</sup>. 这些步骤通常能显著改善模型在问答、内容生成等交互性任务中的稳定性与安全性. 对于电力系统这类对可靠性要求极高的场景, 应在部署前结合充分的验证机制与约束手段, 确保模型输出的专业性与准确性<sup>[33]</sup>. LLM 基于 Transformer 架构获得了强大的语言建模能力, 但在微电网等特定应用中, 还需要通过领域化的微调与知识注入来增强决策能力. 可利用历史调度记录、技术规程等专业语料进行进一步训练, 以嵌入电

力领域知识<sup>[34]</sup>, 并结合强化学习方法, 使模型在遵循电网运行约束的前提下优化其决策与规划过程<sup>[35]</sup>.

## 1.2 训练范式与性能演进

随着模型参数规模与训练语料的持续扩张, LLM 在语言理解与生成基准上的表现总体呈持续提升趋势<sup>[25]</sup>. 为了梳理技术演进, 简要回顾近年的代表性模型及其核心指标<sup>[36]</sup>. 表2汇总部分主流 LLM 的参数规模、训练数据与典型基准结果. 自 BERT 引领大规模预训练后, GPT-3 将通用语言模型推至百亿量级, 并展现突出的少样本、上下文学习能力<sup>[26]</sup>. 2022年的 PaLM 将参数扩大到 5400 亿, 在多个基准上取得突破. 2023年, GPT-4 引入图文多模态并在推理问答上接近人类水平. 同年, 大语言模型元宇宙应用 (large language model meta AI, LLaMA) 提供了开放权重的 65 亿参数模型, 以较小规模实现了出色性能. 清华的 GLM-130B 则作为国内开源的百亿级模型, 在中英文处理上取得突破. Grok 3 在推理与数学、编码上明显进步, GPT-5 则在代码、智能体式工具调用与可控输出上强化开发者体验. 这一系列模型的演进体现出模型规模、训练数据与范式改进的协同作用, 不断拓展 LLM 的应用边界.

总体而言, 参数规模与数据量的提升, 叠加更合算的训练与对齐策略, 使零样本、小样本能力显著增强. 这意味着模型无需针对每个任务单独监督训练, 也能依据任务描述与少量示例完成新任务, 对需求多变的微电网场景尤其应用价值<sup>[37]</sup>. 性能提升并不局限于纯语言类基准, 部分模型在数学、科学问答等

表2 部分主流 LLM

模型	年份	参数规模	主要进展
BERT	2018	3.4亿	Transformer双向编码, 预训练范式奠定NLP新基础
GPT-3	2020	1750亿	上下文少样本学习能力, 零样本推理表现突出
PaLM	2022	5400亿	超大规模预训练, 模型能力持续提升
GPT-4	2023	未公开	支持图像等多模态输入, 复杂推理接近人类水平
LLaMA	2023	65亿	开放模型权重, 中等规模下性能优异
GLM-130B	2023	1300亿	中英双语预训练, 国内开源大模型突破
Qwen2.5-72B	2024	720亿	在语言理解、推理、数学与代码等广泛基准上接近一流模型
DeepSeek-R1	2025	6710亿	通过大规模强化学习, 推理任务表现可与OpenAI o1相当
Grok 3	2025	未公开	在推理、数学、编码与指令跟随上显著提升
GPT-5	2025	未公开	新一代通用大模型, 代码与智能体式工具调用增强

技术性任务上也取得显著进展<sup>[38]</sup>. 但也有工作对其度量基础提出质疑<sup>[39]</sup>, 例如多步推理与复杂数学题的链式提示效果, 往往要求较大模型规模才能稳定发挥<sup>[40]</sup>. 因此, 在面向微电网的应用中, 模型规模可能影响模式识别与决策质量, 较小模型或难以覆盖复杂运行约束, 而大模型蕴含更丰富的隐式知识. 但受制于边缘算力与实时性, 直接部署超大模型并不现实, 需借助量化、剪枝与蒸馏等压缩技术以满足落地<sup>[41]</sup>.

综上所述, 大语言模型经过近几年的飞速发展, 已经奠定了强大的通用智能基础. 从 Transformer 架构、预训练范式到模型规模演进可以看到 LLM 在语言理解与生成方面的卓越能力. 然而, 要将其引入微电网负荷优化领域, 还需跨越专业知识和实际应用环境的鸿沟.

## 2 微电网负荷分配技术现状

微电网负荷分配是能量管理系统 (energy management system, EMS) 的核心目标, 其本质是如何在保证电力平衡与供电可靠性的前提下, 将局域负荷合理分配到分布式电源、储能及可调负荷之间, 实现经济性与灵活性最优. 该类问题常被形式化为多目标、多约束的优化调度模型, 并因随机性与场景不确定性而呈现较高复杂度<sup>[42]</sup>.

已有综述研究指出, 微电网负荷分配往往依赖于若干关键环节的支撑, 包括电源调度、负荷预测、储能管理和需求响应<sup>[43-44]</sup>. 电源调度决定各电源出力方案, 负荷预测为分配提供未来需求信息, 储能管理通过充放电平衡波动, 需求响应则通过用户侧调节增强灵活性. 因此, 本文在后续讨论中亦以这4个方向为主线, 对大语言模型在相关任务中的应用进行系统梳理和评述.

虽然 LLM 在微电网与输配电系统中的应用均涉及预测和决策辅助, 但两者的底层逻辑差异显著.

大电网调度更侧重于跨区域的潮流协调与规程一致性, 而微电网则需解决分布式电源、储能与负荷之间的紧耦合问题. 在微电网场景下, 系统不仅要实时吸收波动剧烈的新能源, 还需频繁处理并网与孤岛模式切换下的多目标权衡, 这对模型的局部协同能力提出了更高要求.

这种特有的运行环境也为 LLM 的应用带来了实际困难. 首先, 微电网规模虽小, 但运行状态切换极快, 模型必须能精准识别当前情境并迅速调整策略. 其次, 储能 SOC 或用户侧热舒适度等非线性约束高度交织, 任何由于模型产生的调度偏差, 都可能直接危及局部供电的经济性甚至安全性. 此外, 考虑到微电网对实时性的严苛要求, LLM 的角色应定位在慢时间尺度的能量管理与策略支持上, 而非介入底层的闭环控制逻辑. 因此, 本文在后续评述中, 将针对这些特定约束, 深入探讨 LLM 在微电网语境下的适用边界.

### 2.1 微电网负荷分配概述

微电网的负荷分配是其能量管理的核心任务. EMS 负责监测分布式电源、储能与负荷状态, 并基于对未来负荷和可再生出力的预测, 求解优化调度方案<sup>[45]</sup>. 通常情况下, 优化目标包括运行成本最小化或污染排放最小化, 决策变量涵盖各时段机组出力 and 储能充放电功率, 约束条件涉及功率平衡、设备容量与网络安全<sup>[46]</sup>. 在并网与孤岛两种运行模式及多重不确定性背景下, 调度一般采用日前计划与实时滚动修正相结合的策略.

在 LLM 介入前, 学术界与工业界主要采用数学规划、元启发式、数据驱动预测等方法<sup>[43]</sup>. 其中, 混合整数线性规划 (mixed integer linear programming, MILP) 等数学规划方法长期占据主导<sup>[47]</sup>, 但由于线性化近似与维度增长, 难以直接表达复杂非线性机理, 常需将混合整数非线性规划 (mixed integer

nonlinear programming, MINLP) 进行松弛或近似处理<sup>[48]</sup>. 为应对非线性与不确定性, 遗传算法、粒子群、蚁群等元启发式被广泛用于寻优<sup>[49]</sup>, 但在高维与强约束问题中可能出现收敛缓慢或解不可行等问题. 鲁棒与随机优化可显式处理不确定性, 但鲁棒优化可能带来保守性, 而随机优化依赖分布假设<sup>[50]</sup>.

除优化模型外, 预测技术是负荷分配的关键支撑<sup>[51]</sup>. 高精度的负荷与可再生出力短期预测是实现优化调度的前提, 近年来支持向量机 (support vector machine, SVM)、随机森林、LSTM、卷积神经网络 (convolutional neural network, CNN) 等方法较传统自回归积分滑动平均模型 (autoregressive integrated moving average, ARIMA) 等统计模型在多场景下取得显著精度优势<sup>[52]</sup>. 在执行层面, EMS 通过下达分布式电源出力与储能充放电策略来落实优化分配<sup>[44]</sup>. 常见策略包括优先级控制、下垂控制以及集中式和分布式协调控制<sup>[53]</sup>. 在孤岛运行条件下, 为保证电压与频率稳定, 储能与可调负荷通常作为关键调节手段<sup>[54]</sup>.

图 2 以概念图形式描绘了微电网典型架构. 可以看到, 在微网内部, 不同能源通过能源路由器进行交互, 分布式电源、储能和负荷通过控制器实现协同.

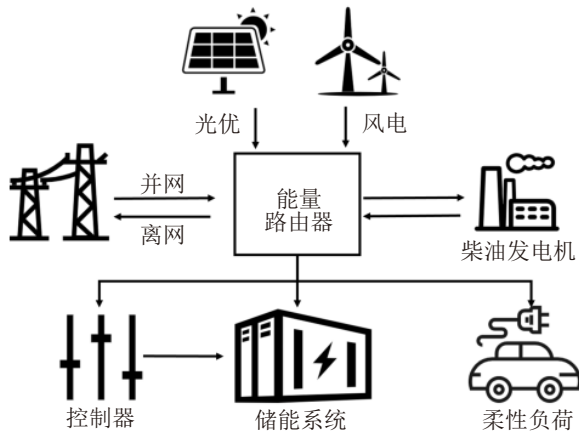


图2 单个微电网结构

表 3 总结了现有方法的特点: 数学规划方法能够保证全局最优解, 但高度依赖精确模型, 难以直接处理非线性和大规模不确定性; 元启发式算法在非线性问题上具有优势, 然而难以保证全局最优且高维强约束下易出现收敛困难; 数据驱动的强化学习等方法无需精确机理模型, 具有一定自适应能力, 但其解的最优性尚难以保证; 基于经验的规则控制实现简单、实时性好, 但缺乏全局优化能力, 难以应对复杂未知情形. 综上所述, 现有方法各有优劣, 难以兼顾微电网优化的高精度与高灵活性, 这凸显了引入更智能的 LLM 决策手段的必要性.

表3 微电网调度方法对比

性质/方法	数学规划	元启发式算法	数据驱动	规则与经验控制
全局最优解保障	√[43][47]	×[4][49]	×[3][4]	×[44][53]
非线性建模能力	×[4][47]	√[4][49]	√[5][50]	×[44][53]
不确定性适应性	×[43]	×[49]	√[3][4]	×[44]
实时自适应能力	×[42][43]	×[49]	√[3][4]	√[44][53]
需精确模型	√[43][47]	√[4][49]	×[3][5]	×[44][53]

## 2.2 负荷预测技术现状

负荷预测是微电网能量管理的关键环节, 其精度直接影响调度决策与运行绩效<sup>[55]</sup>. 现有方法可概括为统计模型与机器学习两大类<sup>[56]</sup>. 统计模型如 ARIMA、季节指数平滑、季节性自回归积分滑动平均模型 (Seasonal AutoRegressive Integrated Moving Average, SARIMA) 与周期分解等应用广泛, 依托历史序列的自相关与周期性建立线性关系进行预测<sup>[57]</sup>, 但在受天气与节假日等外生因素影响且存在非线性的场景中, 精度常受限<sup>[58]</sup>. 机器学习在短期负荷预测中整体表现优于纯统计法, SVM 借助核函数拟合非线性、神经网络 (Artificial Neural Network, ANN) 可自适应学习复杂映射, 融合气象等外生变量的多元建模可提升预测准确度. 然而, 传统机器学习依赖特征工程与调参与再训练策略, 其在极端或漂移场景下泛化仍受挑战<sup>[59]</sup>.

深度学习显著提升了负荷预测性能, RNN、LSTM 和门控循环单元 (Gated Recurrent Unit, GRU) 通过记忆单元刻画长短期依赖<sup>[60]</sup>. 大量研究显示, LSTM 在日级与分钟级预测中常优于 ANN、SVM. 同时, CNN 可提取局部与周期模式, 与 LSTM 结合形成 CNN-LSTM 可增强稳定性. 对多周期、非平稳负荷, 小波分解与预测的组合能进一步提升精度<sup>[61]</sup>. 深度学习亦便于融合天气、节假日等多源信息. 不过, 其对数据规模与可解释性的要求, 在工程应用中仍面临一定挑战.

表 4 分类总结了负荷预测领域从现有方法到引入大模型的新趋势. 统计模型依赖线性假设, 难以捕捉非线性负荷模式, 对天气、节假日等外部因素的融入能力有限. 机器学习方法能够拟合一定非线性并结合多种特征, 但特征设计工作量大, 在极端状况下泛化能力不足. 深度学习显著提升了预测精度, 能自动提取特征并融合多源数据, 但通常需要大量训练样本, 并且可解释性表现相对有限.

