

电力系统大模型的关键科学问题、挑战与展望

杜博骏¹, 钟非池², 侯庆春^{2*}, 贾宏阳¹, 张宁¹, 康重庆¹

(1. 清华大学电机工程与应用电子技术系, 北京市 海淀区 100084;

2. 浙江大学, 浙江省 杭州市 310058)

Key Scientific Issues, Challenges and Prospects of Power System Foundation Models

DU Bojun¹, ZHONG Feichi², HOU Qingchun^{2*}, JIA Hongyang¹, ZHANG Ning¹, KANG Chongqing¹

(1. Department of Electrical Engineering, Tsinghua University, Haidian District, Beijing 100084, China;

2. Zhejiang University, Hangzhou 310058, Zhejiang Province, China)

ABSTRACT: Driven by the “dual carbon” goals and energy security imperatives, China is accelerating the construction of the new power system. This transition urgently necessitates the integration of intelligent methodologies to enhance system flexibility, security, and economic efficiency. The AI revolution, exemplified by foundation models, holds immense potential to revolutionize traditional analysis and decision-making paradigms. However, the core scientific problems in applying foundation models to the power domain require systematic investigation. Also, application paradigms remain to be clarified, and the overall research framework has yet to be established. Thus, this paper first elucidates the necessity of introducing foundation model technologies into the new-type power systems and systematically analyzes two key scientific problems. Then, from the perspectives of general large models and domain-specific foundation models, it explores the overall architecture and prospective application paradigms in power systems. Based on these technological pathways, an evolutionary road-map is proposed to advance from single-model intelligence toward multi-agent collaborative intelligence. Furthermore, a suite of domain-specific base models is established to support core power-system functions, including forecasting, operational planning, stability analysis and control, and panoramic simulation. On this basis, the paper analyzes the major challenges confronting large-model research for power systems and proposes a prospective research framework and key technological directions for future development.

KEY WORDS: new-type power system; artificial intelligence; general large model; domain-specific foundation model

摘要: 在“双碳”目标和能源安全问题的背景下,我国正加

速构建新型电力系统,亟需引入智能化手段提升系统的灵活性、安全性和经济性。以大模型为代表的人工智能技术革命具有颠覆传统分析与决策范式的巨大潜力,但构建电力系统大模型的核心科学问题亟需系统梳理、应用形态仍待清晰、研究框架尚未明确。因此,该文首先阐明新型电力系统的构建中引入大模型技术的必要性,系统分析构建电力系统大模型亟待解决的两个关键科学问题;进而,从通用大模型和专有基座大模型两条技术路线出发,探讨电力系统大模型的总体架构与应用形态;基于此,提出由“单模型智能”向“多智能体协同”的通用大模型演进路径,并构建面向电力预测、运行规划、稳定分析与控制及全景仿真等核心任务的专有基座大模型体系;在此基础上,分析电力系统大模型研究所面临的主要挑战,并针对性地提出未来电力系统大模型的研究框架与关键技术。

关键词: 新型电力系统; 人工智能; 电力系统通用大模型; 电力系统专有基座大模型

0 引言

当下,全球能源安全形势日趋严峻、气候变化与环境污染问题突出,推动清洁能源转型已成为实现全球可持续发展的关键任务^[1]。中国提出双碳目标,将绿色低碳转型上升为国家战略^[2]。根据国家发展和改革委员会、国家能源局联合发布的《关于促进新能源消纳和调控的指导意见》,到2030年,中国将基本建立协同高效的多层次新能源消纳调控体系,新增用电量需求将主要由新增新能源发电满足^[3]。截至2025年9月,全国风电和太阳能发电装机容量已达17.08亿kW,占全国发电装机的46%;风光发电量合计达1.73万亿kW·h,占同期全社会用电量的22.3%^[3]。随着新能源渗透率的快速提升和电力电子设备的高速增长,电力系

基金项目: 国家重点研发计划项目(2022YFB2403300)。

National Key R&D Program of China (2022YFB2403300).

统面临平衡消纳压力增加、安全稳定问题加剧等挑战，迫切需要构建以新能源为主体的新型电力系统^[4]。

与此同时，以大语言模型为代表的通用人工智能技术正引领数字技术革命和产业变革的新方向，成为推动产业升级和社会变革的核心动力^[5]，我国也将人工智能上升为国家战略^[6]。2025 年 8 月，国务院发布了《关于深入实施“人工智能+”行动的意见》，明确提出到 2027 年率先实现人工智能与六大重点领域广泛深度融合，新一代智能终端、智能体等应用普及率超过 70%^[7]。在政策鼓励和规范引导下，我国人工智能大模型相关创新成果加速涌现。截至 2025 年 7 月，我国已发布 1509 个大模型，数量位居全球首位；人工智能企业数量超过 5100 家，约占全球 15%^[8]。在此背景下，电力系统作为关系国计民生的关键基础设施，正成为人工智能技术落地的重要场景之一。随着新能源占比不断提高，传统依赖人工经验的调度与控制模式已难以应对日益增加的复杂性与不确定性。因此，我国在政策层面持续推进“人工智能+”赋能新型电力系统建设^[9]，发展能够适应复杂电力系统场景的智能化技术成为迫切需求。

然而，现阶段的人工智能方法在复杂电力系统任务中仍存在明显局限。一方面，传统的小规模深度学习模型往往依赖于特定场景下的有限样本，缺乏跨任务、跨场景的泛化能力，难以适应电力系统高维耦合与动态演化的特征^[10-11]；另一方面，这类模型的可解释性弱、可迁移性差、可靠性不足，限制了其在电网核心业务中的实际落地^[12]。大模型范式的兴起为电力系统智能化提供了全新路径。相较于传统的机器学习和深度学习技术，大模型改变了模型的训练与应用模式，将依赖人工标注数据的监督学习变革为利用海量无标签数据的自监督学习，再通过少量标注数据进行微调。这一变革显著提升了训练数据利用效率与模型参数规模，实现了模型精度和泛化能力的跃升^[13-16]，并展现出了涌现能力^[17-18]。

顺应这一趋势，面向电力系统的垂类大模型研究正处于由初步探索走向体系化构建的关键阶段。本文所探讨的电力系统大模型是指以海量多源电力数据为基础，通过超大规模自监督预训练技术、或对通用大模型进行领域适配后训练，构建具有庞大参数规模的人工智能模型，其核心特征在于能够

学习电力系统物理规律与运行知识，并借助微调、提示工程等轻量化适配机制，高效迁移至不同场景和多种下游应用任务。具体而言，电力系统大模型主要包括两类：一类是基于通用大模型（如 DeepSeek、ChatGPT 等大语言模型），通过领域知识注入后训练、电力智能体构建等方式，构建能够理解与处理专业问题的电力系统通用大模型；另一类则是完全基于电力系统运行数据进行预训练，从而构建能够深刻刻画电力系统内在机理的电力系统专有基座大模型。这两类大模型是面向不同需求、各具优势的互补性技术路线。电力系统通用大模型长于处理自然语言交互与任务流程自动化，扮演智能交互中枢角色；专有基座大模型则专注于内化物理规律、保障数值计算精确性与可靠性，充当专业分析引擎。二者将共同支撑 AI for Science (AI4S) 在电力系统领域的落地应用。

针对第一类电力系统通用大模型，现有研究主要探索了大语言模型的生成与模拟能力在电力系统的应用。表 1 总结了主要的研究成果，包括最优潮流的直接求解、电力系统任务助手等领域。在电力任务直接求解方面，有研究尝试使用大语言模型进行潮流求解^[41-42]、电力预测^[19-25]、负荷管理^[26]、网络安全防护^[27]等。然而大语言模型缺少对复杂电力问题机理的深入理解，也缺少基于电力数据的大规模预训练。因此在这类问题中，大语言模型相对于传统端到端小模型并未展现出显著优势。虽然不适合直接求解复杂电力任务，但大语言模型具有强大的自然语言分析能力和代码编写能力，可以作为助手实现智能客服^[33-34]高效处理电力文档、解答电力专业知识，进一步可以建模复杂电力问题并调用专用软件进行求解，例如最优潮流等^[36]；并且通过大语言模型可以实现人机交互，例如进行交互式负荷预测^[28]和主动配电网管理^[32]等任务。此外，随着多模态技术的发展，有研究展望了多模态大语言模型在山火识别、设备检修等方面的应用潜力^[46-47]。

针对第二类电力系统专有基座大模型，现有研究主要探索了电力负荷预测和潮流求解领域。例如，文献[48]基于多层级、不同主体的用电数据进行预训练，所得模型可以用于各类负荷预测任务、用户行为分析和用电异常辨识等；文献[49]展望了电力潮流基座大模型的构建思路，并讨论了其在潮流计算、应急分析和最优潮流等方面的应用前景。

表 1 电力系统大模型现有研究

Table 1 Existing research on large model for power systems			
细分领域	文献	年份	应用目标
电力预测	[19]	2024	短期负荷预测
	[20]	2025	建筑级短期负荷预测
	[21]	2024	高渗透率负荷预测
	[22]	2025	分布式光伏预测
	[23]	2024	风速预测
	[24]	2024	电力需求预测
	[25]	2024	中长期负荷预测
电力系统控制与调度	[26]	2025	建筑用能管理
	[27]	2024	智能逆变器网络攻击检测
	[28]	2025	交互式负荷预测
	[29]	2025	用户驱动的调度策略
	[30]	2025	电动汽车充电行为模拟
	[31]	2025	电网异常行为检测与纠正
	[32]	2025	配电网能源管理
电力系统智能助手	[33]	2024	智能客服、电力巡检等
	[34]	2024	电力巡检
	[35]	2025	科研自动化助手
	[36]	2025	电力系统潮流仿真
	[37]	2025	电力系统安全规章判别
	[38]	2023	电力系统安全教育
	[39]	2023	大模型辅助解决电力问题
电网模拟与潮流计算	[40]	2025	电力预测与潮流计算辅助
	[41]	2024	最优潮流求解、电动汽车调度等
	[42]	2025	最优潮流求解
	[43]	2025	电网算例生成
	[44]	2024	机组组合初始解生成
	[45]	2025	市场用户行为模拟

然而，这一类研究主要面临物理机理嵌入不足与高质量预训练数据集缺失等瓶颈。尽管现有的物理信息神经网络(physics-informed neural network, PINN)尝试将电力系统的代数-微分方程嵌入损失函数以增强物理一致性^[50]，但在向大规模系统扩展时面临显著的技术瓶颈^[11]。1) 高维非凸优化问题导致收敛困难，现有研究多局限于小系统，在大规模电网场景下模型难以收敛^[51-52]；2) 处理含慢动态系统时易出现梯度消失，难以捕捉长周期的系统演化规律^[53]；3) 连续可微假设限制了对继电保护、机组启停等离散事件与混合整数变量的有效处理^[54-55]；4) 缺乏严格的误差界与鲁棒性理论支撑，损失函数权重设定依赖启发式调整，物理可解释性与可靠性不足^[12,56]。

目前电力系统大模型的发展仍面临诸多关键问题。如图 1 所示，本文 1 节首先分析电力系统大模型研发所面临的关键科学问题，解答“为什么需要电力系统大模型”；2 和 3 节分别从电力系统通用大模型和电力系统专有基座大模型两个角度，解答“电力系统需要怎样的大模型”及“构建这样的大模型难在哪里”；4 节进一步讨论未来研究框架与关键技术，解答“如何构建电力系统大模型”；5 节进行总结和展望。

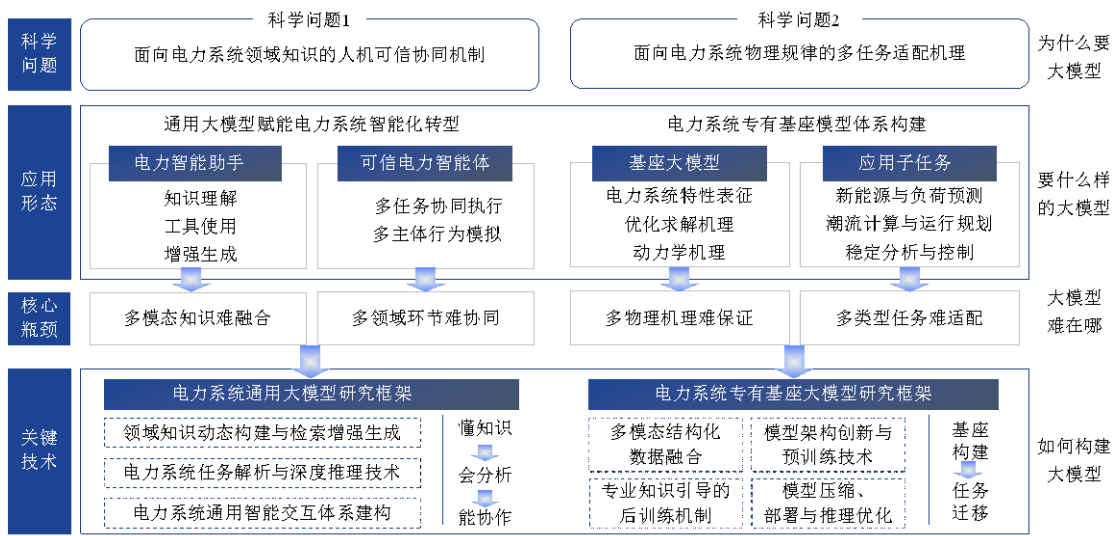


图 1 电力系统大模型应用形态与研究框架展望

Fig. 1 Outlook for application forms and research framework of power system foundation models

