**基于LLM大语言模型的电力工程前期工作智能化系统的落地研究**

**一、课题目标**

**1、近期目标**

1）启动LLM系统的本地化部署工作

根据公司资源情况，对LLM系统的本地部署方案进行技术经济分析，解决开源模型、GPU及其硬件平台的选型问题，为后续模型的高速推理、领域预训练和指令微调以及Agent体系建设打好软硬件基础。

2）启动测试驱动的LLM应用研究工作

以针对开源LLM模型开展的多维度、多层级的评估测试工作为抓手，充分挖掘电力工程前期工作等业务领域的人工智能应用空间，为“十五五”期间通用人工智能(AGI)在电力系统的应用做好技术储备。

3）启动人工智能人才梯队的建设工作

本轮人工智能技术的突破与前一轮2012年左右以卷积神经网络为代表的技术不同，以自注意力(Self-Attention)为核心的Transformer深度学习技术，通过数据和参数的规模化，实现了模型在海量背景知识、语言界面和逻辑推理三大能力上的涌现，业界普遍认为3-5年内将实现AGI，大量重复性甚至创新性的工作都可能被人工智能所取代。建立一支业务和AI技术双一流的队伍是当务之急。

**2、中期目标**

1）大语言模型数据工程

开源LLM模型具备海量的通用型知识和较强的基础推理分析能力，但是缺乏垂域知识和经验，无法完成依托行业一线信息的文本或多模态信息处理、数字人交互以及决策辅助等工作，因此首先需要完成对本领域TB级别数据的整理、清洗和打标工作，为训练垂域大语言模型打下基础。

2）垂域大语言LLM模型的建设

垂域大语言模型的训练需要依托海量数据和一定的算力资源，目前这类模型的预训练和指令微调技术栈已基本成熟，在数据、算力和人才等资源到位的前提下，可以及早开展垂域模型的训练，占领业务智能化的高地。

3）垂域大推理LRM模型的建设

大推理模型LRM与大语言模型LLM在概念和能力上有所区别。

LLM具备语言和逻辑能力，但其能力更接近于人类的口头即兴表达和快速直觉反应，输出信息的可靠性和逻辑性较差。

LRM通过基于探索、试错和回溯等手段的框架机制，具备了AGI第二个关键能力维度，即慢思考能力，可以对复杂的规划、分析甚至数学问题进行逻辑严密和相对完备的解答。垂域LRM模型的建设，应成为“十五五”业务智能化工作的重中之重。

**二、研究内容设想**

**1、LLM系统的本地化部署研究**

1）LLM本地部署的硬件系统

2）LLM本地部署的软件生态

3）LLM本地部署的技术经济分析

**2、基于开源LLM的前期工作智能化应用研究**

1）LLM的基本能力分析

2）基于LLM的工具调用体系

3）基于LLM的增强检索生成

4）基于LLM的Agent体系

5）基于开源LLM的前期工作智能化应用方案

**3、垂域LLM的训练研究**

1）数据工程研究

2）垂域继续预训练研究

3）垂域指令微调研究

**4、垂域LRM技术研究**

1）强化学习、树搜索和自博弈

2）LRM模型训练技术研究

3）LRM模型推理技术研究