## 2.3 储能与需求响应优化

储能系统和需求响应 (Demand Response, DR) 是微电网实现负荷灵活调配的两个重要手段, 在调度中分别承担蓄水池和可控负荷的作用. 传统上, 对

表4 负荷预测技术对比

方法	统计模型	传统机器学习	深度学习
非线性模式拟合	×[56][57]	√[58][59]	√[60][61]
自动特征提取	×[56][57]	×[58][59]	√[60][61]
融合多源信息	×[56][57]	√[58][59]	√[60][61]
小样本、零样本能力	×[56][57]	×[58][59]	×[60][61]

储能的优化管理和对可调负荷的激励主要依赖规则和简单优化模型,但随着新能源和电动车的大规模接入,储能与需求响应的调度愈发复杂,需要引入智能化决策方法<sup>[62]</sup>.

储能优化方面,早期研究多采用阈值式充放电策略以平滑负荷<sup>[63]</sup>,但其难以充分发挥储能容量潜力<sup>[64]</sup>.随后,研究普遍将储能充放电功率作为优化变量,纳入微电网调度统一求解,并用于削峰填谷、降低购电成本<sup>[65]</sup>.部分模型进一步考虑电池寿命衰减成本,通过优化充放电深度等延缓退化<sup>[66]</sup>.在孤岛运行情形,为兼顾经济性与频率稳定,常采用计划-实时双层优化或分层控制框架<sup>[67]</sup>.总体而言,确定性优化对随机波动的实时适应性有限<sup>[68]</sup>,引入强化学习等方法可使储能策略随状态自适应调整,相关研究已快速增多<sup>[69]</sup>.

需求响应旨在高峰时段削减用电、低谷时段增加用电以平衡负荷<sup>[70]</sup>.典型机制包括分时电价(Time-of-Use Pricing, TOU)、实时电价(Real-Time Pricing, RTP)与直接负荷控制(Direct Load Control, DLC)<sup>[71]</sup>.其中,TOU通过预设峰谷电价引导用户错峰用电<sup>[72]</sup>;RTP随系统状态动态定价,但用户实际响应可能存在滞后<sup>[73]</sup>;DLC则通过远程控制可中断负荷进行直接调节<sup>[74]</sup>.由于上述方案往往依赖用户被动响应或高集成度,激励效果与可持续性受限.为提升效果,智能DR结合家庭能源管理系统,依据价格与用户偏好自动调节如空调与电动车充电等负荷<sup>[75]</sup>,亦有研究采用博弈论或多智能体协商以实现聚合商与用户的互动<sup>[76]</sup>.近年兴起的竞价平台与虚拟电厂框架将DR视为可交易的资源,通过市场化机制参与调度<sup>[77]</sup>.然而,精准刻画用户行为并设计个性化激励依然困难<sup>[78]</sup>.在此背景下,LLM基于历史用电与反馈信息可辅助建模弹性与舒适偏好,从而生成更

个性化的DR策略推荐<sup>[79]</sup>.

需要明确的是,本文所讨论的负荷最优分配是指广义上的微电网能量优化目标.在微电网语境下,它与能量管理、经济调度等概念高度重叠.负荷最优分配是EMS的核心决策任务,其本质是通过协调电源侧(出力调度)、储能侧(充放电控制)及需求侧(负荷响应),在满足系统约束的前提下实现功率的实时平衡与经济最优.因此,本文后续章节讨论的电源调度、负荷预测等,均被视为支撑负荷最优分配这一核心目标的关键技术环节.

### 3 LLM在微电网中的应用研究

LLM在微电网中的应用并非孤立存在,而是通过赋能不同的子环节来共同支撑负荷最优分配.本章将从支撑该目标的四个关键维度,电源调度、负荷预测、储能管理和需求响应,评述LLM的最新研究进展.

#### 3.1 LLM在电源调度中的应用

电源调度优化是微电网负荷分配的核心任务,旨在确定分布式电源出力与主网购售电策略以满足供需平衡并兼顾经济与安全目标<sup>[43]</sup>.传统做法以数学优化为主,近年来引入LLM以其推理与决策能力直接或辅助生成调度方案.

当前探索大致分为两路,如表5所示.其一是将LLM用作辅助决策工具,在调度场景中生成初步方案或提供情境化分析,由调度员进行审核与裁决后再实施.这种方式下,人类专家可以对LLM建议进行把关,确保不违反电网物理约束,提高方案安全性.同时,LLM擅长从海量非结构化信息中提取有用情境,例如根据运行记录和天气文本提供辅助分析,从而减轻人工信息处理负担.其不足在于最终效果仍依赖人工经验,LLM给出的建议质量不稳定,调度员需要投入精力甄别和调整.

需要指出的是,表5中的局限性并非单一问题,而是由多重因素共同导致.首先,现有电力调度类LLM研究多基于仿真数据和离线样本开展训练与评测,极端天气、设备故障、通信异常等长尾工况覆盖不足,容易造成部署阶段的分布偏移.其次,通用LLM的生成机制以语言概率建模为主,若未将功率

表5 LLM在电源调度中的应用模式

方式	人机协同调度(LLM辅助决策)	LLM自主调度(自动决策)
应用模式	LLM生成调度建议,人工审核执行	LLM直接生成调度方案并下达执行 <sup>[80]</sup>
优势	人工把关确保安全合规,避免物理违规	全自动快速决策,响应速度快可直接输出可行方案,减少人工参与
局限	依赖调度员经验,结果受人因素影响. LLM建议质量有限仍需专家判断	模型可靠性不足,可能违反约束.缺少人为校验,错误风险需额外监控
典型实例	eGridGPT调度助手 <sup>[14]</sup>	GAIA自动调度模型 <sup>[81]</sup>

平衡、设备容量、网络安全等物理约束显式嵌入推理或解码过程,则可能出现“语义上合理但物理上不可行”的方案。再次,即使模型输出格式正确,也不代表数值结果满足工程可执行性,因此仅依赖文本一致性或问答准确率难以充分表征调度可靠性。

从现有研究路径看,领域微调可提升模型对电力术语、拓扑与典型操作规程的适配能力,但其效果仍受训练数据质量与覆盖范围限制;检索增强与知识注入有助于降低事实性错误,但对连续变量优化的约束作用有限;“LLM+求解器”协同范式能够将场景理解与连续优化分离,在可行性方面更具优势,但系统集成复杂度与实时调度链路时延也相应增加。因此,在微电网场景下,相较于完全自主决策,现阶段更适宜将 LLM 优先用于场景理解、策略生成与人机协同辅助,并通过求解器校核、仿真验证和人工放行构建分层安全机制。

其二作为全自动模式由 LLM 直接生成调度方案。如 GAIA(Grid-Aware Intelligent Agent) 针对电力调度进行领域微调,融合负荷、成本与网络拓扑等多维数据,并配套场景化提示与 ElecBench 评测<sup>[81]</sup>。实证表明,GAIA 在 ElecBench 多项指标上优于 LLaMA2,并在运行监控等场景的安全性评分上超过 GPT-3.5,且能在电压越限案例中给出更准确的诊断与处置建议。其有效性源于对电力专业知识与模拟数据的充分注入,使模型在专业约束下具有更强的任务适配能力。

LLM 在电力调度中的一项重要潜能,是对非结构化文本(如规程、日志与告警)的建模与利用<sup>[34]</sup>。在实际运行中,调度除依赖 SCADA 等数值量测,也需参考运维记录与气象信息等文本材料。据此,LLM 可把文本纳入决策链路,提升情境感知。在极端天气等场景中,可提出预防性调度建议,并通过仿真校核。面向未来,数字孪生+AI 的跨模态智能调度被视为控制室的重要发展方向<sup>[82]</sup>。

与此同时,LLM 用于调度仍面临可靠性与安全性挑战。从任务类型看,现阶段应用多集中于离散、规则型操作与应急处置,而对大规模连续调度量的直接输出仍存难点。一条可行路径是让 LLM 完成场景理解与策略模板生成,再调用外部优化求解器计算连续量,形成 LLM+Solver 的协同范式<sup>[83]</sup>。此外,受限于推理时延与算力开销,通用 LLM 难以直接应用于对实时性要求极高的调度环节,更适合在负荷分配与人机协同决策中提供支持。通过蒸馏、量化以及端侧部署等技术,可在提升响应效率的同时兼顾数据隐私与可用性<sup>[84]</sup>。

**LLM 在电源调度应用中的典型案例:**在电源调度领域,电力专用大模型 GAIA<sup>[81]</sup> 是目前最具代表性的自主调度方案。该研究通过将仿真数值数据与电力专业文献相结合,在 LLaMA2 基础上通过 LoRA 微调,使模型具备了直接生成满足复杂物理约束的调度策略的能力。

该研究的训练流程如图 3 所示。首先,在标准电力系统算例上构造多工况数据,覆盖运行调整、运行监测等典型情形,得到可结构化的状态、动作与结果记录。同时,从教材、论文与技术规范中抽取与调度相关的定义、约束与推导,形成文本知识。随后,对两类数据进行清洗与标准化,统一单位与时间索引,并据此将样本转写为问答与指令格式,输入与输出字段采用固定的 JSON 结构。在完成一致性与可解性检查后,基于基座语言模型开展有监督指令微调。训练过程中监控验证集损失并采用早停策略选择最优权重。训练完成后,对离线推理结果进行格式与物理约束校核,并在既有基准上开展对比评测。

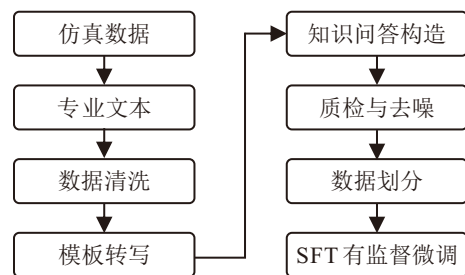


图3 GAIA 训练流程

如图 4 所示,调度人员仅需以自然语言描述电网运行状态及约束条件,模型便可依托训练阶段融合的仿真数据与电力系统专业知识,自动生成机组出力方案和成本评估结果,并同步识别潜在风险及相应调整建议。

与现有的调度优化方法相比,该模式不再依赖复杂的数学建模过程。调度员能够直接提出任务需求并获得结构化输出,结果除包含数值解外,还附带风险分析与改进意见,从而提升人机协同决策的效率与可解释性。

### 3.2 LLM 在负荷预测中的应用

基于大语言模型的负荷预测是当前研究的一个新颖方向,主要思路是将负荷时间序列数据转化为语言,利用 LLM 的序列建模能力进行预测。这一方法通常包括两个步骤:(1)提示设计:将历史负荷数值序列编码成文本形式的提示;(2)LLM 序列续写:让预训练好的语言模型读取提示,生成对后续时刻负荷的预测描述。

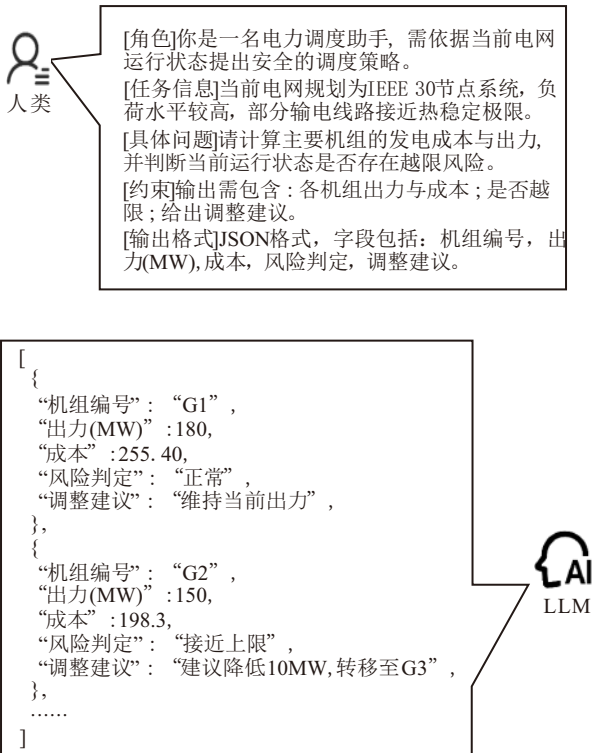


图4 LLM在电源调度中的应用

提示设计方面,常用做法是设计一个模板句,将时间和负荷值嵌入其中。例如,可采用将每个时刻与用电量改写为文本或数字串的做法,使模型以自然语言或符号序列的形式接收历史轨迹。对一段长度为 $n$ 的历史窗口,得到 $n$ 个句子作为提示输入LLM,从而实现由数值到语言序列的转换<sup>[85]</sup>。这种转换的合理性在于,负荷序列的模式(涨落、周期)可在文本模式中显性化,LLM具备延续此类模式的能力。提示中通常还包括时间信息,以帮助模型捕捉日周期、周周期等特征<sup>[86]</sup>。此外,为突出关键趋势,也有工作将连续数值离散化(如分段或高、低负荷)以便于建模,但这一过程往往会牺牲部分数值精度,仅能在区间或等级层面反映负荷变化<sup>[87]</sup>。