1 关键科学问题

1.1 面向电力系统领域知识的人机可信协同机制

随着新型电力系统建设的深入，在面对海量异构数据、强时效要求、区域差异性和权限依赖性时，

传统依赖人工的规程理解和执行流程的运行模式在操作效率和人力资源方面遇到瓶颈。以大型语言模型为基础的通用人工智能技术，为破解这一难题提供了新范式。其核心在于如何让人工智能像人类

专家一样理解并“活用”庞大的领域知识，并与人类形成高效、可信的协同关系。而实现这一目标，亟需解决以下核心瓶颈：多模态知识难融合、多领域任务难协同。

具体地，在领域规程和知识的统一表征方面，电力系统积累了海量的技术规程、运行规范、事故案例、研究文献以及设备参数等非结构化知识。这些知识来源广泛、格式各异、语境复杂，形成了巨大的知识鸿沟。当前，大模型对电力专业术语、复杂逻辑关系和隐含约束的理解仍存在偏差，难以形成精确且可推理的内部表征。因此，亟需研究一种统一范式，将分散、隐性的领域知识重构为大模型可理解、可推理的结构化知识体系，奠定人工智能在电力领域深度应用的认知基石。

在面向复杂任务的人机可信协同方面，对于调度运行、故障处置、规划评审等敏感场景，完全依赖人工智能黑箱决策存在较大风险。大模型需要进一步从“知识库”升级为能够执行复杂任务、并与人类专家协同工作的“智能体”。这要求大模型不仅能准确解析自然语言指令，还能自主调用外部工具，同时具备自校验与自修正功能，最终将决策过程与结果以可解释的方式输出给人类专家。如何构建这种人机可信的协同决策机制，是实现人工智能在电力系统中落地应用的关键。

1.2 面向电力系统物理规律的多任务适配机理

随着电力系统新能源渗透率的增加和电力电子化提高，其运行与控制呈现出强不确定性、多时间尺度耦合、强非线性、高维异构等复杂特征。传统以精确物理模型和数值迭代算法为核心的分析方法，在面对超大规模计算、实时决策响应和高维不确定性场景时，面临维度灾难与速度瓶颈；而现有端到端深度学习方法在可解释性、泛化性等方面存在明显不足。以基座大模型为代表的新一代人工智能技术，为突破这一困境提供了新路径，其核心在于如何让机器从海量系统运行数据中“学习”并“掌握”底层物理规律，并作为一个通用基座灵活适配多样化的下游应用子任务。为实现这一目标，亟需解决以下核心瓶颈：多物理机理难保证、多类型任务难适配。

具体地，在底层机理提取和物理一致性保证方面，电力系统的运行规划基于潮流方程和大规模优化问题的求解，动态过程由微分-代数方程组描述。当前，数据驱动模型的参数规模较少、训练样本量

不足，难以有效提取底层机理以保证端到端求解的物理一致性。亟需探索能够融合物理机理与数据驱动的新型建模范式，构建兼具泛化能力、物理一致性与计算可扩展性的电力系统专有基座大模型，为实现下一代电力系统智能计算奠定理论基础。

在面向电力系统核心任务的高效适配方面，针对稳定评估、实时调度、故障预警等对准确性与时效性要求极高的场景，一个通用的预训练模型需要被快速、精准地定制以适应特定任务的需求。这要求基座大模型不仅能通过少量样本或提示指令快速激发出其在特定任务上的优异性能，还能保证其在预训练数据所含场景外的泛化能力和鲁棒性。如何设计高效的微调方法与提示机制，实现基座大模型灵活迁移和规模化应用，是释放基座大模型在电力系统核心业务中巨大潜力的关键。

2 通用大模型赋能电力系统智能化转型及挑战

如图2所示，本节提出了未来电力系统通用大

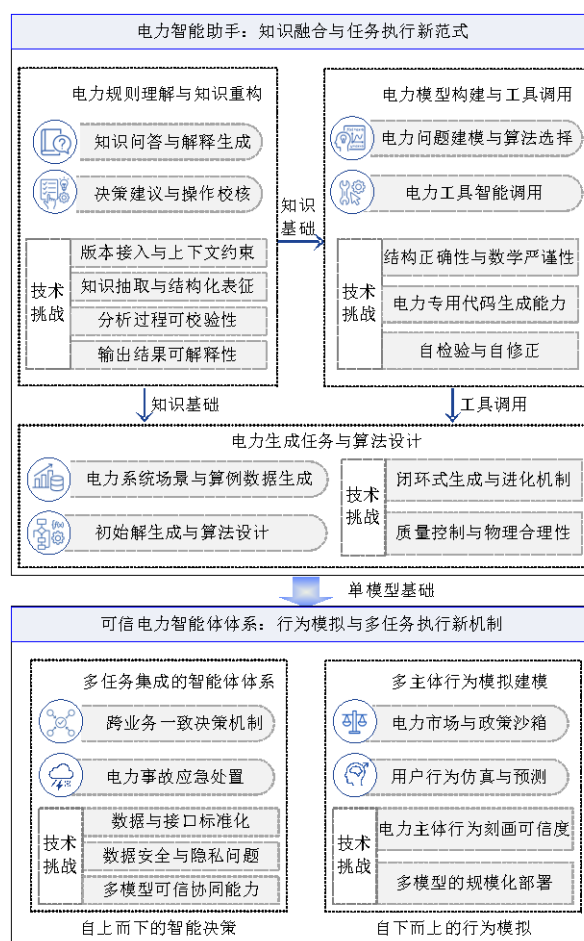


图2 电力系统通用大模型应用形态展望

Fig. 2 Outlook for application forms of general large models in power systems

模型的应用蓝图,并分析所面临的技术挑战。2.1节聚焦于单个大模型能力的纵向提升,强调如何使大模型从“能回答问题”走向“能参与决策流程”。其中,2.1.1节围绕电力规则理解与知识重构,让大模型“懂电力”;在此基础上,2.1.2节讨论优化建模与算法设计,要求大模型不仅具备领域知识,还能根据问题智能选择解决方法与工具,实现“会用工具”;2.1.3节面向生成类和启发式任务,强调大模型在“懂知识、会用工具”的基础上发挥快速生成与探索的优势,让大模型“会创新”。2.2节从单个大模型走向体系化协作,其中2.2.1节提出多任务集成的智能体体系,自上而下地使多个具备专业分工的大模型协同工作;2.2.2节面向多主体行为模拟建模,支撑市场机制检验、需求响应评估等新型分析范式。电力系统通用大模型将从“回答者”演进为“协调者”和“组织者”,重塑电力系统的知识获取、决策生成和执行协同方式,有望成为新型电力系统运行范式变革的智能中枢。

2.1 电力智能助手:知识融合与任务执行新范式

2.1.1 电力规则理解与知识重构

电力系统的运行调度、规划决策与市场交易依赖于运行规程、安全约束、并网标准、故障处置预案以及市场交易与出清规则的复杂知识体系。这些知识具有非结构化、分散性强等特性。现有的检索与决策支持手段仍主要依赖关键词匹配或人工编码规则,缺乏对约束条件、适用范围、流程依赖关系的语义理解能力和深度分析能力^[57-58]。这导致在调度、运维和市场等场景对操作人员的专业要求极高。面向这一问题,基于大语言模型的电力智能助手有望通过对运行规程、市场规则与操作边界的统一建模与语义重构,实现对电力系统知识的理解,从而在问答咨询、文档生成、行为合规性审查与操作方案推演等任务中发挥核心作用。

通过电力规则理解与知识重构,通用大模型可提供两个层次的智能化。1)在知识问答与解释生成层面,模型可对调度规程、安全约束、市场准入条件、并网技术要求等问题进行结构化分析,提供结论性判断,并对结果合规性进行溯源判定;进而自动生成符合行业口径的文档,显著降低文字性工作负担。2)在决策建议与操作校核层面,模型可以对拟执行的运行策略、控制命令或应急处置步骤进行“规程一致性”和“权限边界”校验,即在输出调控建议的同时给出该建议是否需

要调度许可、是否触及保护定值或 $N-1$ 安全边界、是否满足并网与调度分工的程序性要求,从而在建议层面起到前置审查、而非事后审核的作用。这使得大语言模型不再是单纯的描述型系统,而能在操作建议进入执行链条之前提供一层可解释、可追溯的合规审视。

电力规则理解与知识重构仍面临多方面挑战。首先,电力规程和市场规则等具有强时效性和区域差异性,模型需要在知识接入上具备动态更新与上下文约束能力;其次,规程性知识往往存在大量隐含前提和默示责任边界,亟需发展相应知识抽取与结构化表征方法;再次,电力系统操作的安全责任属性要求模型不仅能够引用规则,还要在推理和生成过程中显式暴露其依据、假设和适用范围,避免“幻觉”。未来研究需要构建可维护、可动态更新的电力领域知识表示体系,将安全边界、流程约束和审批条件显式纳入模型的推理过程,使该类语言智能助手能够以可解释、可监管的方式嵌入电力系统的日常运行与辅助决策流程。

2.1.2 电力模型构建与工具调用

随着新能源和电力电子设备渗透率的增加,电力系统决策建模需要考虑更加复杂的约束,通常需要经验丰富的工程师或研究者完成模型搭建、约束处理及算法选择,人工依赖强、工程成本高、复用性差^[36]。大语言模型提供了新的建模范式,通过将任务目标、运行边界条件和系统约束以自然语言的形式描述给模型,驱动大模型自动解析问题结构,建立相应的数学模型,并进一步调用仿真分析工具与求解器完成计算和结果分析,实现复杂场景下快速建模、智能求解和高效验证。

在模型构建与工具调用方面,电力系统通用大模型可支持两个层次的智能化应用模式。1)在模型构建层面,大模型可智能解析非格式化输入数据,并根据用户的自然语言描述自动抽取关键要素,基于问题场景和用户需求智能选择建模方式,形成标准数学规划形式,进一步生成可执行的求解脚本。2)在求解流程集成与工具调用层面,大模型能够根据建模结果自动调用适配的外部电力分析工具,并对不同建模方式与配置的输出结果进行对比分析。通过这一模式,大模型将显著降低电力系统任务的工程门槛与研究周期。

然而,基于大模型的智能建模与工具调用仍存在多方面挑战。首先,生成模型的结构正确性与数

学严谨性难以保证，特别是在处理复杂约束时，容易产生形式上完整但物理上不成立或数学上不可解的表达式；其次，现有大模型在电力专业建模语言的代码生成能力有限，语法错误与逻辑偏差频发，严重影响生成代码的可用性；最后，模型构建与代码生成必须具备可验证性与可纠错性，大模型应能自动生成测试算例、边界工况和可行性检验步骤；一旦检验未通过，大模型需要基于测试算例结果反馈进行修正。未来研究需重点围绕建模过程的数学验证、领域专用代码生成能力提升及自检验与自修正3个方向展开。

2.1.3 电力生成任务与算法设计

电力系统中存在大量生成型任务，用于支撑后续的求解与评估流程，其典型输出包括极端运行场景、测试算例、初始可行解及剪枝与割平面策略等。通过引入领域知识并具备对求解器等外部工具的调用能力，电力系统通用大模型能够在这些任务中发挥高效探索与生成能力^[43]。

针对生成型或探索型任务，电力系统通用大模型可以在两个方面提供关键支撑。1) 数据生成方面，大模型用于构造典型极端运行场景和算例集合，包括高比例新能源出力波动场景、极端负荷场景等，从而为后续的控制策略评估、韧性分析和应急预案推演提供针对性更强的测试集。2) 算法设计方面，大模型可以根据电力系统任务特性、系统约束及历史运行信息，自动生成可行的初始调度方案、候选机组启停组合，或提出可能的割平面集合与启发式算法。这些生成结果可作为优化求解的启动点或剪枝策略，用于缩小搜索空间、加快收敛速度。此类用法不要求大模型直接求出最终优化解，而是将其作为高质量起点和搜索指导信号，替代传统依赖人工经验和手工调参的启发式设计过程，从而显著提升电力系统问题求解的计算效率。

然而，构建面向电力生成任务与算法设计的智能体仍面临若干挑战。首先，大模型可能生成统计上合理但物理上不可行或工程上无意义的场景、解或规则；其次，更为深层的挑战在于探索效率与多样性的平衡，如何在避免陷入局部重复生成的同时，有效引导模型探索高价值、低概率的候选解或策略空间。未来研究需要围绕物理引导的探索策略自适应调控方法、“生成-验证-修正”的闭环式生成机制，发挥大模型的高效探索与可靠生成价值。

