

《电力系统的大语言模型微调数据准备规范》编制说明

(征求意见稿)

一、工作简况

1 主要工作过程

起草（草案、调研）阶段：

国网信息通信产业集团有限公司于 2023 年 9 月成立了标准编制工作起草小组，组织标准编制组织工作。标准编制工作起草小组在 2023 年 10 月积极组织筹备和征集标准起草单位，成立标准起草工作组。

标准起草工作组制定了标准编制工作计划、编写大纲，明确任务分工及各阶段进度时间。同时，标准起草工作组成员认真学习了 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准的结构和编写规则》，GB/T20000.2—2009《标准化工作指南 第 2 部分：采用国际标准的规则》，结合标准制定工作程序的各个环节，进行了探讨和研究。

标准立项阶段：标准立项阶段：2023 年 11 月，在北京召开了第一次标准的专家立项评审会，经中国电工技术学会标准工作委员会专家组审议，批准《电力系统的大语言模型微调数据准备规范》标准立项；

2 主要参加单位和起草工作组及其所做的工作

本标准由国网信息通信产业集团有限公司、国家电网有限公司大数据中心、中国电力科学研究院有限公司、国网智能电网研究院有限公司、北京国网信通埃森哲信息技术有限公司、四川中电启明星信息技术有限公司、国网福建省电力有限公司共同负责起草。

主要成员：李强、赵峰、赵永生、邱镇、陈振宇、李博、刘识、李炳森、黄晓光、秦余、王晓东、张琳瑜、张国梁、刘园园、崔迎宝、王兴涛、卢大玮、吴迪、赵如意、宋卫平、杨帆、高攀、王红蕾、董梅、李欢欢、徐小云、叶林峰、赵林林、王誉博、李扬笛、杨彦、林晨翔等人。

所做的工作：标准编写组收集了电力系统中大语言模型应用以及数据准备等方面的相关材料，通过对比整理分析，确定了标准主要技术内容，主要由国网信息通信产业集团有限公司牵头完成标准初稿编制，其他参与单位配合编制，并负责收集相关资料、提出建议。

二、标准编制原则和主要内容

1、标准编制原则

本标准按照 GB/T1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构与起草规则》的编写原则制定，定位为团体标准是对国家标准的补充，与相关技术领域的国家现行法律、法规、规章、政策及相关标准保持一致。

本标准遵循科学性、先进性、经济性，坚持实事求是，以符合标准化工作导则，综合国内外经验、公平公正公开、实际可操作、可复制可验证、综合考虑技术安全伦理要求，制定了《电力系统的大语言模型微调数据准备规范》，以电力系统的大语言模型技术和丰富的数据处理经验为基础，遵守国家有关法律法规，符合团体标准要求，目的在于规范在大语言模型训练中的数据处理过程，以保证数据的可用性、一致性和可追溯性。

2、标准主要内容

本标准正文内容分为八个章节，第一章是本标准的适用范围、第二章是规范性引用文件、第三章是术语和定义、第四章是缩略语、第五章是总则

第六章是数据准备要求，主要从数据来源、数据多样性与数据质量三个方面介绍对数据收集过程进行相关的规范性要求，确保大语言模型微调技术应用过程中训练数据符合要求。

第七章是数据准备规范，主要对电力系统中大语言模型的数据准备过程中的关键性步骤进行规范性要求。

第八章是大语言模型微调数据准备流程规范，主要规范了在电力系统中通用的大语言模型微调数据准备流程的相应要求。

3、主要技术差异

本标准为新制度标准，无主要技术差异。

4、解决的主要问题

针对目前电力系统中大语言模型技术应用相关的业务场景，主要解决如何有效地采集和清洗与电力系统相关的数据，包括数据来源的选择和数据质量的评估；定义必要的数据库标签和注释，以确保模型了解电力系统领域的相关概念和关系；规定数据的格式和结构，以适应大语言模型的微调需求，包括数据的输入和输出格式；确定数据的安全存储方式，并定义数据共享政策，以确保数据的合规和安全性。

三、主要试验（或验证）情况

《电力系统的大语言模型微调数据准备规范》的主要试验情况包括场景化验证、数据质量验证、微调模型性能验证、可信度验证和安全性验证。这些试验旨在验证电力系统中大语言模型微调过程中数据准备在特定场景下的数据可用性、一致性和可追溯性，通过严格的验证和评价流程，提供可靠的基于实际需求的验证和评价体系，确保大语言模型在电力系统中的应用具备准确性、可信度和安全性，并为推广和应用大语言模型提供坚实的数据基础。

四、标准中涉及专利的情况

本标准不涉及专利问题。

五、预期达到的社会效益、对产业发展的作用等情况

电力系统是一个复杂的系统，准确预测和优化能够对能源供应和消费产生深远的影响。通过大型语言模型的微调，可以提高对电力系统中各种情况的预测和优化能力，从而降低能源浪费，提高效率。大型语言模型的微调能够为电力系统的运营和管理提供更智能的决策支持，解决电力系统中很多场景问题，包括预测负荷需求、优化能源分配、改善供电稳定性等方面，有助于管理者做出更明智、更基于数据的决策。对电力系统的大型语言模型微调需要涉及大量的数据和技术，将促进相关领域的技术创新和发展。通过更好地理解 and 利用数据，可以推动电力系统相关技术的进步，从而推动整个产业的发展。

六、与国际、国外对比情况

未检索到国际同类标准。

七、在标准体系中的位置，与现行相关法律、法规、规章及相关标准，特别是强制性标准的协调性

本标准与相关技术领域的国家现行法律、法规和政策保持一致。

八、重大分歧意见的处理经过和依据

标准编制过程中充分征集了专家意见，所有意见均按照标准编制程序进行了是否采纳，不存在重大分歧意见。

九、标准性质的建议说明

建议本标准的性质为团体标准。

十、贯彻标准的要求和措施建议

建议本标准批准发布7天后实施。

十一、废止现行相关标准的建议

无。

十二、其他应予说明的事项

无。