在LLM预测生成阶段,可视作文章续写,给定前文提示,让模型生成后续的负荷描述。模型会基于训练中对语言与模式的学习,推断下一步的合理输出<sup>[88]</sup>。目前研究多关注短期负荷预测,因此生成的句子序列长度通常与历史窗口相当。对于多步预测,常用自回归策略:一次生成一个时段,将其加入提示再预测下一个,逐步外推。该自回归方法与传统时间序列逐步外推一致,能利用最新预测刷新上下文并刻画多步依赖关系<sup>[89]</sup>。

表6概括了将大型语言模型用于短期负荷预测的优势和局限。与现有方法相比,LLM利用预训练获得的通用知识,在训练样本很少甚至零样本的情况下也能给出合理的负荷预测,展现出强大的迁移

泛化能力。这避免了每个新场景都需重新训练模型的开销。同时,LLM模型可以统一适配多种任务类型,不局限于特定电力数据格式。然而,当前LLM用于数值时序预测还存在明显不足。首先,通用模型对数值的敏感度有限,可能出现数字精度问题,需要在训练与提示中加强约束与校准<sup>[90]</sup>。其次,若以自然语言生成结果,还需从文本中稳健地解析为数值或借助分布映射,这会引入额外步骤。再次,LLM微调需要一定算力且数据格式特殊,在小样本下易不稳定<sup>[91]</sup>。

表6 LLM在负荷预测中的优势与局限

优势	局限
小样本数据下预测精度相对更高	数值敏感度有限,易出现数字误差
可直接利用预训练知识,无需针对特定场景重新训练	预测结果以文本形式给出,需额外解析为数值
具备跨领域迁移能力,可一套模型适配多种任务	模型微调成本高,格式设计和对齐需较大算力支持

从表6可以看出,现有研究在任务目标、数据来源、实验场景和评价指标方面差异较大,因此不同方法的结果不宜直接进行简单横向比较。部分研究强调LLM在场景理解、策略生成或人机交互方面的优势,另一些研究则关注调度成本、约束满足率或运行稳定性等工程指标,两类工作的评价口径并不一致。此外,不少研究仍以仿真平台或离线测试为主,对在线闭环运行、极端工况响应和长期稳定性的验证相对不足。上述差异说明,当前文献总体上仍处于概念验证阶段,后续研究需要在统一测试基准、标准化指标体系和可复现实验流程方面进一步加强,以支持更有说服力的比较分析。

**LLM在负荷预测应用中的典型案例:**针对负荷预测任务,不同于专用模型路线,文献<sup>[85]</sup>探索了利用通用大语言模型通过提示工程处理时序数据的可能性。该方案通过数值转写技术,将连续负荷数值转化为自然语言片段嵌入提示语,成功将负荷预测转化为LLM擅长的文本序列补全任务。

其整体流程如图5所示。首先,将连续时间序列离散化并编码为数字字符串,通过控制空格与逗号的位置以确保语言模型能够逐位识别数值变化。随后,对输入数据执行线性缩放与偏移,使其数值分布集中于统一尺度区间,从而提高模型的可泛化性。该预处理后的序列被输入到已预训练的大型语言模型,模型无需额外参数更新即可基于上下文进行自回归预测。预测阶段通过温度与核采样等机制生成多组可能的时间序列扩展,以其统计分布构建确定性和概率性结果。为实现精确的连续密度估计,提出将离散数字序列的条件概率层级化处理,形成混合型连

续分布,用于似然计算与不确定性评估.

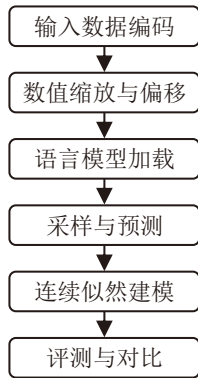


图5 LLM 在负荷预测中的训练流程

如图6所示,该案例展示了利用 LLM 进行微电网负荷预测的提示构建与结果输出过程. 首先,在提示中嵌入站点信息、历史数据以及任务约束,通过自然语言的形式将数值型时序转写为语言化片段. 随后, LLM 在提示引导下生成与能源管理系统分辨率一致的未来若干时段预测结果,并附带自然语言解释. 最终,所有预测结果以可解析 JSON 格式输出,其中每个时段包含“时段、预测负荷、解释”三要素,既保证了结果的结构化,便于后续调度调用,又保留了解释性文本.

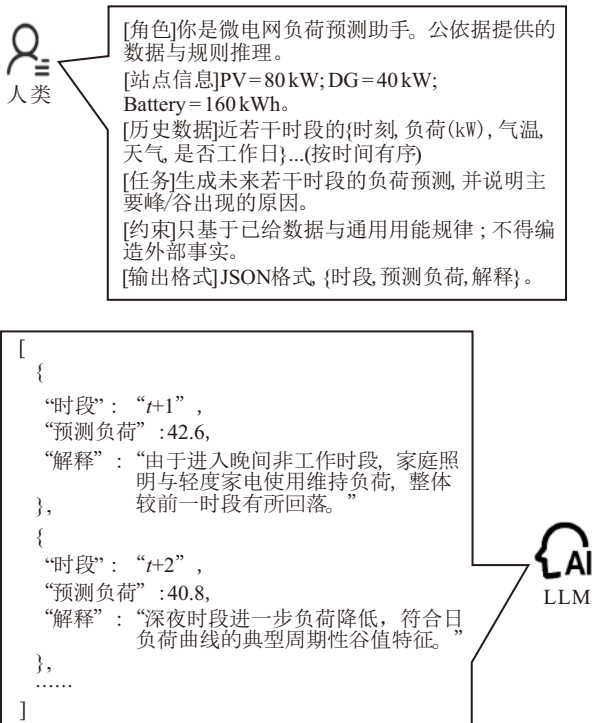


图6 LLM 在负荷预测中的应用

### 3.3 LLM 在储能管理中的应用

储能管理涵盖电池、超级电容等装置的优化调度与寿命管理,在微电网中发挥蓄能缓冲作用. 合理运用储能可削峰填谷、提高供电可靠性并降低运行

成本<sup>[92]</sup>. 传统储能调度多采用规则或优化框架,引入大语言模型为从数据与知识中学习运行策略提供了新路径<sup>[93]</sup>.

目前将 LLM 直接用于储能调度的公开研究仍然有限,但已有相关探索可供借鉴<sup>[94]</sup>. 一方面,研究已尝试利用 LLM 的跨域泛化能力开展电池状态估计;另一方面,基于时间生成式预训练 Transformer (Time Generative Pre-trained Transformer, TimeGPT) 等基础时序模型的微调可在较少针对性训练的情况下,准确表征电池健康状态 (State of Health, SOH) 随循环演化的趋势<sup>[95]</sup>. 由此可见, LLM 有潜力作为统一的储能建模工具,在大规模电池数据上预训练以抽取老化规律与效率特性,并迁移用于微电网场景的储能预测与调度决策.

LLM 还能辅助储能充放电的优化决策生成. 例如在能源管理的决策序列生成与校验中, LLM 可结合负荷、电价等预测,给出何时充、放电的候选方案,再借助强化学习或搜索机制进行验证与改进<sup>[96]</sup>. 凭借长上下文与长序列推理能力, LLM 能够在全天负荷轨迹与多约束下给出较现有方法更为灵活的时序策略,且在非常规情景 (如突发停电、极端天气) 中可根据训练中学到的相似事件描述快速调整<sup>[93]</sup>. 当然,针对 LLM 用于储能优化的系统性公开验证仍然不足,尚需进一步实验与评估.

隐私保护方面,可将 LLM 与联邦学习结合,微电网的储能数据分散且敏感,不宜集中汇聚,而联邦学习通过交换模型参数而非原始数据实现数据可用不可见<sup>[97]</sup>. 已有研究将 LLM 微调过程联邦化,使各微网共享参数更新,得到兼具泛化与隐私保护的全局模型. 该框架同样适用于包含储能调度在内的多类任务<sup>[98]</sup>. 当各微网储能配置与策略存在差异时,联邦微调的 LLM 可以学习跨场景共性并输出个性化建议<sup>[99]</sup>.

需要强调的是,储能优化的关键难点在于经济收益与寿命退化的权衡<sup>[100]</sup>. 考虑到目前 LLM 对老化机理的精确刻画能力仍有待提升,应在策略生成中引入物理信息约束以避免深循环导致的过度退化<sup>[101]</sup>. 从中长期看,可采用“生成-验证-反馈”的闭环机制,使 LLM 生成的策略在验证与反馈基础上不断优化,并结合强化学习迭代形成更稳健的技术路线<sup>[102]</sup>.

**LLM 在储能管理应用中的典型案例:** 在储能管理环节,文献<sup>[93]</sup>提出了一种由 LLM 担任高层决策引擎的 Transformer 与 LSTM 协同架构. 在该体系中,大语言模型不直接参与低层控制,而是利用其强

大的上下文关联能力分析电池运行规律,生成具备可解释性的储能调度建议。

如图7所示,该模型的训练流程包括预训练、领域适配和任务评估三个阶段。首先,模型在通用语料上通过自监督学习获得语言理解与生成能力;随后,研发团队构建了涵盖能源系统运行日志、监测记录与技术文档的领域数据集,并通过数据标准化与指令模板化实现结构化与非结构化信息的统一表示;最后,采用有监督微调完成领域适配,从而在控制计算复杂度的同时减少过拟合风险。经过训练的模型在能源预测、运行优化和智能调度等任务中进行了验证,多项指标均显示其具备较好的泛化性能。

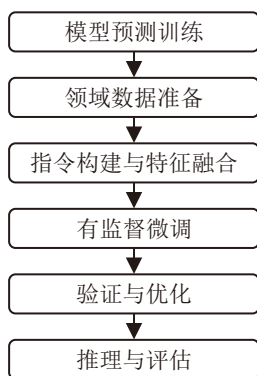


图7 LLM在储能管理中的训练流程

图8展示了该模型在储能调度中的实际应用。模型能够综合光伏出力预测、电价变化及SoC状态等多源信息,在自然语言提示的引导下自动生成符合运行约束的充放电策略,并以结构化形式输出功率分配与SoC演化结果。与传统仅输出数值最优解的方法相比,该模型在提供可执行方案的同时,还给出策略解释说明,使削峰填谷与经济优化过程更具透明性与可理解性。

### 3.4 LLM在需求响应中的应用

需求响应通过经济激励或直接调度引导用户调整用电行为,是微电网中削峰填谷与提升可靠性的重要手段<sup>[103]</sup>。传统DR多依赖预设价格机制与用户被动响应,而大型语言模型有望通过更智能的人机交互与行为建模提升DR的精细化与有效性<sup>[104]</sup>。

文献<sup>[105]</sup>提出的两层电动汽车需求响应框架利用LLM模拟用户充电决策,该框架划分为上层聚合商定价与下层用户响应两个层次,其中上层依据收益目标设定动态电价,下层大规模用户结合电价与个体需求作出充电选择。为逼真刻画用户决策,研究在下层引入LLM决策生成模块,使其读取用户画像与外部情境生成充电意向与行动。上述结果说明LLM在模拟人类多因素决策并生成合理行为方面

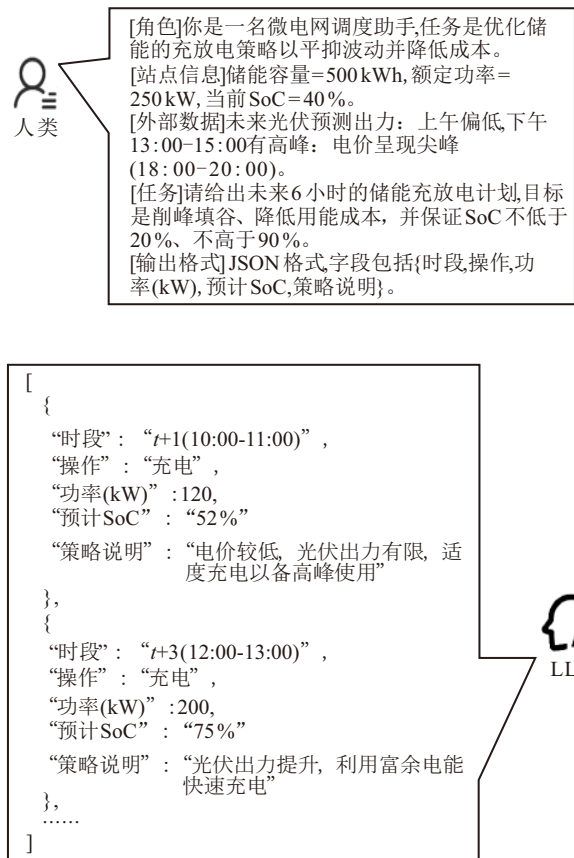


图8 LLM在储能管理中的应用

具有潜力,这对难以用解析模型刻画的人因问题尤为关键。

除行为模拟外,LLM还能支持DR的人机交互,可构建智能能源助理与用户对话,在高峰电价时主动给出可执行建议并按偏好调整<sup>[106]</sup>。该类对话智能体由LLM驱动,能够理解自然语言并提供个性化、可解释的节能、移峰建议,从而提升参与度与用户体验<sup>[107]</sup>。已有学术研究与试点系统验证了ChatGPT式对话在能效建议中的可行性,为微电网运营方后续开展本地化微调与部署提供了依据<sup>[108]</sup>。

同时,应重视伦理与隐私,用户用电与偏好数据敏感且具再识别风险,DR交互需遵循隐私政策与最小化原则<sup>[109]</sup>,建议设计需兼顾公平性,避免仅对部分用户有利<sup>[110]</sup>,并通过数据与算法层面的偏差缓解,提升对不同群体的建模准确性。在工程实现上,可构建反馈-学习闭环,结合强化学习等方法让模型依据真实DR效果持续修正策略,若发现某类用户长期不响应特定激励,则针对该类用户自适应调整策略与沟通内容<sup>[111]</sup>。

**LLM在需求响应中的应用:**对于需求响应场景,文献<sup>[105]</sup>展示了LLM在用户行为语义建模方面的独特优势。研究者利用大语言模型解析电动汽车用户的心理偏好与历史用电记录,构建了个体化的

响应预测模型. 这证明了在不需大规模电力数据微调的情况下, 通用大模型亦能通过对非结构化文本的理解, 动态生成最优充电决策.