2.2 电力智能体体系：多任务执行与行为模拟新机制

2.2.1 多任务集成的智能体体系构建

电力系统运行具有强流程性、强耦合性和强时效性，调度、稳定控制、市场出清等任务往往并行发生并相互制约。这一过程目前仍高度依赖人工介入和多部门协同，表现为信息分散、响应链条长、决策一致性难以保证，特别是在高比例新能源并网、运行状态快速变化和异常工况高频出现的背景下，上述模式的局限愈加突出^[59-60]。在此背景下，基于电力系统通用大模型构建的智能体体系有潜力围绕具体运行目标自动完成任务解析、流程编排、实时监视、策略建议与结果解释，从而将传统的不同人员依次操作多个系统、转变为智能体统一协调多个系统的协同决策范式。

多任务集成智能体的应用前景主要体现在两个方向。1) 跨业务场景的一致性决策代理：智能体可在同一逻辑框架下同时接入运行控制、市场交易与资源协调等多任务场景，生成具备物理可行性、经济一致性和合规边界说明的联合策略。2) 高速应急处置：在大扰动或极端工况下，智能体可与实时数字孪生系统联动，自动识别故障影响域与连锁失稳风险，生成多套应急控制路径，实现不同环节的联动，面向秒级和分钟级的处置窗口提供策略候选与辅助决策。

多任务集成的智能体体系的构建面临多方面挑战。1) 外部工具协同与接口标准化问题。智能体的有效性依赖于其稳定、准确地调用潮流计算、优化求解、稳定分析、市场出清等异构求解器的能力，这要求建立统一的数据与约束描述格式、标准化的任务编排与调用协议。2) 数据安全与隐私合规问题。调度指令、状态量测、市场交易信息等关键数据高度敏感，跨系统共享与在线推理必须满足分级管控、最小可见范围和合规留痕的要求，这对智能体的记忆管理、上下文保留方式及多区域协同模式提出了额外约束。未来的研究可围绕可信交互架构与多智能体自主协同决策机制展开，推动电力系统多任务集成智能体向安全、可靠与高效等方向发展。

2.2.2 多主体行为模拟建模

随着新型电力系统的新能源渗透率逐步提升、新型负荷不断增长、电力市场日益发展，系统运行越来越受到市场主体的响应策略所影响。传统仿真

方法对市场主体行为的静态建模难以真实再现其在价格波动、激励政策、极端扰动等场景下的异质反应，难以捕捉真实市场中参与者因风险偏好、信息不对称等因素导致的复杂甚至非理性的决策行为^[45,61]。针对该问题可引入以大模型为认知内核的多主体智能体，为用户、负荷聚合商、零售商、发电商等不同主体分别构建具备偏好刻画、风险认知与记忆机制的行为体，生成可解释的行动决策轨迹，并在多轮交互中不断更新自身策略。多主体行为模拟不只是对“未来会发生什么”的预测工具，而是面向“制度是否稳健、策略是否可行、响应是否可用”的前置验证平台，这将成为市场规则设计、源荷互动调度以及高冲击场景应急预案评估的关键支撑能力。

多主体行为模拟的应用前景主要体现在两个方向。1) 市场模拟与政策沙箱，通过在统一仿真环境中引入发电企业、售电主体、储能运营商、需求响应等多个具备自主偏好的智能体，可用于评估不同出清机制、价格信号、补偿/惩罚策略以及需求响应激励方案将如何被各类参与者实际采纳，市场主体间的互动和决策将如何演化，从而提前识别可能的投机行为、价格极端波动、资源错配或者在极端工况下的非协同反应。2) 用户行为预测，构建具备异质偏好、动态决策特征的用户智能体，可用于刻画电力用户对价格、气象等外部因素的响应规律。进一步可通过将大模型与强化学习或博弈学习机制结合，使用户智能体具备长期学习与策略演化

能力，以模拟新型负荷形态(如电动汽车、数据中心、可调工业负荷)在未来电力市场中的行为。该方向不仅有助于提高负荷预测与需求响应评估的精度，也为用户侧能量管理、分布式灵活性聚合与新型商业模式设计提供了重要支撑。

多主体行为模拟建模仍存在两类突出挑战。1) 行为刻画的可信度，智能体都需要表现出“可解释、可追溯、可验证”的决策逻辑，这要求引入历史运行数据、典型应急场景记录以及人为干预痕迹对智能体进行持续校准。2) 规模化与合规问题，大规模并行智能体推演会带来显著的计算与通信开销，同时调度记录、实时量测、市场交易信息等数据高度敏感，要求在跨主体协同建模中同时满足最小披露、分级可见与留痕审计。未来的关键方向，是将行为智能体从“模拟用户个体”提升为“可与调度、市场、规划模型协同”的标准化行为模块，使人在回路成为系统仿真和运行规划流程中的重要支撑。

3 面向电力系统核心任务的专有基座大模型体系构建及挑战

如图 3 所示，面向新型电力系统的复杂性与不确定性，本节以“感知-优化-评估-集成”的层次化逻辑，构建面向未来电力专有基座大模型的系统蓝图：1) 预测基座大模型作为感知层，提供负荷、新能源与电价等多要素、多尺度的预测结果；2) 运行规划基座大模型在上述预测与系统约束下，求解机组组合等时序优化，输出可执行的计划与控制设

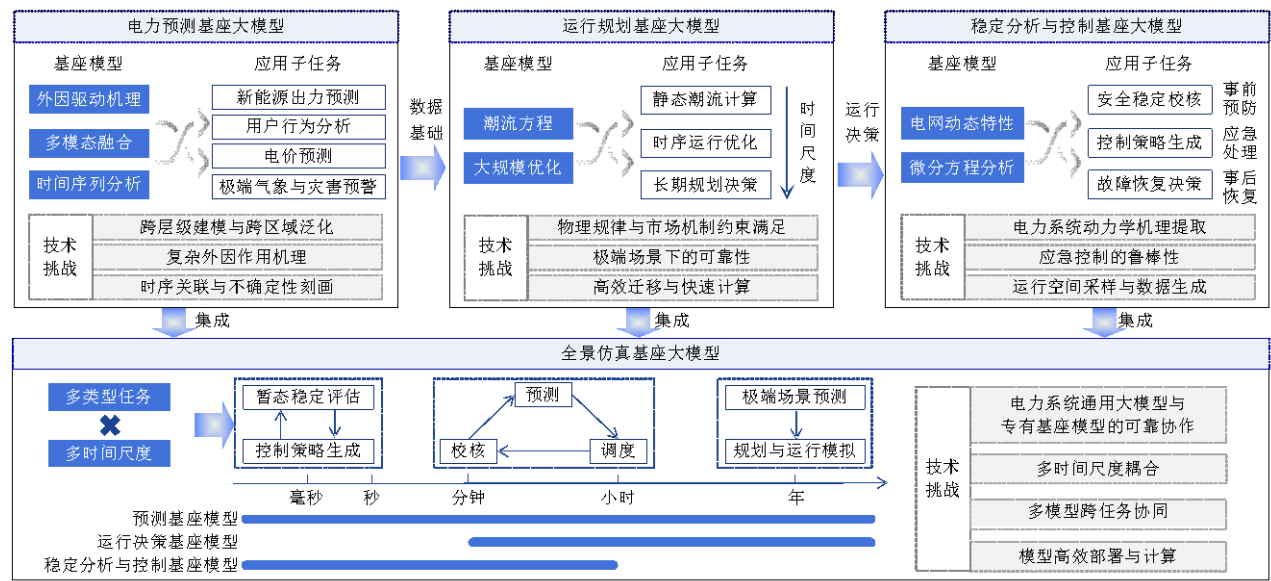


图 3 电力系统专有基座大模型应用形态展望

Fig. 3 Outlook for application forms of domain-specific foundation models in power systems

定；3) 稳定分析与控制基座大模型从暂态稳定分析角度校核决策方案的安全可行性，并生成扰动场景下的快速控制策略，将稳定裕度与安全约束反馈至决策层；4) 全景仿真基座大模型在统一表示与时间轴上耦合预测-决策-评估，实现电力系统运行的端到端闭环推演。由此，四者之间形成自洽的因果链与数据流，最终实现从碎片化小模型到跨域统一基座的跃迁，并完成由传统解析建模到人工智能范式的深层转化。

3.1 电力预测基座大模型

预测是电力系统运行与规划的关键环节，但现有方法通常以单一要素或固定尺度为对象，依赖局部历史样本，难以在不同要素、不同尺度间建立统一的表示与约束传递，从而在复杂耦合和极端扰动时暴露出泛化与鲁棒性不足。未来，预测专有基座大模型以超大规模预训练为基础，学习跨要素、跨尺度、跨任务、跨模态的通用表征。进一步通过小样本有监督微调，为不同区域电网与业务场景提供即插即用的高质量预测能力。电力预测基座大模型的定位是电力基座大模型体系的感知层，为运行规划、稳定控制与全景仿真提供高精度的预测输入，实现从“局部经验驱动”迈向“全域知识统一”的关键技术跃迁。

预测专有基座大模型的主要应用场景包括新能源出力预测、负荷预测及需求响应能力评估、电力市场电价预测及极端气象与灾害预警。在新能源出力预测方面，基座大模型结合电力时序与外部观测数据，综合利用气象雷达、卫星云图、社会活动强度等多源异构信号，精准预测新能源出力概率分布，为系统灵活性评估和源网协调运行提供多时间尺度先验信息。在负荷预测及需求响应能力评估方面，基座大模型在统一框架下提取不同层级、不同类型的用户用能行为特征，实现负荷曲线与需求响应资源的协同预测，为精细化负荷管理提供数据支持。在电力市场电价预测方面，基座大模型利用多模态信息融合能力，将负荷与新能源预测结果与燃料价格、碳价、购售电合同、交易行为特征及政策和市场规则文本等外生变量共同纳入时间序列建模，提升日前、日内及实时电价预测的精度。在极端气象与灾害预警方面，基座大模型深度融合灾害监测和气象信息等多源异构数据，显著增强对极端高温低温、台风洪涝等事件的动态感知能力，实现对极端气象及其对新能源出力、负荷、电价冲击的

鲁棒预测。依托多任务泛化与快速迁移机制，上述各类应用可在同一预测基座上通过少样本微调实现跨区域、跨季节、跨业务的高效适配，展现出区别于传统“小模型”的系统性优势。

预测专有基座大模型的构建仍面临一系列关键挑战。首先，不同区域和电网的数据分布差异显著，数据覆盖有限且标注稀缺，如何在数据不均衡条件下实现跨域泛化仍有待突破；其次，天气、经济与社会行为等外生变量对预测结果具有重要影响，但其时空特征复杂多变，如何实现与电力时序数据的高效融合仍是难点；再次，预测结果必须满足从用户侧到系统侧的层级一致性并符合物理约束；最后，在新能源高比例渗透以及极端气候事件等场景下，模型需要具备足够的鲁棒性，避免预测偏差引发系统性风险。针对上述问题，未来研究亟需在自监督预训练、跨层级图结构建模与物理机理嵌入等方面持续探索，从而推动预测基座大模型由单任务驱动走向全局化、由局部应用走向多任务与可迁移化。

3.2 运行规划基座大模型

运行规划是电力系统的重要任务，其核心是以潮流方程为约束，解决包括最优潮流、经济调度、机组组合在内的一系列大规模优化问题。现有人工智能方法已在潮流近似计算^[62-63]、运行约束提取^[64-65]或优化求解加速^[66-67]等方面取得进展，但在物理约束满足、不确定性建模能力、多任务泛化和跨系统适应性等方面仍存在较大不足^[68-70]。未来，运行规划专有基座大模型有望通过超大规模预训练与物理约束嵌入，学习电力系统潮流方程和优化求解的底层机理，统一支撑从单时间断面潮流计算到跨时段的复杂调度决策与长尺度的运行模拟。其定位是电力基座大模型体系的决策引擎，能快速近似传统优化器的求解过程，在不确定性条件下生成面向多场景多任务的最优决策方案。

通过小样本的有监督微调，运行规划基座大模型的下游应用可分为3个层次。1) 在静态优化层面，能够高效替代传统交流最优潮流求解器，实现安全约束下的快速最优潮流计算，并为大规模灵敏度分析和风险评估提供近似解。2) 在时序优化层面，它能够扩展到机组组合、储能调度与需求响应的跨时段优化，通过捕捉时序耦合与物理约束，融合电碳市场机制，支撑日前、日内及实时调度的统一建模，实现经济、安全、低碳多目标一体化决策。

3) 在规划与运行模拟层面, 运行规划基座大模型可以作为统一的优化内核, 与长期电源规划、网架扩建、碳约束情景分析等相结合, 支撑长时间尺度、高时空分辨率的电力系统运行模拟。

运行规划基座大模型的构建仍面临诸多挑战。首先, 电力系统优化问题具有高度非凸和规模庞大的特性, 如何在保证解的物理约束满足和优化质量的是亟需解决的难题。现有 PINN 方法因依赖软约束惩罚而缺乏最坏情况保障, 机组启停等离散事件导致的不可微性, 更使基于梯度的训练机制面临失效风险。其次, 跨时段优化涉及状态延续、启停约束和不确定性传播等复杂因素, 如何在基座大模型中刻画时序关联与源荷不确定性仍需进一步突破。运行规划结果直接关系系统安全, 模型必须具备防范预测误差与优化偏差所引发运行风险的能力。再次, 电力市场引入了经济与博弈因素, 因此需要在模型中统一表达物理规律与市场机制。最后, 针对不同网架拓扑和系统运行状态, 实现模型的高效迁移与快速计算是提升基座大模型普适性与实时性的主要方向。未来研究需在电力潮流与时序优化的物理机理嵌入、跨任务自监督预训练以及决策风险量化等方面持续探索, 以推动运行规划基座大模型走向实用化。

