

《电力人工智能样本增广技术架构要求》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：2023年10月开始，国家电网有限公司大数据中心牵头各单位成立标准编写组，讨论确定了标准的主要内容及分工，同时进行调研分析，收集资料，形成标准草案稿，准备立项审查答辩；

立项阶段：2023年11月开始进行专家函审，其中，七位专家同意立项并提出部分意见，详见标准提案立项评估函审结论表与立项建议汇总表。

中期稿评审阶段：2023年12月报2023年中国电工技术学会标准提案申报项目（第五批），经专家组审阅同意立项并提出意见，修改后形成征求意见稿。

2 主要参加单位和起草工作组成员及其所做的工作

本标准由国家电网有限公司大数据中心、国网信息通信产业集团有限公司、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、安徽继远软件有限公司、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、邱镇、陈振宇、李博、刘识、李炳森、黄晓光、张琳瑜、秦余、王晓东、张国梁、周逸平、苏勇、朱署光、李小宁、徐凡、郑碧煌、李黎、余江斌、郭庆、浦正国、薛濛、黄旭东、聂文萍、刘晓飞、刘健、李扬笛、林晨翔、谢炜。

所做的工作：标准编写组查阅收集人工智能领域样本增广技术方面的相关资料，结合电力人工智能场景，经分析讨论和对比整理确定了标准主要技术内容，由国家电网有限公司大数据中心牵头完成标准草案编制，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》的规定起草，遵循科学性、先进性、经济性，坚持实事求是，以电力人工智能样本增广技术为基础，遵守国家有关法律、法规，符合团体标准要求，

目的在于明确电力人工智能样本增广技术架构要求，提高电力人工智能模型的泛化能力和准确性，提升公司电力人工智能样本增广技术及应用水平，促进电力人工智能技术的标准化和规范化，为电力行业的数字化转型和智能化升级提供有力的技术支持和保障。

在标准编制过程中，主要参考了《信息技术 词汇 第 29 部分：人工智能 语音识别与合成》(GB/T 5271.29—2006)、《信息技术 词汇 第 31 部分：人工智能 机器学习》(GB/T 5271.29—2006)、《纸质档案数字复制件光学字符识别 OCR 工作规范》(DA/T 77-2019)等标准。

2、标准主要内容

本标准分为 11 个章节：（1）范围；（2）规范性引用文件；（3）术语和定义；（4）符号、代号和缩略语；（5）样本增广技术总体架构；（6）图像类样本增广技术；（7）文本文档类样本增广技术；（8）音频类样本增广技术；（9）样本增广效果评价要求；（10）样本增广策略制定要求；（11）样本增广算子编排技术和功能要求。主要内容如下：

图像类样本增广技术。规范图像类样本增广技术：包括基本图像增广、混合图像增广和虚拟图像生成。

文本文档类样本增广技术。规范文本文档类样本增广技术：包括标签无关方法、标签相关增广方法、用于 OCR 文档的样本增广技术。

语音类样本增广技术。规范语音类样本增广技术：包括回译技术、词汇替换技术、随机噪声引入技术和生成式技术。

样本增广效果评价要求。规范样本增广效果评价要求：包括通用评价要求、图像类样本增广效果评价要求、文本类样本增广效果评价要求、音频类样本增广效果评价要求。

样本增广策略制定要求。规范样本增广策略制定要求：包括样本增广目标、样本增广算子选择、样本增广算子的顺序、样本增广程度、样本增广的随机性。

样本增广算子编排技术和功能要求。规范样本增广算子编排技术和功能要求：要求增广算子编排技术具备可扩展性、并行性和容错性，数据流管理、优化、调度、可视化和管理。

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

1) 样本数量不足和质量层次不齐：在电力人工智能领域，训练样本的数量和质量对模型的准确性和泛化能力有着至关重要的影响。由于电力行业的特殊性和复杂性，收集到的训练样本往往数量不足、质量层次不齐，这给模型训练和测试带来了很大的困难；

2) 模型泛化能力不足：电力人工智能模型在训练过程中往往会出现过拟合和欠拟合的问题，导致模型在测试集上的表现不佳，泛化能力不足。这主要是因为训练样本的数量和质量、模型复杂度、训练时间等因素的影响；

3) 架构设计不规范：电力人工智能样本增广架构设计是影响样本增广效果的关键因素之一，目前电力人工智能样本增广技术的架构设计往往缺乏规范和标准，给行业应用和推广带来了一定的困难。通过规定电力人工智能平台中样本增广技术的架构设计、技术要求和评估指标，提高电力人工智能模型的泛化能力和准确性，促进电力人工智能样本增广技术的标准化和规范化。

三、主要试验（或验证）情况

本标准不涉及试验（或研制）情况。

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

- (1) 填补电力人工智能样本增广技术领域标准的空白；
- (2) 规范电力人工智能样本增广技术架构设计，提升电力人工智能样本增广效果；
- (3) 提高电力人工智能模型的泛化效果。

六、与国际、国外对比情况

本标准不涉及国际标准。本标准在编制过程中参考了《信息技术 词汇 第 29 部分：人工智能 语音识别与合成》(GB/T 5271.29—2006)、《信息技术 词汇 第 31 部分：人工智能 机器学习》(GB/T 5271.29—2006)、《纸质档案数字复制件光学字符识别 OCR 工作规范》(DA/T 77-2019)等标准。目前，国内缺少电力人工智能样本增广技术相关标准。本标准属于国内先进水平。本标准与现行的

相关法律、法规、规章与相关标准保持一致。

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

本标准与现行的相关法律、法规、规章与相关标准保持一致。

八、重大分歧意见的处理经过和依据

无。标准编制过程中广泛征集了专家意见，所有意见均按照标准编制程序进行了采纳，不存在重大分歧意见。

九、标准性质的建议说明

建议本团体标准的性质为推荐性团体标准。

十、贯彻标准的要求和措施建议

(1) 规定相关从事电力人工智能领域工作的人员或团体，按照此标准相关要求，开展电力人工智能样本增广相关工作；

(2) 建议在本标准的指导下，进行电力人工智能样本增广相关工作，规范电力人工智能样本增广技术流程；

(3) 建议本标准批准发布 2 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。