训练流程如图9所示. 首先, 通过收集用户的社会属性、心理特征及充电偏好等多源数据, 利用LLM的语义解析能力生成高维用户画像, 从而刻画个体差异性. 随后, 引入环境状态模块以表征动态外部条件, 包括电池荷电状态、气象因素及实时电价. 基于用户画像与环境输入, LLM通过语义推理生成个性化的充电决策逻辑, 决定用户在不同时段的充电意图、时间选择与电量需求. 该决策模型在多智能体系统中运行, 各智能体独立作出响应并形成整体负荷曲线. 上层聚合器利用粒子群优化算法实时调整零售电价, 以实现利润最大化与电网负荷平衡. 通过多轮交互仿真与迭代优化, 模型实现了“电价—行为—负荷”的动态闭环.

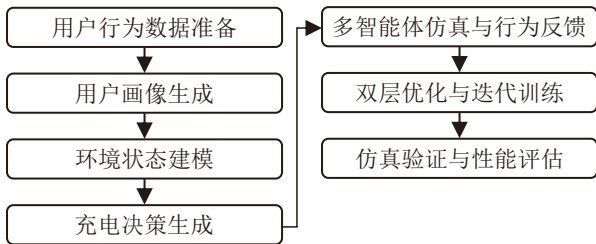


图9 LLM在需求响应中的训练流程

如图10所示, 该研究展示了大语言模型在电动汽车需求响应场景中的应用实例. 模型在接收包含用户画像、车辆运行状态与实时电价等信息的提示后, 通过语义理解与推理生成个性化的充电决策, 并输出结构化的结果与相应的解释说明. 与传统依赖固定规则或数值优化的方法相比, 该框架能够根据价格波动与用户偏好变化自适应调整策略, 从而更准确地刻画不同用户的响应差异. 此类人机协同的建模方式显著增强了需求响应的灵活性与可解释性.

#### 4 关键技术问题

尽管LLM在微电网负荷分配中展现出显著潜力, 但其从理论模型向工程实践的转化仍受制于多重技术瓶颈. 尤其在微电网特有的并网与离网频繁切换、源储荷深度耦合以及分层控制的严苛约束下, 系统对算法失效的容错空间极小, 这使得模型的可解释性、多模态融合、实时性及安全隐私等问题变得愈发复杂.

本章将结合微电网的实际运行逻辑, 深入探讨这些关键挑战. 不同于通用场景下的性能评估, 本文将重点分析在电网安全边界限制下, LLM如何满足调度决策的物理一致性、多源信息的数据权重分配、

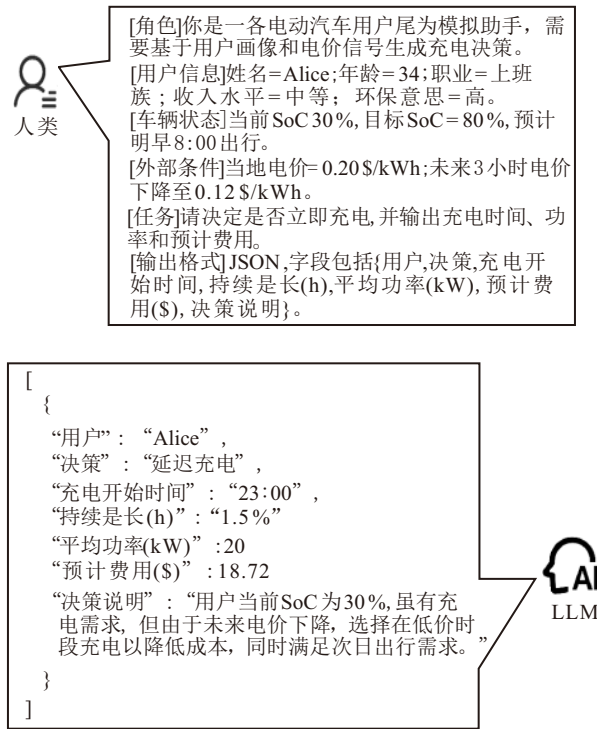


图10 LLM在需求响应中的应用

跨尺度控制的时间响应要求, 以及针对执行指令的风险管控, 探讨这些技术在具体落地时的性能边界与工程约束.

#### 4.1 可解释性与可信 AI

可解释性已成为人工智能在关键基础设施中应用时的核心议题<sup>[112]</sup>. 由于电力系统对运行安全和决策透明度具有严格要求, 运行人员及监管机构必须能够理解智能系统的判断依据, 以确保调度与控制过程的合理性<sup>[113]</sup>. 然而, 大语言模型作为一种高维度且结构复杂的模型, 其内部推理机制难以直接呈现, 这在一定程度上限制了其在微电网等关键场景中的推广与应用<sup>[114]</sup>.

为提升 LLM 在微电网应用中的可信度, 研究者提出多种技术路径. 其一是在生成调度方案时引入物理可解释性与一致性约束, 例如对功率平衡等关键约束进行校核, 不满足则拒绝或重生成, 以确保物理可行<sup>[115]</sup>. 其二是可视化模型依据, 如利用注意力权重分析其关注的输入区域<sup>[116]</sup>; 对文本输出可诱导其产生链式思考的中间推理步骤以提升可审计性<sup>[117]</sup>. 在调度语境下, 这意味着模型需显式说明“因负荷和电价处于高位, 故建议启用储能放电”等推理链条.

此外, 融合领域知识图谱亦能增强解释与约束, 通过将电网拓扑与故障因果等专家知识并入 LLM 的检索与输入通道, 可在收敛输出空间的同时提供基于知识的可追溯证据<sup>[118]</sup>. 例如, 当建议切除某支路以隔离故障时, 可引用图谱中该支路与故障设备的

关系作为解释支撑<sup>[119]</sup>. 在微电网情境下, 可据此构建包含拓扑、设备参数与历史规则的知识库以供查询与对齐<sup>[120]</sup>.

最后, 人机协同的审议放行机制同样关键, LLM 的建议先由人工或传统算法进行校核与完善, 再决定执行. 这种安排提供社会-技术层面的可解释保障, 由经验调度员据专业知识做出最终裁决<sup>[121]</sup>. 在此过程中, LLM 通过交互式问答提供理由与证据, 增强信任与可追溯性, 最终决策权仍由人掌握.

表7对比了提升 LLM 可信性的四类主流手段, 物理一致性校核可降低物理违规风险; 可视化与链式思考提升审计可读性; 知识图谱、检索增强提供可追溯证据; 人机协同确保关键场景的人为把关.

表7 可解释性与可信 AI

路径	作用环节	代价与风险
物理一致性校核	调度控制	计算量提高
可视化与链式思考	预测告警解释	推理时延提高
知识图谱、检索增强	诊断处置理由	构建成本提高
人机协同	最终决策	人工负担增加

需要注意的是, 可解释性增强方法并不等同于可靠性提升. 链式推理与解释文本可提高审计可读性, 但其解释内容本身仍可能与真实决策依据不一致; 相较之下, 物理一致性校核和外部求解器验证虽然增加计算开销, 却更直接作用于工程可执行性. 因此, 在微电网调度场景中, 应将“解释增强”与“可行性校核”区分开来, 前者提升可审计性, 后者保障安全边界, 两者缺一不可.

在微电网负荷分配中, LLM 的可解释性必须直接转化为辅助决策的实质工具. 一套具备参考价值的方案说明应能清晰还原决策背后的数据支撑与物理逻辑: 从当前的负荷水平、新能源波动及储能 SOC 等边界条件出发, 阐明机组启停或储能动作的具体触发诱因; 同时, 方案需明确交代对功率平衡、断面极限及爬坡约束等安全红线的校验结果, 并量化外购电成本、弃电风险与响应补偿之间的经济博弈.

这种解释能力在不同的业务环节中应有不同的侧重. 链式推理利用其逻辑递进的特性, 能够帮助值班调度员在实时任务中快速理清“现状—策略—约束”的闭环过程; 而知识图谱通过关联设备台账、运

行规程与历史案例, 更适合用于离线状态下的事故溯源与合规性审计. 此外, 为了弥补生成式模型在数值精度上的短板, 应引入物理一致性校核与外部求解器验证, 将 LLM 的语义逻辑与硬性的物理约束进行比对. 通过这种逻辑推导、规程关联与可行性验证的深度耦合, 才能在复杂的运行环境下构建起调度员对模型输出的工程信任.

## 4.2 多模态数据融合

微电网运行数据呈现多模态与多源特征, 既包含结构化量测数据, 也涵盖红外热像巡检等非结构化文本与图像信息<sup>[122]</sup>. 如何使大语言模型在决策过程中有效融合多模态证据, 是当前研究的关键议题<sup>[123]</sup>. 在复杂场景下, LLM 需并行整合负荷时序、气象文本与设备告警等异构信息, 方能形成稳健判断.

现有多模态 LLM 已支持图像-文本联合输入与理解<sup>[124]</sup>. 面向微电网场景, 结合文本、时序与表格的统一建模框架被广泛看好. 一条可行路径是将非文本信号语言化, 以自然语言概括负荷曲线等关键特征<sup>[125]</sup>. 在该流程下, LLM 可由语言化描述中提取时序要点, 实现对核心动态的把握, 该语言化策略已在时序预测基准得到验证. 同时可借助视觉-语言模型将图像转述为文本再交由 LLM 分析<sup>[126]</sup>.

进一步地, 可让 LLM 通过工具调用执行专业数值计算, 并将返回结果纳入上下文<sup>[127]</sup>. 该模式本质上是函数与工具调用, 由外部算法库完成计算<sup>[128]</sup>. 可向 LLM 提供负荷数据入口, 由其触发预测函数并结合语言知识形成建议<sup>[129]</sup>. 借助工具链的 LLM 框架, 可在一定程度上弥补纯文本模型在连续数值与高精度计算上的短板.

提升模型可靠性需要更好地结合多源数据与外部工具支持<sup>[130]</sup>. 电力系统长期形成了潮流与短路等标准化分析体系<sup>[131]</sup>. 若缺乏物理约束, LLM 生成的方案可能违反电气定律与运行边界. 因此, 应将物理仿真与计算结果纳入 LLM 的决策约束与提示.

表8归纳了语言化时序、视觉-语言转述、工具函数调用三条可行路径, 前两者侧重把异构感知转为可被 LLM 处理的文本证据, 后两者通过外部计算与物理校核弥补纯文本模型的数值短板.

多模态融合并非简单增加信息源即可提升性能.