3.3 稳定分析与控制基座大模型

稳定分析与控制是电力系统安全运行的基石, 其核心在于刻画由扰动引起的暂态过程, 并保证系统在受到冲击后能够迅速恢复至稳定状态^[71]。传统方法依赖微分方程的解析建模与数值积分, 通过李雅普诺夫法或特征值分析实现稳定性辨识, 或采用时域仿真计算动态响应。尽管精度较高, 但在大规模系统和复杂扰动场景下计算开销巨大, 且对参数测量依赖较强^[72-73]。现有基于人工智能的小模型通常聚焦于特定系统的单一稳定问题进行评估或特定控制策略的学习^[74-75], 缺乏对动力学方程底层机理的表达能力, 难以实现跨区域、跨扰动类型的统一建模与泛化分析。未来, 稳定分析与控制专基座大模型有望通过超大规模预训练与物理约束嵌入, 直接学习电力系统的微分方程表示, 从频率响应、功角动态、振荡波形、阻尼特性等多维动力学特征中提取可泛化的稳定性表征。该模型不仅能够快速仿真系统暂态过程, 还可在多时间尺度上捕捉惯量响应、一次调频、二次调频及虚拟同步机控制等动态机制。其定位是电力基座大模型体系的“动

力学引擎”, 在保障系统稳定性与韧性方面发挥核心作用, 为未来自主、智能、可信的电力系统控制奠定理论与技术基础。

稳定分析与控制基座大模型主要有3类下游应用场景。1) 在暂态稳定评估方面, 能够快速预测故障扰动下系统的动态响应, 评估频率跌落、电压崩溃和失步解列等风险, 提高传统数值积分方法的计算精度和效率。2) 对于在线控制与调节, 可以结合广域测量与控制信号, 依据系统运行状态与扰动特征生成控制动作建议。通过引入自适应参数调整机制与强化学习, 模型能够在扰动发生后实现毫秒级的快速闭环控制, 支撑自动电压调节、虚拟同步控制、频率响应和二次调频等多控制层级的协调优化。3) 在安全稳定校核与运行支撑方面, 该模型能够嵌入到运行规划与调控决策环节中, 为电网运行提供满足安全稳定物理机理的校核依据与前瞻性调整策略。通过这些应用, 基座大模型能够将稳定分析从“事后评估”拓展到“事前预测”与“预防性控制”。

稳定分析与控制基座大模型的构建面临诸多挑战。首先, 电力系统动力学由复杂高维的非线性微分-代数方程描述, 因此学习同时具备物理一致性与跨场景泛化能力的动力学模型极具挑战。电力系统动态过程兼具刚性特征与离散事件, 导致传统 PINN 方法面临梯度病态与连续可微假设的双重限制。其次, 暂态过程的时间尺度跨度大, 从毫秒级电磁暂态到数十秒至数分钟的频率与电压恢复过程, 如何在单一模型中同时捕捉不同时间尺度的动力学特性是核心难题之一。最后, 暂态仿真时间较长, 数据往往稀缺且难以覆盖所有极端扰动场景, 如何通过仿真实验、数据生成及不确定性建模来扩展训练样本空间, 仍是工程化落地的瓶颈。未来研究需在电力系统动态特性的物理机理嵌入、多时间尺度的动力学特征提取和电力系统暂态仿真数据生成等方向开展深入探索, 以推动稳定分析与控制基座大模型从理论构想走向工程应用。

3.4 全景仿真基座大模型

全景仿真是电力系统数字化与智能化发展的高级阶段, 其目标在于通过统一的模型与算力平台, 对系统的源-网-荷-储全链条过程进行端到端推演, 从而在一个框架下同时支撑预测、运行规划与稳定控制。传统方法往往依赖分模块的解析建模与专用仿真工具, 各模块之间存在接口割裂、时间尺

度不一致以及数据交互受限等问题,难以满足新型电力系统对全局一致性和实时性的要求^[76-77]。全景仿真专有基座大模型有望通过集成预测、运行规划与稳定控制3个子基座大模型,在统一表示空间中学习电力系统跨时间尺度的物理机理与多任务关联,实现从静态规划到动态运行的全链条一体化仿真。其定位是电力基座大模型体系的“综合操作平台”,是迈向电力系统通用人工智能的重要形态。

全景仿真基座大模型的应用主要体现在3个方面。1)在全链条一致性分析中,能够在同一仿真框架下统一考虑预测误差、调度优化与暂态响应,实现从日前计划到实时运行的闭环推演,提升系统安全裕度与运行透明度。2)在电网数字孪生应用中,该模型可以构建接近真实的电网数字镜像,实现对不同扰动、极端气候事件的快速推演,支撑风险防控与应急演练。3)在跨尺度综合优化方面,能够在同一模型内处理长周期的电源规划、中期的电网扩展、短期的调度优化以及秒级的暂态控制,实现电力系统规划-运行-控制的全景式衔接。这些功能表明,全景仿真基座大模型不仅是对单一任务的强化,更是对电力系统复杂机理的整体抽象与智能化再现,为构建新型电力系统的智能化体系奠定基础。

然而,全景仿真基座大模型的构建与落地仍面临4大挑战。1)通用大模型与专有基座大模型的可信协作机制尚不完善。如何将通用大模型的语义解析和知识推理能力、与专有基座大模型的计算优势深度融合,使通用模型能够理解电力语义、调用专有模型并对结果进行分析解释,是迈向通用智能体系的关键。2)跨尺度建模面临严重的数学异构性与系统刚性挑战。电力系统毫秒级暂态过程本质为高维刚性微分代数方程,而中长期规划则属于大规模混合整数优化,两者数学形式差异巨大。如何在单一基座中设计特定架构,以兼容这种计算范式的异构性并解决跨尺度难题,是实现统一建模的核心难点。3)多模型的跨任务协同尚缺乏机制支撑。在同时考虑预测、调度、控制等任务时,如何维持全局一致性是推动多任务智能协同的关键。4)基座大模型的实时响应性能面临严峻挑战,电力系统的稳定与控制任务则要求毫秒级的决策与指令生成,如何实现模型轻量化与高效计算的仍是工程瓶颈。未来研究需在通用大模型与基座大模型可信交互、跨时间尺度建模、多任务协同机制与算力支撑

体系等方面持续探索,以推动全景仿真基座大模型真正成为电力系统通用人工智能的重要载体。

4 未来研究框架与关键技术

4.1 高质量数据构建与标准化评价体系

电力系统大模型的构建与落地应用面临着数据获取困境、样本极度失衡与评测标准缺失的复合挑战。首先,在数据隐私与获取层面,真实电力系统的运行数据属于关键基础设施信息,导致研究人员难以获取高质量的真实数据进行模型训练^[78]。其次,在样本分布层面,电网绝大部分时间处于稳定运行状态,极端天气或级联故障等高风险场景极其稀缺^[79]。这种严重的样本失衡使得直接基于真实数据训练的模型难以捕捉高影响、低概率事件的特征,限制了其应对极端事故的泛化能力。最后,在测试标准层面,行业内尚缺乏公认的标准化基准与评价体系。现有的通用大模型评测框架尚未充分覆盖电力领域独特的业务需求与专业知识,而针对电力图数据等面向人工智能模型评价的公开基准数据集也相对匮乏且没有形成行业共识^[80],大多数研究仍采用私有或自建数据集,这导致不同技术路线的模型难以在物理一致性、安全性与逻辑性等关键维度上进行公平的对比与迭代优化。

针对电力系统通用大模型的数据集构建,应聚焦于“知识补齐-推理增强-工具调用”的三维能力跃迁。在知识补齐层面,需构建覆盖海量行业标准、学术文献、设备手册与事故预案的领域语料库,以修正通用模型在电力专业术语上的偏差与知识幻觉;在推理增强层面,开发基于人工标注的高质量指令微调与强化学习数据集,通过构造复杂业务意图、思维链推理和专家级决策的对齐样本,激活模型对调度规则的深度理解与任务拆解能力;在工具调用层面,亟需构建自然语言与专业分析软件交互的API映射数据集,赋予模型精准调用外部求解器以解决复杂计算问题的能力。

针对电力系统专有基座大模型的数据集构建,需突破真实数据稀缺与样本分布不均衡的瓶颈,建立高可靠仿真与生成式人工智能合成的数据增强范式。一方面,通过设计高效的采样策略,驱动仿真平台生成覆盖多时空尺度、关键运行方式与典型故障路径的样本,以最小数据量揭示最全面的系统动态;另一方面,则需利用物理约束下的生成式模型,合成在分布与机理上均具有高可信度的虚拟样

本。所有合成数据必须经过严格的物理一致性校验,以确保模型能够从海量数据中准确内化电网的底层物理规律。

针对电力系统通用大模型的标准测试集与评价体系构建,应深度融合领域知识的核心诉求。标准测试集的构建需侧重于模型对专业知识的深度理解与复杂任务的逻辑推理能力,可设计为由多层次专业问答库、代码生成任务及专业文档处理基准共同构成的综合评价基准。其中,问答库应系统覆盖电力系统基础理论、运行规程与典型故障案例;代码任务需要模型生成电力系统专用工具软件的脚本及调用指令;文档基准则用于检验其对技术标准、调度报告等非结构化文本的解析与可控生成能力。在评价指标方面,应系统性地引入多维评价标准,包括专业知识准确率,以量化其知识掌握程度;逻辑安全性与物理一致性,评价输出结果对领域规则与物理定律的遵从度;生成代码的功能正确性与执行效率;推理过程的可解释性与逻辑链条的完备性;以及面向实际部署的响应速度与计算资源消耗情况,从而全面、量化地评价模型在专业场景下的可靠性及工程可行性。

针对电力系统专有基座大模型的标准测试集与评价体系构建,应系统涵盖数值精度、扰动鲁棒性、极端场景可靠性、工况泛化能力及响应速度等关键维度。在应用场景方面,应全面覆盖超短期至短期的新能源出力预测、电价与负荷预测、安全约束机组组合、最优潮流计算与电网规划等运行优化

任务及暂态功角稳定、电压/频率稳定、小干扰振荡分析和紧急控制策略等稳定控制任务。在测试数据构建方面,需基于高保真仿真生成从常规运行到极端扰动、大干扰乃至级联故障的全场景时序数据,并引入高影响、低概率的极端样本。此外,测试集还应包含由继电保护动作、开关分合等离散事件触发的混合动态序列,以检验模型对复杂系统混合动态行为的建模能力。在评价指标体系方面,除数值精度指标外,还需重点考察模型在少样本条件下的快速泛化能力、面向实时应用的推理速度及在未见工况与噪声数据下的鲁棒性,从而系统保障模型在物理一致性、数值精确性与工程可用性方面的综合性能。

4.2 通用大模型:面向电力认知与智能交互的研究框架与关键技术

如图4所示,基于3节所提电力系统通用大模型面临的关键挑战,本节提出针对性的未来研究框架与解决方案,聚焦电力认知与智能交互的3类核心能力:知识增强、深度推理与智能体交互。三者构成了电力系统大模型能力演进的递进链条:知识增强解决“知识从何而来”,为推理提供领域知识支撑;深度推理解决“复杂问题如何分析”,提升模型对复杂电力系统问题理解与分析能力;智能体交互体系解决“人机如何协同”,构建人-机与多智能体间可信交互、决策协同的通用智能体系。它们共同奠定了电力系统大模型从语言理解向系统智能跃迁的技术基础,为实现可解释、可验证、可协

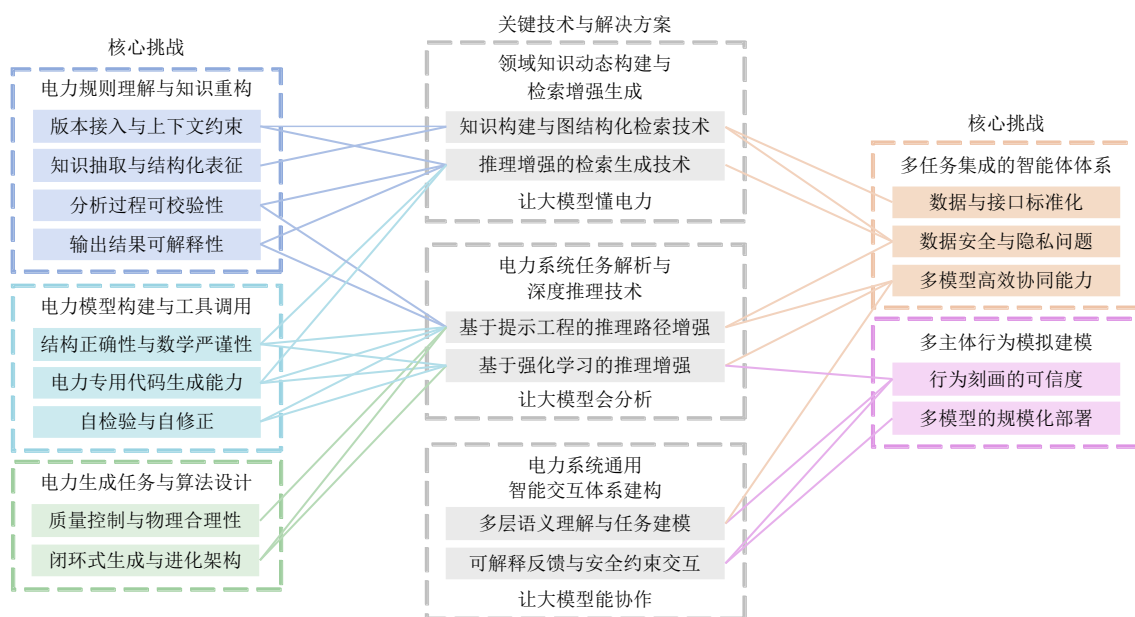


图4 电力系统通用大模型研究框架与关键技术

Fig. 4 Research framework and key technologies of general large models for power systems

同的电力认知智能提供系统化支撑。

4.2.1 领域知识动态构建与检索增强生成

电力系统知识与数据具有高度的专业性、时效性和隐私敏感性。通用大模型虽然具备强大的语言与推理能力,但受限于预训练的静态语料,其对电力领域的知识更新与分析能力存在天然不足^[80];同时,电力行业的数据受隐私与安全限制难以全部加入模型预训练。为此,引入检索增强生成技术(retrieval-augmented generation, RAG)^[81],通过“外部知识动态检索+内部语义生成”的协同机制,使模型能够在生成阶段集成领域知识库和本地数据,从而实现知识的专业扩展、动态更新、可信引用与受控访问。

RAG框架在电力系统中的实现主要包括两条关键路径。知识构建与图结构化检索技术面向电力系统的多源异构数据特征,形成跨层级、跨模态的电力知识图谱。在检索阶段,通过引入图结构检索机制,将知识组织为节点与边的形式,使模型能够沿语义或逻辑路径进行关联查询,从而显著增强对复杂任务的因果关联挖掘与跨模态信息检索能力^[82]。

推理增强的检索生成技术针对复杂电力任务中存在的逻辑推理与约束判断需求,将RAG与链式推理、工具增强及符号推理相结合,形成RAG-Reasoning一体化框架^[83-85]。该机制可进一步结合自适应RAG与迭代RAG等策略^[84-86],基于推理动态更新检索结果,同时根据检索结果增强专业知识推理能力。

4.2.2 电力系统任务解析与深度推理技术

电力系统的运行与控制对逻辑严谨性、过程可解释性与结果高可靠性具有极高要求。通用大模型固有的端到端生成范式,使其内部推理过程难以被审视与验证^[87],这构成了其在电力关键任务中应用的核心障碍。为此,可引入推理增强与思维链技术,提升模型在电力系统复杂场景下的逻辑一致性、决策透明度与指令遵从能力。

两种互补的推理增强路径能显著提升大模型在电力系统决策中的适用性。其一为基于提示工程的显式推理路径增强,该路径通过结构化的提示工程,引导模型生成逻辑连贯的推理思维链,从而将其隐式的“思考”过程外化为可溯源、可验证、可干预的逻辑链条,赋予模型进行深度分析、多步规划与约束满足的能力^[88]。进一步,通过引入树状或图结构的推理拓扑,模型能够系统性地探索多条并

行的推理路径,从而在多分枝、多目标的复杂决策空间中提升决策质量与鲁棒性^[89-90]。该技术路线为解决需要严谨程序和清晰逻辑的电力系统分析与规划类任务提供重要技术支撑。

其二为基于强化学习的推理增强技术,旨在推动大模型由被动的静态推理器演化为能够与环境交互、自主优化决策策略的闭环智能体^[91-92]。核心思想是将大模型视作智能体的认知核心,通过强化学习机制在仿真环境中实现“感知-决策-反馈-优化”的循环迭代。典型方法包括基于人类反馈的强化学习(reinforcement learning from human feedback, RLHF)^[93]与直接偏好优化^[94],可用于对齐模型生成结果与物理可行性及经济性等系统目标。奖励函数可作用于中间思维链条或决策过程,从而在复杂约束条件下实现多步逻辑验证与策略修正^[95-98]。

4.2.3 电力系统通用智能交互体系建构

传统的人机交互方式多以规则化指令或固定界面为主,缺乏语义理解、上下文记忆与推理反馈能力,难以支撑多目标、多场景下的智能决策。构建通用智能交互体系的目标在于使大模型能够具备语义理解、意图识别、可解释反馈与安全可控对话能力,从而成为电力系统的智能助手与决策代理。该体系不仅要求模型理解调度、运行与市场任务背后的逻辑关系,还需在语义空间内实现对运行状态、风险告警与控制指令的解释与生成,最终形成“人-机-系统”一体化的智能交互范式。

构建电力系统的智能交互体系可沿两条技术路径展开。多层语义理解与任务建模的重点在于将自然语言理解与结构化任务建模相结合,使模型能够在语义层面解析操作指令、运行报告与市场策略等多源文本。代表方法包括基于指令调优的任务语义抽取^[99]、基于记忆增强的上下文追踪^[100]及基于混合提示的多轮任务协同机制^[101]。通过语义解析与状态映射,模型能够实现从自然语言到控制变量、运行指标和优化约束的映射,为交互式决策提供可执行的结构化理解。

可解释反馈与安全约束交互旨在提升大模型在决策闭环中的可信度与可控性^[102]。具体而言,首先对模型的推理路径和控制输出进行结构化追踪,实现从生成结果到推理依据的可验证映射^[103]。在此基础上,可进一步发展强化学习驱动的交互优化框架与人在回路的协同机制^[104],通过对话式强化学习持续修正模型响应偏差,在实时运行、异常

处置与应急决策等高风险场景中显著提升模型的可靠性、自适应性与操作安全性。

4.3 专有基座大模型的系统化研究框架与技术路线

如图5所示,针对3节所提电力系统专有基座大模型面临的关键挑战,本节围绕4类关键技术提出相应的研究架构与解决方案:多模态结构化数据

融合建立电力系统海量异构数据的通用表征;模型架构创新与高效预训练技术学习电力系统运行的底层机理,是模型智能的“认知起点”;专业知识引导的后训练机制实现从机理解析到下游应用的转化;模型压缩、部署与推理优化技术为基座大模型的工程化落地提供算力与系统支撑。

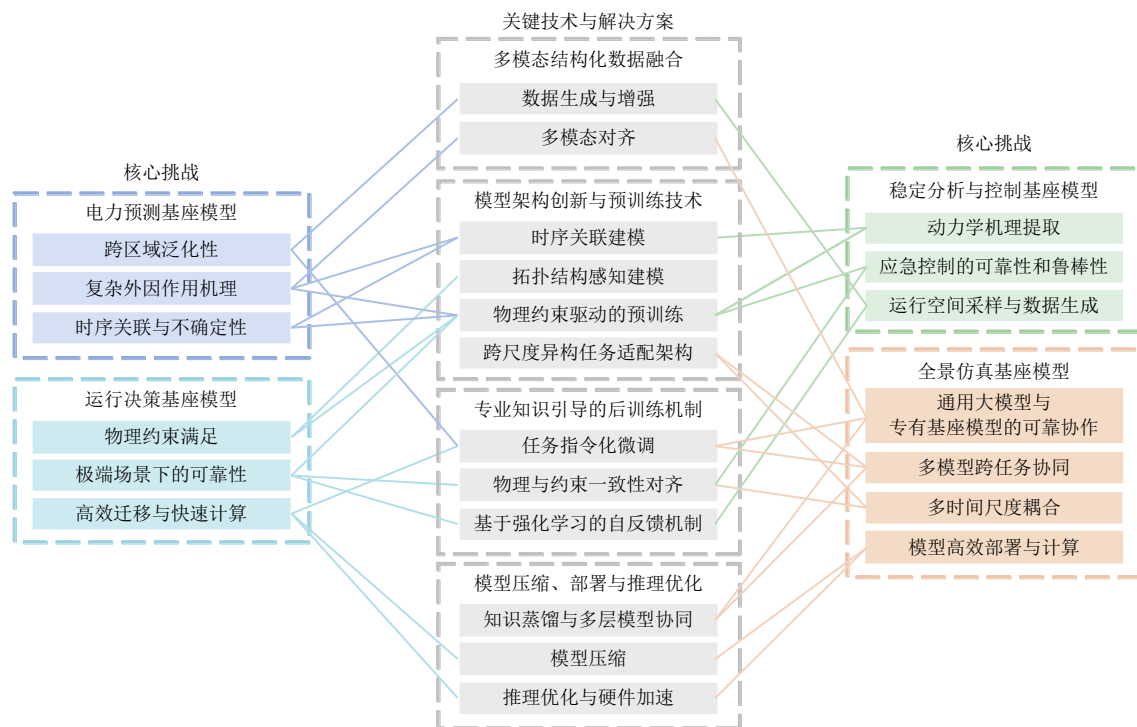


图5 电力系统专有基座大模型研究框架与关键技术

Fig. 5 Research framework and key technologies of domain-specific foundation models in power systems

4.3.1 多模态结构化数据融合

电力系统的运行与控制涉及多源异构数据,涵盖时序测量、空间拓扑、地理气象、文本规则及图像观测等多模态信息。传统的深度学习模型通常仅针对单模态数据进行建模,难以全面刻画电力系统中多尺度、强耦合的动态特性,导致预测与决策模型在复杂场景下的表现受限。多模态结构化数据融合旨在构建能够同时理解电气特征、空间关联与外部环境因素的统一表征空间,从而实现面向多场景、多任务的全局感知与智能决策。

当前多模态结构化数据融合技术的研究路径主要可分为两个层面:数据生成与增强、多模态对齐。为应对电力系统数据的稀缺性、标签不平衡及场景覆盖不足等难题,面向电力系统大模型的数据生成与增强技术成为不可或缺的一环。该类技术主要包括合成数据生成、结构化扰动增强以及标签扩展与重映射。首先,合成数据生成通过深度生成模型,如生成对抗网络^[105]、变分自编码器^[106]等,模

拟电力系统中极端工况、罕见故障、负荷和新能源出力突变等场景,从而拓展训练集的覆盖范围;其次,结构化扰动增强针对时序测量、拓扑结构、气象数据等多模态电力数据,在保持物理一致性的基础上引入噪声、时间缩放、拓扑扰动等变换^[107-109],以提高模型对异常、扰动、迁移场景的鲁棒性;最后,标签扩展与重映射则针对电力系统中标签缺失或类别严重不平衡问题,采用半监督生成^[110]、伪标签扩充^[111]、时序标签插值^[112-113]等技术,增强模型对少数工况类别的感知能力。

针对电力系统的大量多模态异构数据,模态对齐技术旨在通过建立跨模态的一致表示空间,通常遵循“独立预训练-语义映射-跨模态统一”的总体框架^[114-115],首先对不同模态的模型分别进行独立预训练,使其在各自数据域内学习高质量、任务相关的表征能力。例如,利用时序 Transformer 进行电力负荷特征建模,采用图神经网络提取电网拓扑嵌入,使用视觉 Transformer 编码气象卫星图像,

或通过语言模型理解运行规程与市场文本。随后，通过跨模态映射与对齐机制，将这些模态嵌入投射至共享的语义空间，以建立“同一物理事件在不同模态下的等价表示”。典型实现途径包括基于对比学习的模态配对与跨模态注意力机制^[116]，它们能够在保持各模态独立表达能力的同时，实现语义层面的关联与信息互补，从而为后续的联合预训练奠定统一的表示基础。

4.3.2 模型架构创新与预训练技术

预训练是基座大模型构建的核心环节，其目标在于通过大规模数据驱动的学习，获得能够跨任务、跨场景迁移的通用表征。然而，电力系统的复杂性和专业性决定了直接套用自然语言处理或计算机视觉中的模型架构和预训练范式存在显著局限：一方面，电力运行数据具有明显的时序性、空间拓扑性与物理约束特性，传统方法难以捕捉其中的结构化规律；另一方面，电力系统数据的获取受限于隐私保护和仿真实验成本，难以形成与互联网规模相当的训练语料。因此，亟需发展兼顾物理机理表达、先验知识嵌入与算力效率的模型架构与高效预训练体系，为下游任务应用奠定统一表示基础。

预训练的技术路线可从 3 个方面展开。1) 物理约束驱动的预训练以强化模型的物理一致性为核心，在训练目标中显式嵌入潮流方程、节点功率平衡等电力系统约束，通过约束正则、残差最小化或物理一致性对比等机制，使模型在自监督阶段即可学习可行域边界与系统状态约束，从而提升下游优化任务的可迁移性与可解释性。2) 拓扑结构感知建模与预训练技术强调对电力网络结构关系的编码，基于图神经网络、图 Transformer 或谱卷积结构对电网拓扑进行结构化建模，结合连通性保持与层级聚合机制，使模型能够捕捉电气拓扑、支路耦合及区域间能量传递规律，为潮流预测、区网划分与稳定分析等任务提供拓扑一致的嵌入表征。3) 时序建模与预训练技术旨在捕捉电力系统跨时间尺度的动态演化规律，利用时序 Transformer、时空图神经网络及差分方程嵌入层，对功率波动、频率响应与电压暂态等动态过程进行建模，通过自回归预测、时序掩码建模或动力学一致性正则，学习系统在不同时间尺度下的动态演化机理。4) 跨尺度异构任务适配架构旨在解决不同任务数学建模的异构性难题。例如全景仿真基座大模型中，电力系统暂态问题分析与中长期规划的数学本质上存在