表8 多模态数据融合

融合方式	输入模态	关键做法	优点	限制
语言化时序	文本	将时间-负荷曲线摘要成短句供 LLM 读写	兼容通用 LLM; 少样本友好	数值精度需后处理
视觉-语言转述	图像+文本	设备巡检、热像转文字描述再交由 LLM	融合非结构化证据	转述误差累积
工具函数调用	文本+工具输出	调用外部预测、求解器, 将结果回填上下文	数值精度高; 可控	集成复杂度提高

当时序数据被语言化压缩时,模型获得了更好的通用兼容性,但也可能损失峰值、突变等对调度高度敏感的数值细节;视觉转述路径可扩展非结构化感知能力,但图像到文本的转述误差可能在后续推理中被放大。相比之下,工具调用与外部计算可显著提升数值精度与约束一致性,但其接口稳定性、调用时延与系统耦合复杂度成为新的工程瓶颈。因此,多模态方案的优劣并非绝对,而取决于任务目标是“情境理解”还是“高精度数值决策”。

在处理微电网负荷分配这类任务时,多模态融合的实际收益很大程度上取决于输入数据与特定调度环节的匹配程度。在日前经济调度或滚动优化阶段,决策核心是电力功率的实时平衡,此时高频次的结构化数据(如预测出力、储能SOC、实时电价等)占据主导地位,数值精度直接决定了分配方案的优劣。相比之下,对于异常工况处置或调度复盘任务,告警日志、检修记录及规程条文等非结构化文本则提供了关键的背景支撑,有助于模型理解复杂的运行环境。

这种信息特性的差异要求多模态方案必须具备明确的任务导向。在对数值精度敏感的功率优化环节,应侧重于时序数据与外部求解工具的联动,以确保计算结果满足物理约束;而在风险预警或人机辅助环节,则应侧重于对文本知识和历史案例的关联分析,提升模型对突发状况的理解力。盲目增加模态数量不仅可能导致计算冗余,还可能在多源信息权衡中稀释关键控制信号。因此,多模态技术的应用重点应从单纯的“数据接入”转向根据调度任务需求进行的信息筛选与权重动态分配,从而确保融合后的方案在工程层面更具执行力。

总体而言,多模态融合和物理结合是LLM迈向实用智能调控的必由之路。为此,需要在模型训练中引入多源数据集,并使用联合训练或多任务学习策略,让模型同时掌握语言理解和数理推算。实现这一点需要电力与AI专家通力合作,开发新颖的网络结构或训练范式。但可以预见,其成果将极大提升AI接管复杂电力系统的能力。

### 4.3 实时性与模型高效部署

微电网负荷分配具有较强的实时性要求,需要在运行过程中快速形成并执行调度决策<sup>[132]</sup>。然而,大规模语言模型的体量与自回归推理往往带来明显的延迟与算力开销,制约其在现场侧的高效应用<sup>[133]</sup>。在这种情况下,如何降低推理延迟并提升计算效率,成为实现其在负荷分配环节中应用的关键<sup>[134]</sup>。为此,

可通过蒸馏、剪枝与量化等压缩技术在较小精度损失下显著降低模型尺寸与推理成本。结合轻量化与指令微调,可在减少计算的同时维持对话与任务能力<sup>[135]</sup>。在架构层面,建议采用边端分工:本地运行蒸馏后的轻量模型以满足快速响应需求,云端保留大模型进行离线评估与校正<sup>[136]</sup>。同时,专用硬件与加速库对时延与能耗优化至关重要。随着现场可编程门阵列(Field-Programmable Gate Array, FPGA)、专用集成电路(Application-Specific Integrated Circuit, ASIC)等边缘平台的进展,Transformer推理可显著提速降耗<sup>[137]</sup>。

在调度策略上,事件触发与分层控制可在保证稳定性的同时降低对上层智能体的调用频率<sup>[138]</sup>。LLM适宜定位于复杂、异常情景的决策支持,上层赋能、下层稳定。当出现负荷突变或设备故障等事件时,由阈值触发LLM生成应急策略,从而减少其参与频率并弱化硬实时要求<sup>[139]</sup>。该做法契合分层、分区控制理念,将LLM作为上层认知-决策大脑而非高速环内控制器。工程实现中可设置触发条件或守护规则,在预测违约和约束失配时再调用LLM。

在本地化部署方面,边远或弱联网微电网需要可在本地执行的高效模型以降低对云端依赖<sup>[140]</sup>。可采用边云协同,将预测与评估等重计算迁移至云端,快速闭环控制留在本地,通过通信完成协同与同步,从而兼顾本地负载与时延优化<sup>[141]</sup>。同时,应重视模型生命周期管理,若在数据特征随时间变化的情况下缺乏持续学习与域适配,模型性能易随时间退化<sup>[142]</sup>。可构建中心聚合,本地更新的持续学习与参数下发流程,并结合联邦学习在不出站点的数据约束下实现跨域协同更新,从而实现现场AI模块的滚动升级与稳定运行<sup>[143]</sup>。

表9给出了低时延与可落地的组合策略,通过模型压缩与硬件加速降低推理延迟;以边云协同在本地闭环、云端重算;采用事件触发、分层控制把LLM置于上层认知层,从而适当降低对实时性的要求;再借助联邦学习处理数据特征随时间变化及隐私保护问题。

表9 实时性与模型高效部署

策略	延迟	资源	隐私	典型场景
蒸馏、量化、剪枝	降低	降低	—	本地推理
GPU、FPGA、ASIC加速	降低	提高	—	边缘侧加速
边云协同	降低	—	提高	本地闭环+云端重算
事件触发、分层控制	降低	—	—	上层LLM异常、复杂场景触发
持续、联邦学习	—	—	提高	适应数据分布变化

现有研究表明,模型压缩与硬件加速能够有效降低推理时延,但压缩后的能力退化并不均匀,复杂推理、长上下文与罕见场景处理能力往往先于常规问答能力下降.因而,在微电网场景中,不宜仅以平均延迟作为部署依据,还应同时评估关键工况下的最坏响应时间与策略质量.相较于“全时在线调用大模型”,事件触发与分层控制更符合当前技术成熟度,可在保证底层稳定控制的前提下降低对上层LLM的实时性压力.

微电网运行涉及从毫秒级物理波动到分钟级能量平衡的多时间尺度特性,这直接决定了LLM在不同控制层级中的应用边界.微电网的运行涉及从瞬时物理平衡到长周期能量优化等多个时间尺度,因此讨论实时性必须落实到具体的控制层级.

对于逆变器本地调节或下垂控制等一次控制环节,由于响应时间需达到毫秒级,受自注意力机制计算密度的限制,目前的LLM难以满足高频闭环控制的实时性与确定性需求.在这一领域,大模型更实际的价值在于离线阶段,例如辅助整定控制参数初值或提炼运行规则.进入秒级的二次协调层,如频率恢复或电压修正,系统对响应速度虽有一定容忍度,但仍极其强调动作的稳定性.此时,部署在边缘侧的轻量化模型可以发挥场景识别与告警分析的作用,通过辅助决策而非直接控制的方式,为底层逻辑提供参考.

相比之下,经济调度、滚动优化等能量管理任务通常处于分钟级尺度,这与LLM目前的处理能力最为契合.在这个层面,大模型能够整合多源异构信息,并与求解器或预测模型配合,在复杂的演化模拟和人机交互中体现出明显优势.综合来看,微电网的LLM部署应当采取一种分层架构:底层快速响应环节仍交由传统控制算法负责,以确保物理层面的安全稳健;高层慢速决策环节则引入大模型,通过这种逻辑上的分工,在不牺牲系统实时性的前提下提升运行的智能化程度.

#### 4.4 安全与隐私保护

作为关键信息基础设施,微电网运行须以安全与隐私为首要原则<sup>[144]</sup>.在接入大语言模型后,应从调度安全、网络安全与数据隐私三个维度系统识别并管理风险<sup>[145]</sup>.在调度安全方面,LLM的概率生成特性可能产生“幻觉”,即输出与事实不符的内容<sup>[146]</sup>.此类错误在聊天语境影响有限,但在调度环节可能转化为运行风险,需要在控制室场景对模型建议进行验证与人机协同.若模型误判系统状态并生成错

误调度方案,应将其拦截在执行链之外.为此,有必要建立面向AI决策的安全围栏<sup>[147]</sup>,如设置双通道校验并在结论不一致时触发人工复核,以及以策略约束与在线监控实现对越界或敏感操作的拦截<sup>[148]</sup>.同时,应采用对齐与安全训练方法,使模型在面对危险请求时具备拒答能力.指令生成式预训练Transformer(Instruct Generative Pre-trained Transformer, InstructGPT)等工作已证明人类反馈强化学习在降低有害输出方面的有效性<sup>[149]</sup>.据此,微电网中的LLM应在高风险场景默认请求人工确认,以人机协作提高可信度.

在网络安全层面,大语言模型面临的主要威胁包括对抗样本攻击与提示注入,这类攻击可能诱导模型生成或执行具有潜在危害的指令<sup>[150]</sup>.为降低风险,系统应强化身份认证、访问权限控制与持续安全监测,并在输入与输出环节设置隔离与校验机制,以防止提示注入与恶意输出的传播.当模型通过API访问业务数据时,宜采用端到端加密和最小化数据暴露策略,避免敏感信息泄露.在具备条件的场景下,可进一步通过网络分区、隔离运行或离线部署的方式,缩减外部攻击面.

在数据隐私保护方面,微电网系统通常涉及用户用电行为及企业运行信息等敏感数据<sup>[151]</sup>.联邦学习可作为一种有效路径,通过仅共享参数或梯度而非原始数据,实现多节点间的协同训练<sup>[152]</sup>.若推理任务在云端执行,应对输入数据进行匿名化或本地预处理,以减少敏感信息上云的风险.对于隐私保护要求更高的应用,可采用本地部署的小型模型,在性能与安全之间取得平衡.

为减少同类错误的复现,可结合离线对齐更新与在线自我校验两类互补策略.离线方面,将经人工确认的错例纳入对齐数据<sup>[30]</sup>,通过基于人类或规则反馈的对齐训练更新参数,使模型在相似情境下趋近更安全的输出<sup>[33]</sup>.在线方面,引入链式思考,对关键中间结论进行可检验化分解并借助外部工具校核,从而抑制幻觉在决策链上的传播<sup>[39]</sup>.同时采用参数高效微调与低成本微调<sup>[91]</sup>,最终通过固定的回归测试与场景化指标对新版本进行核验,确保历史错例不再出现<sup>[36]</sup>.

此外,LLM可能泄露敏感样本<sup>[153]</sup>,解决方式包括数据去标识与差分隐私微调等方法,以抑制模型记忆与泄露能力<sup>[154]</sup>.进一步地,学术界正探索面向大模型的“机器遗忘与选择性遗忘”技术,力求仅保留泛化知识而不暴露个体信息<sup>[155]</sup>.

表10按幻觉、提示注入、数据泄露、合规审计

梳理了对应防护,在推理侧引入检索与物理校核;在训练侧采用差分隐私、联邦学习与选择性遗忘;在运维侧落实日志与常态化演练,形成全生命周期的安全防线。

表10 安全与隐私保护

风险类型	典型场景	主要防护	适用阶段
幻觉、错误结论	生成违规调度建议	检索校验、物理仿真校核、红队测试	推理、集成
提示注入、越权	第三方数据源被投毒	策略隔离、输入清洗	集成、运行
数据泄露	私有负荷、日志外泄	去标识、差分隐私微调、联邦、选择性遗忘	训练、运行
合规与审计缺失	无日志、无复现	全链路日志、版本与审批闭环	全生命周期

在安全与隐私保护方面,现有方法更多体现为“风险缓解”而非“风险消除”。检索校验、物理仿真校核与人工确认可降低幻觉带来的执行风险,但无法从根本上消除生成模型的不确定性;联邦学习、去标识化与差分隐私可降低数据泄露概率,但会带来通信开销、训练不稳定或性能折衷。因此,微电网中的LLM安全治理更适宜采用分层防护与多重校验,而非依赖单一技术手段。

随着调度流程的推进,微电网面临的安全与隐私风险表现出明显的动态演变特征。在策略生成阶段,主要的威胁来自模型幻觉或提示注入攻击,这类干扰可能诱导大模型给出违反运行规程或背离物理约束的指令。当流程进入执行前的校核环节,风险则转向技术实现层面,例如数据采集的非同步性、接口响应异常,以及大模型输出与外部求解器计算结果之间的冲突。而在底层的运行数据管理中,负荷曲线、电源出力等时序信息的高频采集,也显著放大了电力数据的泄露隐患。

这种多维度的风险分布意味着防御手段必须具备针对性。针对策略生成的随意性,需通过严格的规程约束与人工辅助审校进行过滤;而在执行校核环节,则应侧重于物理仿真验证与求解器的可行性检查,并配置相应的异常回退预案以应对误调度风险。对于数据隐私,联邦学习或差分隐私等技术可有效降低敏感信息的暴露概率。从微电网的运行安全出发,防护工作的核心不应仅停留在模型语义层面的审计,而应重点拦截那些具备诱导性但会导致越限运行的策略建议,确保任何输出在转化为物理动作前,均经过了严密的边界验证。

总体来说,只有将安全与隐私保障融入LLM应用的全生命周期,才能真正放心地让其介入微电网调度。在AI安全领域已有很多成熟思路,如沙箱测

试、红队攻击模拟等。将这些应用到电力AI系统,可以提前发现漏洞、补救加固。最终目标是构建一个安全可控、可信可靠的智能微电网调度系统,让大语言模型成为保障系统稳定和用户数据安全的助手,而非新的风险源。

## 5 结论与展望

基于以上综述和案例分析,可以展望大语言模型+微电网这一交叉领域未来的研究方向和面临的挑战。总体而言,未来工作将围绕如何让LLM更加专业、更高效、更可靠地服务于微电网优化展开。以下几点值得重点关注:

(1) 开发专用的大模型和算法框架:当前通用LLM如GPT-5具备强大能力,但未必针对电力任务进行过优化。未来可考虑研发电力能源专用基础模型。如同医疗领域有BioGPT<sup>[156]</sup>、法律领域有LegalBERT<sup>[157]</sup>一样,能源领域也可训练一个EnergyGPT或GridGPT。它在预训练阶段就融合了大量电力系统知识、典型运行数据,从而天然对微电网场景友好。这种模型的训练需要跨学科合作,也需高质量的数据集支持,是一项艰巨工程。同时,算法层面也需创新。例如当前Transformer架构未必对数值运算友好,是否可引入新的模块来增强物理计算能力?又如如何将强化学习更自然地融入LLM训练,使之能在交互环境中边对话边学策略?解决问题将使模型更适应动态优化任务。

(2) 物理与数据双驱动的混合智能:微电网优化不能单靠数据黑箱,物理约束和机制始终重要。未来应发展物理引导的LLM方法,把物理模型、仿真反馈实时融入LLM推理循环。例如构建AI调度员+数字孪生联合体,每次LLM提出方案先经孪生验证再反馈调整。这需要研究如何快速调用仿真供LLM使用,如何在LLM内部嵌入约束求解模块等。此外,多智能体强化学习在微网能量管理中已有应用,未来可考虑LLM与RL结合的新范式。例如用LLM做策略生成器、RL做参数细调,两者相辅相成。又如多个LLM智能体分别代表不同微网或不同角色(发电方、负荷方),通过博弈对话达成全局优化。这些探索有望形成数据+机理+博弈融合的新型智能优化框架。

(3) 评估体系与标准的建立:建立统一的评估体系与标准是推动大语言模型在电力系统中规范化应用的关键环节。目前,该领域尚缺乏系统的性能评价指标与通用测试基准。未来可参考ElecBench等框架,构建覆盖微电网典型运行场景的标准化评测体

系,用以系统比较不同模型与算法的性能差异,从而促进方法改进与结果复现.与此同时,行业标准亦需同步完善,对人工智能参与电网调度的可靠性、人工干预流程及安全冗余机制等提出明确规范.类似于无人驾驶需要遵循的安全法规,智能微电网的调度系统也应明确 AI 决策的法律责任及操作流程.为支撑研究与验证工作,还应发展开放的微电网仿真平台,提供可复现实验环境,支持研究者测试新算法与策略.这些基础性工作将为模型评测、标准建设及成果转化奠定坚实基础,推动该领域从探索走向工程化与规范化.

(4) 跨领域多任务应用:微电网负荷优化只是 LLM 可发挥作用的一个切入点.未来,LLM 在能源领域的应用可扩展至更广泛任务,例如能源系统规划、碳交易策略制定、用户能效咨询等.LLM 的强知识整合能力使其可成为能源系统综合决策的核心 AI.在微电网层面,除了负荷分配,还可涉及保护定值整定、设备运维诊断等方面.一个有趣方向是,LLM 结合机器视觉对微电网现场视频、图像做监控报警分析,比如发现储能电池鼓包、光伏板遮挡等隐患,及时通知维护.这需要融合视觉大模型与语言模型.再如 LLM 助力能源市场交易,自动解读市场规则文本并帮助微电网运营者决策报价策略.总之,可期望未来出现能源 AI 助手,贯穿设计-运营-维护全流程.而在每个子任务中的研究积累,又会反哺提升通用能源大模型的能力.

(5) 社会影响和人才培养:大模型的应用正在改变微电网的运行管理方式.调度岗位对人才的要求也在转变,未来的从业者需要懂电力,也要会使用和理解人工智能.这意味着高校和培训机构应及时更新课程,加入能源与 AI 结合的内容.同时也要看到潜在问题.过度依赖智能系统可能让人员操作能力下降,技术门槛过高又可能扩大地区间的运维差距.应当坚持人机协同,把 AI 作为辅助工具,而不是替代人.在政策上可以建设共享平台,降低中小企业使用大模型的成本,让技术收益更普惠.在隐私保护方面,应提高公众意识,在安全前提下推动数据共享,促进模型优化.只有在这些社会问题得到重视的情况下,微电网智能化才能健康持续地发展.

总体来看,大语言模型为微电网负荷优化提供了新的思路和工具.它能理解和利用非结构化知识,结合人类经验与数据模式,为复杂的能源系统提供更灵活、更智能的优化手段.尽管应用仍处早期阶段,但已有成果验证了其可行性.未来,随着算力和算法的提升,LLM 有望从辅助决策逐步过渡到自主

优化,在安全可控的前提下实现对微电网的持续智能管理,为分布式能源系统的高效运行和“双碳”目标的实现提供支撑.

#### 参考文献 (References)

- [1] 李军祥, 万秋婷, 屈德强, 等. 考虑共享储能的多微电网低碳经济优化研究[J]. *太阳能学报*, 2026, 47(1): 756-764.  
(Li J X, Wan Q T, Qu D Q, et al. Low-carbon and economic optimization of multiple microgrids considering shared energy storage[J]. *Acta Energiæ Solaris Sinica*, 2026, 47(1): 756-764.)
- [2] 林村, 张寿明. 基于混合鲸鱼算法的需求响应微电网最优经济运行[J]. *云南大学学报: 自然科学版*, <https://link.cnki.net/urlid/53.1045.N.20260120.0920.002>.  
(Lin C, Zhang S M. Optimal economic operation of microgrids with demand response based on hybrid whale optimization algorithm[J]. *Journal of Yunnan University: Natural Sciences Edition*, <https://link.cnki.net/urlid/53.1045.N.20260120.0920.002>.)
- [3] 于继文, 刘海鹏, 李耀雯, 等. 基于改进生成对抗网络模仿学习的微电网最优经济运行策略[J]. *现代电力*, DOI: 10.19725/j.cnki.1007-2322.20260224.001.  
(Yu Jiwen, Liu Haipeng, Li Yaowen, et al. Optimal economic operation strategy of microgrid based on imitation learning with improved generative adversarial network[J]. *Modern Electric Power*, DOI: 10.19725/j.cnki.1007-2322.20260224.001.)
- [4] 刘向杰, 陈思安, 黄汉武, 等. 基于近端策略优化的智能微电网能量管理策略[J]. *中国科学: 信息科学*, 2026, 56(4): 937-953.  
(Liu X J, Chen S A, Huang H W, et al. An intelligent energy management strategy for a microgrid based on proximal policy optimization[J]. *Scientia Sinica: Informationis*, 2026, 56(4): 937-953.)
- [5] 田春笋, 祖文静, 李慧璇, 等. 基于 LSTM-CNN-CGAN 的风储微电网日前风险优化调度方法[J]. *电力系统及其自动化学报*, DOI: 10.19635/j.cnki.csu-epsa.001788.  
(Tian C Z, Zu W J, Li H X, et al. Day-ahead risk optimal dispatch method for wind-storage microgrids based on LSTM-CNN-CGAN[J]. *Proceedings of the CSU-EPSSA*, DOI: 10.19635/j.cnki.csu-epsa.001788.)
- [6] Baur L, Ditschuneit K, Schambach M, et al. Explainability and interpretability in electric load forecasting using machine learning techniques – A review[J]. *Energy and AI*, 2024, 16: 100358.
- [7] Wasserkrug S, Boussieux L, Den Hertog D, et al. Enhancing decision making through the integration of large language models and operations research optimization[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39(27): 28643-28650.
- [8] 张浩然, 李君, 邢立宁, 等. 大模型与智能优化算法集成研究综述[J]. *控制与决策*, 2026, 41(4): 871-891.  
(Zhang H R, Li J, Xing L N, et al. A research review on

- integration of large models and intelligent optimization algorithms[J]. *Control and Decision*, 2026, 41(4): 871-891.)
- [9] Chowdhery A, Narang S, Devlin J, et al. PaLM: Scaling language modeling with pathways[J/OL]. 2022, arXiv: 2204.02311.
- [10] Shi H, Fang L R, Chen X Y, et al. Review of the opportunities and challenges to accelerate mass-scale application of smart grids with large-language models[J]. *IET Smart Grid*, 2024, 7(6): 737-759.
- [11] Yao Q, Fang F, Chen Y Y, et al. AI large models for power system: A survey and outlook[J]. *IET Smart Energy Systems*, 2025, 1(1): 3-21.
- [12] Liao W L, Yang Z, Jia M S, et al. Zero-shot load forecasting with large language models[J/OL]. 2024, arXiv: 2411.11350.
- [13] Bonadia R S, Trindade F C L, Freitas W, et al. On the potential of ChatGPT to generate distribution systems for load flow studies using OpenDSS[J]. *IEEE Transactions on Power Systems*, 2023, 38(6): 5965-5968.
- [14] Choi S L, Jain R, Emami P, et al. eGridGPT: Trustworthy AI in the Control Room[R]. Golden: National Renewable Energy Laboratory, 2024.
- [15] Long S F, Tan J J, Mao B M, et al. A survey on intelligent network operations and performance optimization based on large language models[J]. *IEEE Communications Surveys & Tutorials*, 2025, 27(6): 3915-3949.
- [16] Amjad F, Korötko T, Rosin A. Review of LLMs applications in electrical power and energy systems[J]. *IEEE Access*, 2025, 13: 150951-150969.
- [17] 李薇, 乔少杰. 大模型驱动的人工智能: 理论、挑战、应用及展望[J]. *无线电工程*, <https://link.cnki.net/urlid/13.1097.TN.20260212.1433.002>. (Li W, Qiao S J. Large model-driven artificial intelligence: Theory, challenges, applications and prospects[J]. *Radio Engineering*, <https://link.cnki.net/urlid/13.1097.TN.20260212.1433.002>.)
- [18] Xu P, Zhu X T, Clifton D A. Multimodal learning with transformers: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(10): 12113-12132.
- [19] Wen Q S, Zhou T, Zhang C L, et al. Transformers in time series: A survey[J/OL]. 2022, arXiv: 2202.07125.
- [20] Lin T Y, Wang Y X, Liu X Y, et al. A survey of transformers[J]. *AI Open*, 2022, 3: 111-132.
- [21] Dao T, Ermon S, Fu D, et al. Flash attention: Fast and memory-efficient exact attention with IO-awareness[C]. *Advances in Neural Information Processing Systems*. New Orleans: Neural Information Processing Systems Foundation, 2022: 16344-16359.
- [22] Kalyan K S. A survey of GPT-3 family large language models including ChatGPT and GPT-4[J]. *Natural Language Processing Journal*, 2024, 6: 100048.
- [23] Borgeaud S, Buchatskaya E, Cai T, et al. Training compute-optimal large language models[C]. *Advances in Neural Information Processing Systems*. New Orleans: Neural Information Processing Systems Foundation, 2022: 30016-30030.
- [24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017: 30.
- [25] Wu J Y, Gan W S, Chen Z F, et al. Multimodal large language models: A survey[C]. *IEEE International Conference on Big Data*. Sorrento, 2024: 2247-2256.
- [26] Zhao W X, Zhou K, Li J Y, et al. A survey of large language models[J/OL]. 2023, arXiv: 2303.18223.
- [27] Chang T A, Bergen B K. Language model behavior: A comprehensive survey[C]. *Computational Linguistics*. Cambridge, 2024: 293-350.
- [28] Geva M, Caciularu A, Wang K R, et al. Transformer feed-forward layers build predictions by promoting concepts in the vocabulary space[J/OL]. 2022, arXiv: 2203.14680.
- [29] Zhao H Y, Chen H J, Yang F, et al. Explainability for large language models: A survey[J]. *ACM Transactions on Intelligent Systems and Technology*, 2024, 15(2): 1-38.
- [30] Agarwal S, Almeida D, Askell A, et al. Training language models to follow instructions with human feedback[C]. *Advances in Neural Information Processing Systems*. New Orleans: Neural Information Processing Systems Foundation, 2022: 27730-27744.
- [31] Chung H W, Hou L, Longpre S, et al. Scaling instruction-finetuned language models[J/OL]. 2022, arXiv:.
- [32] Kaufmann T, Weng P, Bengs V, et al. A survey of reinforcement learning from human feedback[J/OL]. 2023, arXiv: 2312.14925.
- [33] Bai Y T, Kadavath S, Kundu S, et al. Constitutional AI: Harmlessness from AI feedback[J/OL]. 2022, arXiv: 2212.08073.
- [34] Majumder S, Dong L, Doudi F, et al. Exploring the capabilities and limitations of large language models in the electric energy sector[J]. *Joule*, 2024, 8(6): 1544-1549.
- [35] Gabber H A, Hemied O S. Domain-specific large language model for renewable energy and hydrogen deployment strategies[J]. *Energies*, 2024, 17(23): 6063.
- [36] Chang Y P, Wang X, Wang J D, et al. A survey on evaluation of large language models[J]. *ACM Transactions on Intelligent Systems and Technology*, 2024, 15(3): 1-45.
- [37] Andreassen A, Anil C, Dohan D, et al. Solving quantitative reasoning problems with language models[C]. *Advances in Neural Information Processing Systems*. New Orleans: Neural Information Processing Systems Foundation, 2022: 3843-3857.
- [38] Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models[J/OL]. 2022, arXiv: 2206.07682.