巨大差异，可以引入混合专家模型与多任务解码头设计。通过门控网络实现计算路径的动态选择，将电力系统暂态分析任务分配给擅长处理连续刚性问题的神经算子专家网络，将规划决策任务分配给擅长离散逻辑约束的优化专家网络；同时利用多解码头适配不同问题的输出空间，从而在单一基座中实现跨数学范式的统一建模与高效求解。

4.3.3 专业知识引导的后训练机制

电力基座大模型在预训练阶段已通过结构先验与物理约束学习到系统的底层机理与通用表征。然而这种提取到的机理并不能直接用于下游具体任务的求解。因此，有必要在预训练之后引入专业知识引导的后训练机制，将通用机理表征定向对齐到具体下游任务的目标与约束，支撑电力系统专有基座大模型从“机理表征”走向“任务求解”的关键环节。未来研究应聚焦于任务级优化对齐、可验证安全性与动态自适应性 3 方面的持续突破。

任务指令化微调在预训练模型基础上引入带标签的任务数据或指令模板，使模型从底层机理学习转向面向特定任务的输出。典型方法包括 LoRA^[117]、Prefix-tuning^[118]等，通过在部分网络层嵌入低秩更新或可学习适配器，快速实现模型在机组组合、负荷预测、市场出清等任务上的定向迁移。

物理与约束一致性对齐通过引入强化学习或可微约束优化机制，使模型输出符合物理规律与运行约束。以 RLHF 为代表，可将电力仿真反馈、专家操作偏好或安全运行指标转化为奖励信号，实现模型对系统可靠性、稳定性与经济性的策略对齐；或通过可微求解器与拉格朗日乘子嵌入，使优化目标与约束在训练过程中直接参与反向传播。

基于强化学习的自反馈旨在赋予模型在缺乏外部监督条件下的自我校准与生成约束能力，通过构建“自监督评估—反馈修正—约束生成”的闭环机制，增强模型输出的稳定性与可解释性^[119-121]。具体而言，自一致性训练通过多路径推理结果的相互校验提升物理一致性；自蒸馏与内生奖励建模利用历史预测残差、潮流偏差或能量平衡误差等构造内生反馈信号，引导模型在连续学习中自适应优化；可控生成约束则在解码阶段施加基于物理可行域、安全边界或运行规则的软硬约束，实现生成输出的可验证控制。该机制特别适用于扰动响应建模、暂态稳定分析与智能控制等缺乏标注或反馈稀疏的场景，可提升模型在动态环境中的鲁棒性与物

理可行性。

4.3.4 模型压缩、部署与推理优化

随着专有基座大模型参数规模的迅速增长,其计算与存储开销已成为制约实际部署的重要瓶颈。电力系统应用场景具有显著的实时性和可靠性要求,模型不仅需在毫秒级响应控制信号,还需在电网调度、频率控制和市场运行等关键环节中实现稳定可用。然而,针对电力系统稳定控制毫秒级响应的严苛要求,大模型受限于秒级的推理延迟;且全参量大模型难以直接部署到资源受限的调度中心和边缘终端中。因此,发展面向电力系统任务的高效模型压缩与推理优化技术,成为扩大电力系统基座大模型应用的关键环节。

当前大模型的高效部署主要沿3条技术路径展开。其一是模型压缩,包括权重量化、参数剪枝与低秩分解等方法^[122-123]。权重量化通过将浮点参数映射为低比特定数值,在保证精度可接受的前提下显著降低存储与计算开销。针对电力工控设备计算资源受限且对延迟敏感的特性,量化技术具有部署便捷的优势^[124],无需高成本的重新训练,即可在保持模型精度损失极小的前提下,能够将模型体积压缩至原本的约 $1/20$ ^[125-126],成为在受限资源下部署大模型并实现毫秒级响应的关键技术。此外,参数剪枝通过识别和移除对最终预测贡献较小的权重或神经元连接,以减少模型的冗余计算。低秩分解通过矩阵分解或张量分解将高维参数空间压缩至低秩子空间,从而减少计算量与内存需求。

第2条技术路线为知识蒸馏与多层模型端边云协同。针对电力系统对实时性的严苛要求,首先利用知识蒸馏技术,以云端参数量巨大的“教师模型”为知识源,将其在隐空间中的特征表示、预测分布或注意力权重传递给面向边缘侧的“学生模型”^[127],结合逐层蒸馏、对比蒸馏与任务感知蒸馏等先进方法^[128-129],使学生模型在显著降低参数量与推理延迟的同时,仍能继承主模型的高维泛化能力。在此基础上,进一步构建多层协同架构^[130]:云端大模型负责长周期、全局性的复杂推演与策略更新,而经过蒸馏的边端小模型则部署于安全自动装置等计算节点,专注于本地数据的实时处理与毫秒级控制指令下发。这种机制实现了知识在不同计算节点间的高效共享与分工,有效化解了高维认知智能与实时工业控制之间的矛盾。

第3条技术路线是推理优化与硬件加速。该方

向主要关注如何在保持模型结构完整的前提下,充分利用硬件特性以提升计算效率。在并行化与计算调度方面,研究者提出了数据并行、模型并行与张量并行等策略,通过在多个GPU或分布式节点上分解计算图,实现计算负载的精细划分与流水线化执行^[131]。近年来的专家混合架构仅激活部分专家子网络完成前向计算,特别适用于电力系统中任务区域化、时空分层的特性^[132]。在模型结构优化层面,有研究改进了算子与注意力机制。以FlashAttention等快速注意力算法为代表的内存访问优化技术,通过减少显存读写与中间缓存,实现了在长序列输入下的低延迟高精度推理^[133];而算子融合与图级编译优化能够自动合并线性层与激活层等常见操作,减少冗余计算与内存访问,从而进一步缩短延迟^[134]。

5 结论与展望

新型电力系统在可靠性、准确性、泛化性、可解释性等方面对人工智能技术提出了更高的要求,而大模型有望成为解决此问题的技术范式。为此,本文深入分析大模型在电力系统智能化中的赋能作用与实现途径,围绕电力系统通用大模型和专有基座大模型两类技术路线,探讨了“大模型用在哪”、“大模型难在哪”以及“大模型如何开发”3个关键问题。未来电力系统通用大模型有望重塑电力系统的知识获取、工具调用、决策执行和多任务协同范式,成为电力系统的智能中枢;电力系统专有基座大模型有望实现从碎片化小模型到跨域统一基座的跃迁,推动电力系统分析由传统解析建模到人工智能范式的深层转化,为通用大模型提供分析工具和数据支撑。这两类大模型共同构成了AI4S在电力系统领域应用落地的关键技术支撑。然而实现上述愿景仍面临诸多技术瓶颈和开拓性难题。电力系统通用大模型面临多模态知识难融合、多领域任务难协同等挑战;电力系统专有基座大模型面临底层机理难挖掘、物理一致难保证、多类任务难适配、变化工况难泛化等难题。此外,电力系统大模型的构建还面临数据与标准维度的制约,如何生成高质量数据、如何保证数据隐私、如何建立全面客观的评估体系等仍是电力系统大模型真正实现工程落地的重要问题。电力系统大模型目前仍处于初期探索阶段,希望本文能够为新型电力系统智能化转型的研究提供一定参考。

参考文献

- [1] 张智刚, 康重庆. 碳中和目标下构建新型电力系统的挑战与展望[J]. 中国电机工程学报, 2022, 42(8): 2806-2818.
ZHANG Zhigang, KANG Chongqing. Challenges and prospects for constructing the new-type power system towards a carbon neutrality future[J]. Proceedings of the CSEE, 2022, 42(8): 2806-2818(in Chinese).
- [2] 国务院. 国务院关于印发 2030 年前碳达峰行动方案的通知[EB/OL]. (2021-10-24)[2025-11-25]. http://www.mee.gov.cn/zcwj/gwywj/202110/t20211026_957879.shtml. The State Council of the People's Republic of China. Notice of the State Council on Issuing the Action Plan for Carbon Dioxide Peaking Before 2030[EB/OL].(2021-10-24)[2025-11-25]. http://www.mee.gov.cn/zcwj/gwywj/202110/t20211026_957879.shtml(in Chinese).
- [3] 国家发展和改革委员会, 国家能源局. 关于促进新能源消纳和调控的指导意见(发改能源〔2025〕1360号)[EB/OL]. (2025-10-29)[2025-11-25]. https://www.ndrc.gov.cn/xxgk/zcfb/tz/202511/t20251110_1401469.html.
National Development and Reform Commission, National Energy Administration. Guiding Opinions on Promoting the Integration and Regulation of New Energy (Development and Reform Energy [2025] No. 1360) [EB/OL].(2025-10-29)[2025-11-25]. https://www.ndrc.gov.cn/xxgk/zcfb/tz/202511/t20251110_1401469.html(in Chinese).
- [4] 康重庆, 姚良忠. 高比例可再生能源电力系统的关键科学问题与理论研究框架[J]. 电力系统自动化, 2017, 41(9): 2-11.
KANG Chongqing, YAO Liangzhong. Key scientific issues and theoretical research framework for power systems with high proportion of renewable energy [J]. Automation of Electric Power Systems, 2017, 41(9): 2-11(in Chinese).
- [5] 工业和信息化部, 中央网络安全和信息化委员会办公室, 国家发展和改革委员会, 等. 工业和信息化部 中央网络安全和信息化委员会办公室 国家发展和改革委员会 国家标准化管理委员会关于印发国家人工智能产业综合标准化体系建设指南(2024 版)的通知[EB/OL]. (2024-06-05)[2025-11-25]. https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.htm.
Ministry of Industry and Information Technology, Office of the Central Cyberspace Affairs Commission, National Development and Reform Commission, et al. Notice on Issuing the Construction Guidelines for the National Comprehensive Standardization System for the Artificial Intelligence Industry(2024 Edition) by the Ministry of Industry and Information Technology, the Office of the Central Cyberspace Affairs Commission, the National Development and Reform Commission, and the National Administration of Standardization[EB/OL].(2024-06-05)[2025-11-25]. https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.htm(in Chinese).
- [6] 国务院. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. (2017-07-08)[2025-11-25]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
The State Council of the People's Republic of China. Notice of the State Council on Issuing the Development Plan for the New Generation of Artificial Intelligence [EB/OL].(2017-07-08)[2025-11-25]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm(in Chinese).
- [7] 国务院. 国务院关于印发深入实施“人工智能+”行动的意见[EB/OL]. (2025-08-21)[2025-11-25]. https://www.gov.cn/zhengce/content/202508/content_7037861.htm.
The State Council of the People's Republic of China. Opinions of the State Council on Deepening the Implementation of the “Artificial Intelligence Plus” Action[EB/OL].(2025-08-21)[2025-11-25]. https://www.gov.cn/zhengce/content/202508/content_7037861.htm(in Chinese).
- [8] 新华社. 我国大模型数量超 1500 个[N/OL]. (2025-07-28)[2025-11-25]. https://www.gov.cn/yaowen/liebiao/202507/content_7034073.htm.
Xinhua News Agency. Number of China's Large AI Models Exceeds 1,500[N/OL].(2025-07-28)[2025-11-25]. https://www.gov.cn/yaowen/liebiao/202507/content_7034073.htm(in Chinese).
- [9] 国家能源局. 国家能源局关于加快推进能源数字化智能化发展的若干意见[EB/OL]. (2023-03-28)[2025-11-25]. https://www.gov.cn/zhengce/zhengceku/2023-04/02/content_5749758.htm.
National Energy Administration. Several Opinions on Accelerating the Digital and Intelligent Development of the Energy Sector[EB/OL].(2023-03-28)[2025-11-25]. https://www.gov.cn/zhengce/zhengceku/2023-04/02/content_5749758.htm(in Chinese).
- [10] KARAGIANNOPOULOS S, ARISTIDOU P, HUG G. Data-driven local control design for active distribution grids using off-line optimal power flow and machine learning techniques[J]. IEEE Transactions on Smart Grid, 2019, 10(6): 6461-6471.
- [11] HUANG Bin, WANG Jianhui. Applications of physics-informed neural networks in power systems-a review [J]. IEEE Transactions on Power Systems, 2023, 38(1): 572-588.
- [12] VENZKE A, CHATZIVASILEIADIS S. Verification of neural network behaviour: formal guarantees for power