- [39] Bosma M, Chi E, Ichter B, et al. Chain-of-thought prompting elicits reasoning in large language models[C]. *Advances in Neural Information Processing Systems*. New Orleans: Neural Information Processing Systems Foundation, 2022: 24824-24837.
- [40] Ruan J Q, Liang G Q, Zhao H, et al. Applying large language models to power systems: Potential security threats[J]. *IEEE Transactions on Smart Grid*, 2024, 15(3): 3333-3336.
- [41] Zhu X Y, Li J, Liu Y, et al. A survey on model compression for large language models[C]. *Transactions of the Association for Computational Linguistics*. Cambridge, 2024: 1556-1577.
- [42] Lee J, Joung S, Lee K. Scalable optimization approaches for microgrid operation under stochastic islanding and net load[J]. *Applied Energy*, 2024, 374: 124040.
- [43] Eyimaya S E, Altin N. Review of energy management systems in microgrids[J]. *Applied Sciences*, 2024, 14(3): 1249.
- [44] Ali J S, Ayeng'o S P, Kimambo C Z M. A review of control strategies for optimized microgrid operations[J]. *IET Renewable Power Generation*, 2024, 18(14): 2785-2818.
- [45] 葛磊蛟, 范延赫, 来金钢, 等. 面向低碳经济的人工智能赋能微电网优化运行技术[J]. *高电压技术*, 2023, 49(6): 2219-2238.  
(Ge L J, Fan Y H, Lai J G, et al. Artificial intelligence enabled microgrid optimization technology for low carbon economy[J]. *High Voltage Engineering*, 2023, 49(6): 2219-2238.)
- [46] 钱帮远, 郑鹏远, 单婷婷. 基于多样性场景性能的微电网鲁棒经济优化算法[J]. *现代电子技术*, 2026, 49(5): 107-116.  
(Qian B Y, Zheng P Y, Shan T T. Micro-grid robust economic optimization algorithm based on diverse scenario performance[J]. *Modern Electronic Technique*, 2026, 49(5): 107-116.)
- [47] Minh N Q, Linh N D, Khiem N T. A mixed-integer linear programming model for microgrid optimal scheduling considering BESS degradation and RES uncertainty[J]. *Journal of Energy Storage*, 2024, 104: 114663.
- [48] Maya N, Poolla B K, Srinivasan S, et al. A fast dynamic internal predictive power scheduling approach for power management in microgrids[C]. *IEEE PES Innovative Smart Grid Technologies*. Bengaluru, 2025: 1-6.
- [49] 张叶贵, 李光明, 吉畅, 等. 基于改进粒子群算法的微电网能源优化调度方法[J]. *水电与新能源*, 2026, 40(1): 30-35.  
(ZHANG Y G, LI G M, JI C, et al. Optimized energy dispatching method for microgrids based on improved particle swarm optimization algorithm[J]. *Hydropower and New Energy*, 2026, 40(1): 30-35.)
- [50] Maleki N, Lundström O, Musaddiq A, et al. Future energy insights: Time-series and deep learning models for city load forecasting[J]. *Applied Energy*, 2024, 374: 124067.
- [51] 李浩, 李黄强, 向昆, 等. 计及柔性负荷与电动船舶的港口微电网多时间尺度优化调度[J]. *中国电力*, <https://link.cnki.net/urlid/11.3265.tm.20260114.1642.002>.  
(Li H, Li H Q, Xiang K, et al. Multi-time scale optimal dispatch of port microgrid considering flexible load and electric ships[J]. *Electric Power*, <https://link.cnki.net/urlid/11.3265.tm.20260114.1642.002>.)
- [52] Wan A P, Chang Q, AL-Bukhaiti K, et al. Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism[J]. *Energy*, 2023, 282: 128274.
- [53] Tayab U B, Bin Roslan M A, Hwai L J, et al. A review of droop control techniques for microgrid[J]. *Renewable and Sustainable Energy Reviews*, 2017, 76: 717-727.
- [54] Davoudkhani I F, Zare P, Abdelaziz A Y, et al. Robust load-frequency control of islanded urban microgrid using 1PD-3DOF-PID controller including mobile EV energy storage[J]. *Scientific Reports*, 2024, 14: 13962.
- [55] 孙雪雯, 毛帅. 基于智能预测算法的微电网调度优化方法研究[J]. *电气技术与经济*, 2025(12): 24-27.  
(Sun X W, Mao S. Research on microgrid scheduling optimization method based on intelligent forecasting algorithm[J]. *Electrical Equipment and Economy*, 2025(12): 24-27.)
- [56] Kontopoulou V I, Panagopoulos A D, Kakkos I, et al. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks[J]. *Future Internet*, 2023, 15(8): 255.
- [57] Yin C L, Liu K H, Zhang Q J, et al. SARIMA-based medium- and long-term load forecasting[J]. *Strategic Planning for Energy and the Environment*, 2023: 283-306.
- [58] Yan Q, Cao W P, Yan Y, et al. A multi-factor collaborative electricity load forecasting method based on feature importance and multi-scale feature extraction[J]. *Energy and AI*, 2025, 21: 100579.
- [59] Deng Y, Fan J Q, Jiang H, et al. Behavior testing of load forecasting models using BuildChecks[C]. *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*. New York: ACM, 2022: 76-80.
- [60] Dong Q, Huang R B, Cui C H, et al. Short-term electricity-load forecasting by deep learning: A comprehensive survey[J]. *Engineering Applications of Artificial Intelligence*, 2025, 154: 110980.
- [61] Guenoukpati A, Agbessi A P, Salami A A, et al. Hybrid long short-term memory wavelet transform models for short-term electricity load forecasting[J]. *Energies*, 2024, 17(19): 4914.
- [62] Nebey A H. Recent advancement in demand side energy management system for optimal energy utilization[J]. *Energy Reports*, 2024, 11: 5422-5435.

- [63] Koltermann L, Celi Cortés M, Figgenger J, et al. Improved rule-based power distribution algorithm for hybrid battery storage systems and real-world validation[J]. *Journal of Energy Storage*, 2024, 86: 111360.
- [64] Hossain M I, Shafullah M, Abido M A. Battery power control strategy for intermittent renewable energy integrated modular multilevel converter-based high-voltage direct current network[J]. *Sustainability*, 2023, 15(3): 2626.
- [65] Dey B, Misra S, Garcia Marquez F P. Microgrid system energy management with demand response program for clean and economical operation[J]. *Applied Energy*, 2023, 334: 120717.
- [66] Amini M, Nazari M H, Hosseinian S H. Predictive energy management strategy for battery energy storage considering battery degradation cost[J]. *IET Renewable Power Generation*, 2023, 17(5): 1119-1138.
- [67] Huang L, Sun W, Li Q Y, et al. A two-layer energy management for islanded microgrid based on inverse reinforcement learning and distributed ADMM[J]. *Energy*, 2024, 301: 131672.
- [68] 熊珞琳, 毛帅, 唐漾, 等. 基于强化学习的综合能源系统管理综述[J]. *自动化学报*, 2021, 47(10): 2321-2340. (Xiong L L, Mao S, Tang Y, et al. Reinforcement learning based integrated energy system management: A survey[J]. *Acta Automatica Sinica*, 2021, 47(10): 2321-2340.)
- [69] Li Y R, Chang W G, Yang Q. Deep reinforcement learning based hierarchical energy management for virtual power plant with aggregated multiple heterogeneous microgrids[J]. *Applied Energy*, 2025, 382: 125333.
- [70] Wimalaratna L, Akimov A, Ratnasiri S. Demand response by residential and industrial customers: A rapid systematic review[J]. *International Journal of Green Energy*, 2025, 22(2): 307-319.
- [71] Mahela O P, Khosravy M, Gupta N, et al. Comprehensive overview of multi-agent systems for controlling smart grids[J]. *CSEE Journal of Power and Energy Systems*, 2022, 8(1): 115-131.
- [72] Siahchehre Kholerdi S, Ghasemi-Marzbali A. Effect of demand response programs on industrial specific energy consumption: Study at three cement plants[J]. *International Transactions on Electrical Energy Systems*, 2022, 2022: 8550927.
- [73] Malehmirchegini L, Suliman M S, Farzaneh H. Region-wise evaluation of price-based demand response programs in Japan's wholesale electricity market considering microeconomic equilibrium[J]. *IScience*, 2023, 26(7): 106978.
- [74] Bakare M S, Abdulkarim A, Zeeshan M, et al. A comprehensive overview on demand side energy management towards smart grids: Challenges, solutions, and future direction[J]. *Energy Informatics*, 2023, 6(1): 4.
- [75] Khan M A, Sabahat Z, Farooq M S, et al. Optimizing smart home energy management for sustainability using machine learning techniques[J]. *Discover Sustainability*, 2024, 5(1): 430.
- [76] Jaskólski M, Bućko P, Czapp S. The evolution of low- and medium-voltage distribution system development planning procedures and methods—A review[J]. *Energies*, 2025, 18(13): 3461.
- [77] Pang S M, Xu Q S, Yang Y B, et al. Robust decomposition and tracking strategy for demand response enhanced virtual power plants[J]. *Applied Energy*, 2024, 373: 123944.
- [78] Jurjevic R, Zakula T. Demand response in buildings: A comprehensive overview of current trends, approaches, and strategies[J]. *Buildings*, 2023, 13(10): 2663.
- [79] Zhao J, Zhang X, Chen X, et al. Energy scheduling for multi-energy microgrids: A large language model-assisted reinforcement learning approach[Z].
- [80] Sajjadi Mohammadabadi S M, Entezami M, Karimi Moghaddam A, et al. Generative artificial intelligence for distributed learning to enhance smart grid communication[J]. *International Journal of Intelligent Networks*, 2024, 5: 267-274.
- [81] Cheng Y H, Zhao H, Zhou X Y, et al. A large language model for advanced power dispatch[J]. *Scientific Reports*, 2025, 15: 8925.
- [82] Elfari E M, Rasheed A, San O. Artificial intelligence-driven digital twin of a modern house demonstrated in virtual reality[J]. *IEEE Access*, 2023, 11: 35035-35058.
- [83] Zhang J H, Wang W, Guo S Y, et al. Solving general natural-language-description optimization problems with large language models[J/OL]. 2024, arXiv: 2407.07924.
- [84] Zheng Y, Chen Y H, Qian B, et al. A review on edge large language models: Design, execution, and applications[J]. *ACM Computing Surveys*, 2025, 57(8): 1-35.
- [85] Finzi M, Gruver N, Qiu S K, et al. Large language models are zero-shot time series forecasters[C]. *Advances in Neural Information Processing Systems*. New Orleans: Neural Information Processing Systems Foundation, 2023: 19622-19635.
- [86] Jin M, Wang S Y, Ma L T, et al. Time-LLM: Time series forecasting by reprogramming large language models[J/OL]. 2023, arXiv: 2310.01728.
- [87] Ansari A F, Stella L, Turkmen C, et al. Chronos: Learning the language of time series[J/OL]. 2024, arXiv: 2403.07815.
- [88] Zhang X Y, Chowdhury R R, Gupta R K, et al. Large language models for time series: A survey[J/OL]. 2024, arXiv: 2402.01801.
- [89] Das A, Kong W H, Sen R, et al. A decoder-only foundation model for time-series forecasting[J/OL]. 2023, arXiv: 2310.10688.
- [90] Singh A K, Strouse D. Tokenization counts: The impact of tokenization on arithmetic in frontier