- system applications[J]. IEEE Transactions on Smart Grid, 2021, 12(1): 383-397.
- [13] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020: 159.
- [14] BAHRI Y, DYER E, KAPLAN J, et al. Explaining neural scaling laws[J]. Proceedings of the National Academy of Sciences of the United States of America, 2024, 121(27): e2311878121.
- [15] MOOR M, BANERJEE O, ABAD Z S H, et al. Foundation models for generalist medical artificial intelligence[J]. Nature, 2023, 616(7956): 259-265.
- [16] HOFFMANN J, BORGEAUD S, MENSCH A, et al. Training compute-optimal large language models[J]. arXiv preprint arXiv: 2203.15556, 2022.
- [17] WANG Xuezhai, WEI J, SCHUURMANS D, et al. Self-consistency improves chain of thought reasoning in language models[C]//Proceedings of the 11th International Conference on Learning Representations. Kigali: OpenReview.net, 2023.
- [18] Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models[J]. arXiv preprint arXiv: 2206.07682, 2022.
- [19] LIU Guolong, BAI Yan, WEN Keen, et al. LFLLM: a large language model for load forecasting[J]. Authorea Preprints, 2024.
- [20] ZHOU Yating, WANG Meng. Empower pre-trained large language models for building-level load forecasting[J]. IEEE Transactions on Power Systems, 2025, 40(5): 4220-4232.
- [21] GAO Mengyang, ZHOU Suyang, GU Wei, et al. A general framework for load forecasting based on pre-trained large language model[J]. arXiv preprint arXiv: 2406.11336, 2024.
- [22] LIN Huapeng, YU Miao. A novel distributed PV power forecasting approach based on time-LLM[J]. arXiv preprint arXiv: 2503.06216, 2025.
- [23] WU Tangjie, LING Qiang. STELLM: spatio-temporal enhanced pre-trained large language model for wind speed forecasting[J]. Applied Energy, 2024, 375: 124034.
- [24] LEE J, REW J. Large language models-based feature extraction for short-term load forecasting[J]. Journal of Korea Society of Industrial Information Systems, 2024, 29(3): 51-65.
- [25] GAO Mingyang, ZHOU Suyang, GU Wei, et al. LFPLM: a general and flexible load forecasting framework based on pre-trained language model[J]. arXiv preprint arXiv: 2406.11336v1, 2024.
- [26] ZHANG Minghui, LIU Minhao, WANG Hao, et al. Leveraging large language model for generalization in building energy management[J]. IEEE Transactions on Smart Grid, 2025, 16(6): 4712-4725.
- [27] SELIM A, ZHAO Junbo, YANG Bo. Large language model for smart inverter cyber-attack detection via textual analysis of Volt/VAR commands[J]. IEEE Transactions on Smart Grid, 2024, 15(6): 6179-6182.
- [28] ZUO Yu, QIN Dalin, WANG Yi. Large language model-empowered interactive load forecasting[J]. arXiv preprint arXiv: 2505.16577, 2025.
- [29] MONGAILLARD T, LASAULCE S, HICHEUR O, et al. Large language models for power scheduling: a user-centric approach[C]//Proceedings of 2024 22nd International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks. Seoul: IEEE, 2024: 321-328.
- [30] FENG Junkang, CUI Chenggang, ZHANG Chuanlin, et al. Large language model based agent framework for electric vehicle charging behavior simulation[C]//Proceedings of 2025 10th Asia Conference on Power and Electrical Engineering. Beijing: IEEE, 2025: 2683-2687.
- [31] ZHANG Yan, SABER A M, YOUSSEF A, et al. Grid-agent: an LLM-powered multi-agent system for power grid control[J]. arXiv preprint arXiv: 2508.05702, 2025.
- [32] YANG Xu, LIN Chenhui, LIU Haotian, et al. RL2: reinforce large language model to assist safe reinforcement learning for energy management of active distribution networks[J]. IEEE Transactions on Smart Grid, 2025, 16(4): 3419-3431.
- [33] 新华网. 国家电网发布国内首个千亿级多模态电力行业大模型[EB/OL]. (2024-12-24)[2025-11-25]. <https://www.xinhuanet.com/power/20241224/b37c81368b674fc3978c0703848fddcf/c.html>.
Xinhua Net. State Grid Releases Nation's First 100-Billion-Parameter Multimodal Large Model for the Power Industry[EB/OL].(2024-12-24)[2025-11-25]. <https://www.xinhuanet.com/power/20241224/b37c81368b674fc3978c0703848fddcf/c.html>.(in Chinese)
- [34] 彭雅莹. 南方电网“大瓦特”的故事[N]. 中国能源报, 2024-06-17(15).
PENG Yaying. The story of China Southern Power Grid's "Big Watt" [N]. China Energy News, 2024-06-17(15)(in Chinese).
- [35] LIU Yuxiao, JIA Mengshuo, ZHANG Yongxin, et al. RePower: an LLM-driven autonomous platform for power system data-guided research[J]. Patterns, 2025, 6(4): 101211.
- [36] JIA Mengshuo, CUI Zeyu, HUG G. Enhancing LLMs for

- power system simulations: a feedback-driven multi-agent framework[J]. IEEE Transactions on Smart Grid, 2025, 16(6): 5556-5572.
- [37] LI Bochu, JIANG Yan, XU Jue, et al. The design and implementation of an intelligent Q&A system for electric power safety regulations based on large language model technology[C]//Proceedings of the 6th International Conference on Energy Systems and Electrical Power. Wuhan: IEEE, 2024: 604-607.
- [38] BONADIA R S, TRINDEADE 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.
- [39] LI Ran, PU Chuanqing, TAO Junyi, et al. LLM-based frameworks for power engineering from routine to novel tasks[J]. arXiv preprint arXiv: 2305.11202, 2023.
- [40] 牛泽原, 李嘉媚, 艾芊. 大语言模型在电力系统中的应用初探[J]. 电网技术, 2025, 49(4): 1327-1336.
- NIU Zeyuan, LI Jiamei, AI Qian. Preliminary exploration of the application of large language models in power systems[J]. Power System Technology, 2025, 49(4): 1327-1336(in Chinese).
- [41] HUANG Chenghao, LI Siyang, LIU Ruohong, et al. Large foundation models for power systems[C]//Proceedings of 2024 IEEE Power & Energy Society General Meeting. Seattle: IEEE, 2024: 1-5.
- [42] BERNIER F, CAO Jun, CORDY M, et al. PowerGraph-LLM: novel power grid graph embedding and optimization with large language models[J]. IEEE Transactions on Power Systems, 2025, 40(6): 5483-5486.
- [43] DENG Kaihang, ZHOU Yanzhen, ZENG Hongtai, et al. Power grid model generation based on the tool-augmented large language model[J]. IEEE Transactions on Power Systems, 2025, 40(6): 5487-5490.
- [44] HE S, ZHANG W, CHUNG C. Unveiling optimization potential: using large language models to estimate initial solutions for unit commitment[J]. Social Science Electronic Publishing, 2014, 4(3): 214.
- [45] CHEN Cong, KARADUMAN O, KUANG Xu. Behavioral generative agents for energy operations [J]. arXiv preprint arXiv: 2506.12664, 2025.
- [46] MAJUMDER S, DONG Lin, 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.
- [47] 周远翔, 陈健宁, 赵亮, 等. 电力设备数智化发展历程与大语言模型智能体应用展望[J]. 高电压技术, 2025, 51(8): 4089-4109.
- ZHOU Yuanxiang, CHEN Jianning, ZHAO Liang, et al. Evolution of digitization and intelligence of power equipment and prospects for application of large language model based agent[J]. High Voltage Engineering, 2025, 51(8): 4089-4109(in Chinese).
- [48] TU Shihao, ZHANG Yupeng, ZHANG Jing, et al. PowerPM: foundation model for power systems[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, 2024: 115233-115260.
- [49] HAMANN H F, GJORGIEV B, BRUNSCHWILER T, et al. Foundation models for the electric power grid[J]. Joule, 2024, 8(12): 3245-3258.
- [50] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations [J]. Journal of Computational Physics, 2019, 378: 686-707.
- [51] MISYRIS G S, VENZKE A, CHATZIVASILEIADIS S. Physics-informed neural networks for power systems [C]//Proceedings of 2020 IEEE Power & Energy Society General Meeting. Montreal: IEEE, 2020: 1-5.
- [52] HU Xinyue, HU Haoji, VERMA S, et al. Physics-guided deep neural networks for power flow analysis[J]. IEEE Transactions on Power Systems, 2021, 36(3): 2082-2092.
- [53] STIASNY J, MISYRIS G S, CHATZIVASILEIADIS S. Physics-informed neural networks for non-linear system identification for power system dynamics [C]//Proceedings of 2021 IEEE Madrid PowerTech. Madrid: IEEE, 2021: 1-6.
- [54] MISYRIS G S, STIASNY J, CHATZIVASILEIADIS S. Capturing power system dynamics by physics-informed neural networks and optimization[C]//Proceedings of 2021 60th IEEE Conference on Decision and Control. Austin: IEEE, 2021: 4418-4423.
- [55] NAIR V, BARTUNOV S, GIMENO F, et al. Solving mixed integer programs using neural networks[J]. arXiv preprint arXiv: 2012.13349, 2020.
- [56] STIASNY J, MISYRIS G S, CHATZIVASILEIADIS S. Transient stability analysis with physics-informed neural networks[J]. arXiv preprint arXiv: 2106.13638, 2021.
- [57] 于硕, 王司宇, 王超, 等. 类 ChatGPT 大语言模型在电力系统中的应用前景[J]. 电气时代, 2023(10): 50-53.
- YU Shuo, WANG Siyu, WANG Chao, et al. Application prospects of ChatGPT-like large language models in power systems[J]. Electric Age, 2023(10): 50-53(in Chinese).
- [58] 赵俊华, 文福拴, 黄建伟, 等. 基于大语言模型的电力系统通用人工智能展望: 理论与应用[J]. 电力系统自动化, 2024, 48(6): 13-28.
- ZHAO Junhua, WEN Fushuan, HUANG Jianwei, et al. Prospect of artificial general intelligence for power

- systems based on large language model: theory and applications[J]. Automation of Electric Power Systems, 2024, 48(6): 13-28(in Chinese).
- [59] 杨浩蓝, 刘友波, 李争博, 等. 新型配电系统运行专用大模型关键技术展望[J/OL]. 中国电机工程学报, 2025[2025-10-29]. <https://link.cnki.net/urlid/11.2107.TM.20250820.1724.039>.
- YANG Haolan, LIU Youbo, LI Zhengbo, et al. Key technologies of specialized large language model for operation of next-era distribution system: a prospective study[J/OL]. Proceedings of the CSEE, 2025 [2025-10-29]. <https://link.cnki.net/urlid/11.2107.TM.20250820.1724.039>(in Chinese).
- [60] 孙秋野, 张瑞霞, 陈东岳. 大模型技术赋能电力系统的应用及技术路线展望[J/OL]. 电力系统自动化, 2025[2025-10-29]. <https://link.cnki.net/urlid/32.1180.TP.20250317.1811.006>.
- SUN Qiuye, ZHANG Ruixia, CHEN Dongyue. Application and route prospect of empowering power systems with large language models [J/OL]. Automation of Electric Power Systems, 2025[2025-10-29]. <https://link.cnki.net/urlid/32.1180.TP.20250317.1811.006>(in Chinese).
- [61] LU Xin, QIU Jing, YANG Yi, et al. Large language model-based bidding behavior agent and market sentiment agent-assisted electricity price prediction[J]. IEEE Transactions on Energy Markets, Policy and Regulation, 2025, 3(2): 223-235.
- [62] LOPEZ-GARCIA T B, DOMÍNGUEZ-NAVARRO J A. Optimal power flow with physics-informed typed graph neural networks[J]. IEEE Transactions on Power Systems, 2025, 40(1): 381-393.
- [63] LIU Yundi, LI Yuanzheng, HE Shangyang, et al. A physics-informed graph convolution network for AC optimal power flow via refining DC solution[J]. IEEE Transactions on Power Systems, 2026, 41(1): 438-453.
- [64] ZHANG Ning, JIA Hongyang, HOU Qingchun, et al. Data-driven security and stability rule in high renewable penetrated power system operation [J]. Proceedings of the IEEE, 2023, 111(7): 788-805.
- [65] JIA Hongyang, HOU Qingchun, YONG Pei, et al. Learning multiple convex voltage stability constraints for unit commitment[J]. IEEE Transactions on Power Systems, 2025, 40(1): 125-137.
- [66] GAO Qian, YANG Zhifang, LI Wenyan, et al. Optimality-guaranteed acceleration of unit commitment calculation via few-shot solution prediction[J]. IEEE Transactions on Power Systems, 2025, 40(2): 1583-1595.
- [67] QIN Jingtao, YU Nanpeng. Solve large-scale unit commitment problems by physics-informed graph learning[J]. IEEE Transactions on Power Systems, 2025, 40(6): 4439-4451.
- [68] YU Jianshu, YONG Pei, YANG Zhifang, et al. Stochastic-robust optimal power flow with small-signal stability guarantee under renewable uncertainties[J]. IEEE Transactions on Sustainable Energy, 2025, 16(3): 1711-1722.
- [69] GAO Maosheng, YU Juan, YANG Zhifang, et al. Physics embedded graph convolution neural network for power flow calculation considering uncertain injections and topology[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(11): 15467-15478.
- [70] JIA Yixiong, WU Xian, YANG Zhifang, et al. A two-stage approach for topology change-aware data-driven OPF[J]. IEEE Transactions on Power Systems, 2025, 40(2): 1854-1867.
- [71] HATZIGARYIOU N, MILANOVIC J, RAHMANN C, et al. Definition and classification of power system stability - revisited & extended[J]. IEEE Transactions on Power Systems, 2021, 36(4): 3271-3281.
- [72] YAO Wei, ZHANG Runfeng, ZHANG Yurun, et al. A hybrid deep learning framework with local-global feature extraction for intelligent power system stability assessment[J]. IEEE Transactions on Power Systems, 2025, 40(6): 4718-1731.
- [73] CHEN Qifan, BU Siqi, WANG Huaiyuan, et al. Real-time multi-stability risk assessment and visualization of power systems: a graph neural network-based method[J]. IEEE Transactions on Power Systems, 2025, 40(4): 2955-2968.
- [74] JIA Hongyang, HOU Qingchun, ZHANG Jiawei, et al. Converter-driven stability constrained generation expansion for high IBG-penetrated power systems [J]. IEEE Transactions on Power Systems, 2025, 40(6): 4810-4825.
- [75] XIA Tian, ZHANG Ning, LI Weiran, et al. Efficient embedding of neural network-based stability constraints into power system dispatch[J]. IEEE Transactions on Power Systems, 2024, 39(3): 5443-5446.
- [76] 丁俐夫, 陈颖, 肖谭南, 等. 基于大语言模型的新型电力系统生成式智能应用模式初探[J]. 电力系统自动化, 2024, 48(19): 1-13.
- DING Lifu, CHEN Ying, XIAO Tannan, et al. Exploration of generative intelligent application mode for new power systems based on large language models[J]. Automation of Electric Power Systems, 2024, 48(19): 1-13(in Chinese).
- [77] JIA Mengshuo, HUG G, ZHANG Ning, et al. Generalized data-driven power flow linearization: methodologies, evaluations, and challenges[J]. IEEE Transactions on Power Systems, 2025.