- LLMs[J/OL]. 2024, arXiv: 2402.14903.
- [91] Dettmers T, Holtzman A, Pagnoni A, et al. QLoRA: Efficient finetuning of quantized LLMs[C]. *Advances in Neural Information Processing Systems*. New Orleans: Neural Information Processing Systems Foundation, 2023: 10088-10115.
- [92] 高绪冲, 王朝民, 张展泳, 等. 融合混合储能的微电网多时间尺度优化调度策略[J/OL]. *南方电网技术*, <https://link.cnki.net/urlid/44.1643.tk.20260105.1429.016>. (Gao X C, Wang C M, Zhang Z B, et al. Multi-time scale optimal dispatch strategy for microgrids integrated with hybrid energy storage[J]. *Technology of China Southern Power Grid*, <https://link.cnki.net/urlid/44.1643.tk.20260105.1429.016>.)
- [93] Antonesi G, Cioara T, Anghel I, et al. A systematic review of transformers and large language models in the energy sector: Towards agentic digital twins[J]. *Applied Energy*, 2025, 401: 126670.
- [94] Feng Y Y, Hu G S, Li X D, et al. Adapting amidst degradation: Cross domain Li-ion battery health estimation *via* physics-guided test-time training[J/OL]. 2024, arXiv: 2402.00068.
- [95] Sun W J, Wu C K, Xie C D, et al. Fine-tuning enables state of health estimation for lithium-ion batteries via a time series foundation model[J]. *Energy*, 2025, 318: 134177.
- [96] Papaioannou I, Korkas C, Kosmatopoulos E. Smart building recommendations with LLMs: A semantic comparison approach[J]. *Buildings*, 2025, 15(13): 2303.
- [97] Chen C C, Feng X H, Li Y Y, et al. Integration of large language models and federated learning[J]. *Patterns*, 2024, 5(12): 101098.
- [98] Lai R C, Wang J, Tian Y, et al. FedCBE: A federated-learning-based collaborative battery estimation system with non-IID data[J]. *Applied Energy*, 2024, 368: 123534.
- [99] Wu Y B, Tian C L, Li J G, et al. A survey on federated fine-tuning of large language models[J/OL]. 2025, arXiv: 2503.12016.
- [100] Schade C, Egging-Bratseth R. Battery degradation: Impact on economic dispatch[J]. *Energy Storage*, 2024, 6(2): e588.
- [101] Wang F J, Zhai Z, Zhao Z B, et al. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis[J]. *Nature Communications*, 2024, 15: 4332.
- [102] Hu Y K, Zhao T Q, Yue M. From natural language to solver-ready power system optimization: An LLM-assisted, validation-in-the-loop framework[J/OL]. 2025, arXiv: 2508.08147.
- [103] 石东山, 司起步, 易康, 等. 考虑需求响应负荷的微电网两阶段鲁棒优化调度研究[J]. *机车电传动*, DOI: [10.13890/j.issn.1000-128X.2025.06.300](https://doi.org/10.13890/j.issn.1000-128X.2025.06.300). (Shi D S, Si Q B, Yi K, et al. Two-stage robust optimal dispatch of microgrids considering demand response load[J]. *Electric Drive for Locomotives*, DOI: [10.13890/j.issn.1000-128X.2025.06.300](https://doi.org/10.13890/j.issn.1000-128X.2025.06.300).)
- [104] Papaioannou C, Tzitzios I, Papaioannou A, et al. EnergiQ: A prescriptive large language model-driven intelligent platform for interpreting appliance energy consumption patterns[J]. *Sensors*, 2025, 25(16): 4911.
- [105] Zhang Z Y, Cui C G, Yang N, et al. Research on a two-layer demand response framework for electric vehicle users and aggregators based on LLMs[J/OL]. 2025, arXiv: 2505.20877.
- [106] Sanguinetti M, Atzori M. Conversational agents for energy awareness and efficiency: A survey[J]. *Electronics*, 2024, 13(2): 401.
- [107] Matharaarachchi A, Mendis W, Randunu K, et al. Optimizing generative AI chatbots for net-zero emissions energy internet-of-things infrastructure[J]. *Energies*, 2024, 17(8): 1935.
- [108] Campi R, Giudici M, Pinciroli Vago N O, et al. Enhancing human-AI collaboration through a conversational agent for energy efficiency[J]. *Proceedings of the AAAI Symposium Series*, 2025, 5(1): 52-55.
- [109] Voyez A, Allard T, Avoine G, et al. The privacy cost of fine-grained electrical consumption data[J]. *Scientific Reports*, 2025, 15: 17391.
- [110] Soares J, Lezama F, Faia R, et al. Review on fairness in local energy systems[J]. *Applied Energy*, 2024, 374(C): 123933.
- [111] Xiong B Y, Zhang L L, Hu Y, et al. Deep reinforcement learning for optimal microgrid energy management with renewable energy and electric vehicle integration[J]. *Applied Soft Computing*, 2025, 176: 113180.
- [112] Shadi M R, Mirshekali H, Shaker H R. Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects[J]. *Renewable and Sustainable Energy Reviews*, 2025, 216: 115668.
- [113] Choi S L, Jain R, Feng C, et al. Generative AI for power grid operations[R]. Golden: National Renewable Energy Laboratory, 2024.
- [114] Palikhe A, Wang Z C, Yin Z P, et al. Towards transparent AI: A survey on explainable language models[J/OL]. 2025, arXiv: 2509.21631.
- [115] Benes K J, Porterfield J E, Yang C. AI for energy: Opportunities for a modern grid and clean energy economy[R]. Washington: US Department of Energy, 2024.
- [116] Fantozzi P, Naldi M. The explainability of transformers: Current status and directions[J]. *Computers*, 2024, 13(4): 92.
- [117] Wang X Z, Wei J, Schuurmans D, et al. Self-consistency improves chain of thought reasoning in language models[J/OL]. 2022, arXiv: 2203.11171.
- [118] Rajabi E, Etmiani K. Knowledge-graph-based

- explainable AI: A systematic review[J]. *Journal of Information Science*, 2024, 50(4): 1019-1029.
- [119] Li C, Wang B. A knowledge graph method towards power system fault diagnosis and classification[J]. *Electronics*, 2023, 12(23): 4808.
- [120] Chen Q, Li Q, Wu J Y, et al. Application of knowledge graph in power system fault diagnosis and disposal: A critical review and perspectives[J]. *Frontiers in Energy Research*, 2022, 10: 988280.
- [121] Kovari A. AI for decision support: Balancing accuracy, transparency, and trust across sectors[J]. *Information*, 2024, 15(11): 725.
- [122] Yuan Y X, Dehghanpour K, Wang Z Y, et al. Multisource data fusion outage location in distribution systems *via* probabilistic graphical models[J]. *IEEE Transactions on Smart Grid*, 2022, 13(2): 1357-1371.
- [123] Caffagni D, Cocchi F, Barsellotti L, et al. The revolution of multimodal large language models: A survey[C]. Findings of the Association for Computational Linguistics ACL. Bangkok, 2024: 13590-13618.
- [124] Gao J F, Li C Y, Liu H T, et al. LLaVA-med: Training a large language-and-vision assistant for biomedicine in one day[C]. Advances in Neural Information Processing Systems. New Orleans: Neural Information Processing Systems Foundation, 2023: 28541-28564.
- [125] Tang H, Zhang C, Jin M Y, et al. Time series forecasting with LLMs: Understanding and enhancing model capabilities[J]. *ACM SIGKDD Explorations Newsletter*, 2025, 26(2): 109-118.
- [126] Cao D F, Jia F R, Arik S O, et al. TEMPO: Prompt-based generative pre-trained transformer for time series forecasting[J/OL]. 2023, arXiv: 2310.04948.
- [127] Yao S, Zhao J, Yu D, et al. React: Synergizing reasoning and acting in language models[C]. International Conference on Learning Representations. Kigali, 2023: 2210.03629.
- [128] Cancedda N, Dessi R, Dwivedi-Yu J, et al. Toolformer: Language models can teach themselves to use tools[C]. Advances in Neural Information Processing Systems. New Orleans: Neural Information Processing Systems Foundation, 2023: 68539-68551.
- [129] Gao L Y, Madaan A, Zhou S Y, et al. PAL: Program-aided language models[J/OL]. 2022, arXiv: 2211.10435.
- [130] Zhao Z Y, Ding X Y, Prakash B A. PINNsFormer: A transformer-based framework for physics-informed neural networks[J/OL]. 2023, arXiv: 2307.11833.
- [131] Song J, Fanals-Batlloori J, Marín L, et al. A short-circuit calculation solver for power systems with power electronics converters[J]. *International Journal of Electrical Power & Energy Systems*, 2024, 157: 109839.
- [132] Ojo K E, Saha A K, Srivastava V M. Microgrids' control strategies and real-time monitoring systems: A comprehensive review[J]. *Energies*, 2025, 18(13): 3576.
- [133] Zhou Z X, Ning X F, Hong K, et al. A survey on efficient inference for large language models[J/OL]. 2024, arXiv: 2404.14294.
- [134] Lin C F, Han Z H, Zhang C, et al. Parrot: Efficient serving of LLM-based applications with semantic variable[J/OL]. 2024, arXiv: 2405.19888.
- [135] Wan Z W, Wang X, Liu C, et al. Efficient large language models: A survey[J/OL]. 2023, arXiv: 2312.03863.
- [136] Zhai X Y, Peng Y F, Guo X P. Edge-cloud collaboration for low-latency, low-carbon, and cost-efficient operations[J]. *Computers and Electrical Engineering*, 2024, 120: 109758.
- [137] Kang B J, Lee H I, Yoon S K, et al. A survey of FPGA and ASIC designs for transformer inference acceleration and optimization[J]. *Journal of Systems Architecture*, 2024, 155: 103247.
- [138] Li S J, Oshnoei A, Blaabjerg F, et al. Hierarchical control for microgrids: A survey on classical and machine learning-based methods[J]. *Sustainability*, 2023, 15(11): 8952.
- [139] Zhang J C, Sun B, Zhao D D. A novel event-triggered secondary control strategy for distributed generalized droop control in microgrid considering time delay[J]. *IEEE Transactions on Power Electronics*, 2023, 38(5): 5963-5978.
- [140] Xu J J, Li Z Y, Chen W, et al. On-device language models: A comprehensive review[J/OL]. 2024, arXiv: 2409.00088.
- [141] Yıldırım F, Yalman Y, Bayındır K Ç, et al. Comprehensive review of edge computing for power systems: State of the art, architecture, and applications[J]. *Applied Sciences*, 2025, 15(8): 4592.
- [142] Shi H Z, Xu Z H, Wang H Y, et al. Continual learning of large language models: A comprehensive survey[J]. *ACM Computing Surveys*, 2026, 58(5): 1-42.
- [143] Yang X, Yu H, Gao X, et al. Federated continual learning *via* knowledge fusion: A survey[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(8): 3832-3850.
- [144] Stouffer K, Falco J, Scarfone K. Guide to industrial control systems (ICS) security[J]. *NIST Special Publication*, 2011, 800(82): 16-16.
- [145] Yao Y F, Duan J H, Xu K D, et al. A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly[J]. *High-Confidence Computing*, 2024, 4(2): 100211.
- [146] Huang L, Yu W J, Ma W T, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[J]. *ACM Transactions on Information Systems*, 2025, 43(2): 1-55.
- [147] Dong Y, Mu R H, Zhang Y H, et al. Safeguarding large language models: A survey[J/OL]. 2024, arXiv: 2406.02622.

- [148] Alto V. Building LLM powered applications: Create intelligent apps and agents with large language models[M]. **Birmingham**: Packt Publishing Ltd, 2024.
- [149] Lee J S. InstructPatentGPT: Training patent language models to follow instructions with human feedback[J]. *Artificial Intelligence and Law*, 2025, 33(3): 739-782.
- [150] Greshake K, Abdelnabi S, Mishra S, et al. Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection[C]. Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security. New York: ACM, 2023: 79-90.
- [151] Bibi H, Abolhasan M, Lipman J, et al. A comprehensive survey on privacy-preserving technologies for Smart Grids[J]. *Computers and Electrical Engineering*, 2025, 124: 110371.
- [152] Cheng X, Li C D, Liu X F. A review of federated learning in energy systems[C]. IEEE/IAS Industrial and Commercial Power System Asia. Shanghai, 2022: 2089-2095.
- [153] Abad Rocamora E, Cevher V, Chen Y H, et al. Membership inference attacks against large vision-language models[C]. Advances in Neural Information Processing Systems. Vancouver: Neural Information Processing Systems Foundation, 2024: 98645-98674.
- [154] Yu D, Naik S, Backurs A, et al. Differentially private fine-tuning of language models[J/OL]. 2021, arXiv: 2110.06500.
- [155] Blanco-Justicia A, Jebreel N, Manzanares-Salor B, et al. Digital forgetting in large language models: A survey of unlearning methods[J]. *Artificial Intelligence Review*, 2025, 58(3): 90.
- [156] Luo R Q, Sun L A, Xia Y C, et al. BioGPT: Generative pre-trained transformer for biomedical text generation and mining[J]. *Briefings in Bioinformatics*, 2022, 23(6): bbac409.
- [157] Zheng L, Guha N, Anderson B R, et al. When does pretraining help? Assessing self-supervised learning for law and the CaseHOLD dataset of 53, 000+ legal holdings[C]. Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law. New York: ACM, 2021: 159-168.

### 作者简介

毛志忠 (1961-), 男, 教授, 博士, 博士生导师, 主要研究方向为复杂工业过程智能建模、控制与优化, E-mail: [maozhizhong@mail.neu.edu.cn](mailto:maozhizhong@mail.neu.edu.cn);

刘文程 (1990-), 男, 博士生, 主要研究方向为人工智能、虚拟电厂, E-mail: [ncepu\\_liuwencheng@126.com](mailto:ncepu_liuwencheng@126.com).

### 科研团队简介

毛志忠教授团队隶属东北大学流程工业综合自动化全国重点实验室, 长期与谢植及张殿华等教授合作, 从事对电炉炼钢、连铸及连轧过程进行建模、控制与优化方面的研究. 承担国家重点研发计划项目 1 项, 国家 863 计划项目 3 项, 国家自然科学基金重点项目 2 项. 目前团队正探讨利用大模型实现电炉智能炼钢的技术.

毛志忠教授重点对电炉炼钢过程的建模与优化进行深入研究, 对电炉炼钢的供电曲线优化、炉况判断技术、造渣模型及终点预报技术进行理论探讨和应用研究, 研究成果在全国各大钢铁集团 130 多台炼钢电炉或精炼炉上得到应用. 在国内外杂志及学术会议上发表论文 300 多篇, 获省部级科技进步一等奖 1 项、二等奖 2 项.