- [78] ZHOU Xiyuan, ZHAO Huan, CHENG Yuheng, et al. ElecBench: a power dispatch evaluation benchmark for large language models[J]. arXiv preprint arXiv: 2407.05365, 2024.
- [79] VARBELLA A, AMARA K, GJORGIEV B, et al. PowerGraph: a power grid benchmark dataset for graph neural networks[C]//Proceedings of the Advances in Neural Information Processing Systems 37. Vancouver, 2024: 110784-110804.
- [80] ZHENG Xiangtian, XU Nan, TRINH L, et al. A multi-scale time-series dataset with benchmark for machine learning in decarbonized energy grids [J]. Scientific Data, 2022, 9(1): 359.
- [81] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020: 793.
- [82] XIANG Zhishang, WU Chuanjie, ZHANG Qinggang, et al. When to use graphs in RAG: a comprehensive analysis for graph retrieval-augmented generation [J]. arXiv preprint arXiv: 2506.05690, 2025.
- [83] GAO Yunfan, XIONG Yun, ZHONG Yijie, et al. Synergizing RAG and reasoning: a systematic review[J]. arXiv preprint arXiv: 2504.15909, 2025.
- [84] LI Yangning, ZHANG Weizhi, YANG Yuyao, et al. Towards agentic RAG with deep reasoning: a survey of RAG-reasoning systems in LLMs[J]. arXiv preprint arXiv: 2507.09477, 2025.
- [85] ASAI A, WU Zeqiu, WANG Yizhong, et al. Self-RAG: learning to retrieve, generate, and critique through self-reflection[C]//Proceedings of the 12th International Conference on Learning Representations. Vienna: OpenReview.net, 2024.
- [86] MAO Kelong, LIU Zheng, QIAN Hongjin, et al. RAG-studio: towards in-domain adaptation of retrieval augmented generation through self-alignment[C]//Findings of the Association for Computational Linguistics: EMNLP 2024. Miami: ACL, 2024: 725-735.
- [87] KOJIMA T, GU S S, REID M, et al. Large language models are zero-shot reasoners[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022, 35: 1613.
- [88] WEI J, WANG Xuezhi, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022: 1800.
- [89] YAO Shunyu, YU Dian, ZHAO J, et al. Tree of thoughts: deliberate problem solving with large language models [C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, 2023: 11809-11822.
- [90] BESTA M, BLACH N, KUBICEK A, et al. Graph of thoughts: solving elaborate problems with large language models[C]//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Washington: AAAI, 2024: 17682-17690.
- [91] WANG Shuhe, ZHANG Shengyu, ZHANG Jie, et al. Reinforcement learning enhanced LLMs: a survey [J]. arXiv preprint arXiv: 2412.10400, 2024.
- [92] STIENNON N, OUYANG Long, WU J, et al. Learning to summarize with human feedback[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020: 3008-3021.
- [93] OUYANG Long, WU J, JIANG Xu, et al. Training language models to follow instructions with human feedback[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, 2022: 27730-27744.
- [94] RAFAILOV R, SHARMA A, MITCHELL E, et al. Direct preference optimization: your language model is secretly a reward model[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, 2023: 53728-53741.
- [95] ZHONG Han, SHAN Zikang, FENG Guhao, et al. DPO meets PPO: reinforced token optimization for RLHF[C]//Proceedings of the 42nd International Conference on Machine Learning. Vancouver: OpenReview.net, 2025.
- [96] YE Chenlu, XIONG Wei, ZHANG Yuheng, et al. Online iterative reinforcement learning from human feedback with general preference model[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, 2024: 81773-81807.
- [97] SHANI L, ROSENBERG A, CASSEL A, et al. Multi-turn reinforcement learning with preference human feedback [C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, 2024: 118953-118993.
- [98] HONG I, LI Zichong, BUKHARIN A, et al. Adaptive preference scaling for reinforcement learning with human feedback[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, 2024: 107249-107269.
- [99] JIAO Yizhu, ZHONG Ming, LI Sha, et al. Instruct and extract: instruction tuning for on-demand information extraction[C]//Proceedings of 2023 Conference on Empirical Methods in Natural Language Processing. Singapore: ACL, 2023.

- [100]WANG Weihi, DONG Li, CHENG Hao, et al. Augmenting language models with long-term memory[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, 2023: 74530-74543.
- [101]YAO Shunyu, ZHAO J, YU Dian, et al. ReAct: synergizing reasoning and acting in language models[C]//Proceedings of the 11th International Conference on Learning Representations. Kigali: OpenReview.net, 2023.
- [102]HUANG Xiaowei, RUAN Wenjie, HUANG Wei, et al. A survey of safety and trustworthiness of large language models through the lens of verification and validation[J]. Artificial Intelligence Review, 2024, 57(7): 175.
- [103]JIANG Fengqing, XU Zhangchen, LI Yuetai, et al. SafeChain: safety of language models with long chain-of-thought reasoning capabilities[C]//Findings of the Association for Computational Linguistics. Vienna: ACL, 2025.
- [104]DAI J, PAN Xuehai, SUN Ruiyang, et al. Safe RLHF: safe reinforcement learning from human feedback[C]//Proceedings of the 12th International Conference on Learning Representations. Vienna: OpenReview.net, 2024.
- [105]GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014: 2672-2680.
- [106]KINGMA D P, WELLING M. Auto-encoding variational bayes[C]//Proceedings of the 2nd International Conference on Learning Representations. Banff, 2014.
- [107]ZHANG Hongyi, Cissé M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization[C]//Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018.
- [108]YUN S, HAN D, CHUN S, et al. CutMix: regularization strategy to train strong classifiers with localizable features[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 6023-6032.
- [109]WEN Qingsong, SUN Liang, YANG Fan, et al. Time series data augmentation for deep learning: a survey[C]//Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal: ijcai.org, 2021.
- [110]SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training GANs[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016: 2234-2242.
- [111]SOHN K, BERTHELOT D, LI Chunliang, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020: 51.
- [112]BERTHELOT D, CARLINI N, GOODFELLOW I, et al. MixMatch: A holistic approach to semi-supervised learning[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019: 454.
- [113]YOON J, JARRETT D, VAN DER SCHAAR M. Time-series generative adversarial networks[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019: 494.
- [114]RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021: 8748-8763.
- [115]ZHAI Xiaohua, WANG Xiao, MUSTAFA B, et al. LiT: zero-shot transfer with locked-image text tuning[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 18123-18133.
- [116]LU Jiasen, BATRA D, PARIKH D, et al. ViLBERT: pretraining task-agnostic visiolinguistic representations for vision-and-language tasks[C]//Proceedings of the Annual Conference on Neural Information Processing Systems 2019. Vancouver, 2019.
- [117]HU E J, SHEN Yelong, WALLIS P, et al. LoRA: low-rank adaptation of large language models[C]//Proceedings of the 10th International Conference on Learning Representations. OpenReview.net, 2022.
- [118]LI X L, LIANG P. Prefix-tuning: optimizing continuous prompts for generation[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. ACL, 2021.
- [119]VAN NIEKERK C, VUKOVIC R, RUPPIK B M, et al. Post-training large language models via reinforcement learning from self-feedback[J]. arXiv preprint arXiv: 2507.21931, 2025.
- [120]LIANG Xun, SONG Shichao, ZHENG Zifan, et al. Internal consistency and self-feedback in large language models: a survey[J]. arXiv preprint arXiv: 2407.14507, 2024.
- [121]SHAFAYAT S, TAJWAR F, SALAKHUTDINOV R, et al. Can large reasoning models self-train?[J]. arXiv

- preprint arXiv: 2505.21444, 2025.
- [122]ZHU Xunyu, LI Jian, LIU Yong, et al. A survey on model compression for large language models [J]. Transactions of the Association for Computational Linguistics, 2024, 12: 1556-1577.
- [123]WANG Wenxiao, CHEN Wei, LUO Yicong, et al. Model compression and efficient inference for large language models: a survey[J]. arXiv preprint arXiv: 2402.09748, 2024.
- [124]DETTMERS T, LEWIS M, BELKADA Y, et al. GPT3.int8(): 8-bit matrix multiplication for transformers at scale[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, 2022: 30318-30332.
- [125]EGIAZARIAN V, PANFEROV A, KUZNEDELEV D, et al. Extreme compression of large language models via additive quantization[C]//Proceedings of the 41st International Conference on Machine Learning. Vienna, 2024.
- [126]MA Shuming, WANG Hongyu, MA Lingxiao, et al. The era of 1-bit LLMs: all large language models are in 1.58 bits[J]. arXiv preprint arXiv: 2402.17764, 2024.
- [127]GOU Jianping, YU Baosheng, MAYBANK S J, et al. Knowledge distillation: a survey[J]. International Journal of Computer Vision, 2021, 129(6): 1789-1819.
- [128]ACHARYA K, VELASQUEZ A, SONG H H. A survey on symbolic knowledge distillation of large language models[J]. IEEE Transactions on Artificial Intelligence, 2024, 5(12): 5928-5948.
- [129]FANG Luyang, YU Xiaowei, CAI Jiazhang, et al. Knowledge distillation and dataset distillation of large language models: emerging trends, challenges, and future directions[J]. Artificial Intelligence Review, 2026, 59(1): 17.
- [130]LI Senyao, WANG Haozhao, XU Wenchao, et al. Collaborative inference and learning between edge SLMs and cloud LLMs: a survey of algorithms, execution, and open challenges[J]. arXiv preprint arXiv: 2507.16731, 2025.
- [131]NARAYANAN D, SHOEYBI M, CASPER J, et al. Efficient large-scale language model training on GPU clusters using megatron-LM[C]//Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. St. Louis Missouri: ACM, 2021: 58.
- [132]RAJBHANDARI S, LI Conglong, YAO Zhewei, et al. DeepSpeed-MoE: advancing mixture-of-experts inference and training to power next-generation AI scale[C]//Proceedings of the 39th International Conference on Machine Learning. Baltimore: PMLR, 2022: 18332-18346.
- [133]DAO T, FU D Y, ERMON S, et al. FLASHATTENTION: fast and memory-efficient exact attention with IO-awareness[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022: 1189.
- [134]NIU W, GUAN J, WANG Y, et al. Dnnfusion: accelerating deep neural networks execution with advanced operator fusion[C]//Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation. 2021: 883-898.



杜博骏

在线出版日期: 2025-01-15.

收稿日期: 2025-12-11.

作者简介:

杜博骏(2002), 男, 博士研究生, 研究方向为电力系统规划运行、低碳需求响应、人工智能与大模型在电力系统的应用等, dubj24@mails.tsinghua.edu.cn;

钟非池(2000), 男, 博士研究生, 研究方向为深度迁移学习、时序预测、电力大模型等, zhongfeichi@zju.edu.cn;

*通信作者: 侯庆春(1993), 男, 博士, 研究员, 博士生导师, 研究方向为人工智能与优化决策、高比例可再生能源电力系统运行规划、基础模型与大模型、电商供应链与物流交通等, houqingchun@zju.edu.cn;

贾宏阳(1999), 男, 博士研究生, 研究方向为面向新型电力系统的可信机器学习模型与人工智能算法, 数据驱动的电力系统安全稳定规则提取等, jhy21@mails.tsinghua.edu.cn;

张宁(1985), 男, 博士, 副教授, 博士生导师, 主要从事低碳电力技术、电力系统规划与运行、多能源系统等研究工作, ningzhang@tsinghua.edu.cn;

康重庆(1969), 男, 教授, 研究方向为可再生能源、低碳电力、能源互联网等, cqkang@tsinghua.edu.cn.

(编辑 乔宝榆, 张蕾)