

T/CWDPA

中国西部开发促进会团体标准

T/CWDPA 000—2026

电力行业大模型业务适配能力评估技术规范

Technical specification for business adaptability assessment of large models in the
electric power industry

征求意见稿

2026-XX-XX 发布

2026-XX-XX 实施

中国西部开发促进会团体标准 发布

目 次

前言 III

引言 IV

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总体要求 1

 4.1 评估原则 1

 4.2 评估目标 1

 4.3 评估对象与范围 2

 4.4 评估等级 2

 4.5 评估组织与人员 2

 4.6 评估条件 2

5 评估指标体系 2

 5.1 指标构成 2

 5.2 基础能力 2

 5.3 电力业务知识能力 3

 5.4 业务任务适配能力 3

 5.5 性能与效率 3

 5.6 安全与合规 3

 5.7 可靠性与可维护性 3

6 评估方法 4

 6.1 评估方式 4

 6.2 指标计算与评分 4

 6.3 等级评定 4

 6.4 评估工具 4

7 评估流程 4

 7.1 评估准备 4

 7.2 评估实施 5

 7.3 评估结果与报告 5

 7.4 评估复评与改进 5

 7.5 评估质量保证 5

8 评估结果与应用 5

 8.1 评估结果 5

 8.2 结果应用 5

 8.3 评估档案 5

附录 A （规范性） 业务适配能力评估指标示例 6

参 考 文 献	7
---------------	---

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国西部开发促进会提出并归口。

本文件起草单位：国网四川省电力公司泸州供电公司、国网四川省电力公司电力科学研究院、成都筑川科技有限公司等。

本文件主要起草人：程晗、向思屿、朱川等。

引 言

随着人工智能（AI）快速发展，大模型依托出色的语义理解、内容生成和逻辑推理能力，正日益广泛应用于电力行业的发电、输电、变电、配电、用电及调度等业务场景。业务适配能力不仅决定其行业落地成效，更是电力企业选型与验收的关键依据。

近年来，国内电力企业相继发布面向电力行业的大模型，并在设备运维、故障诊断、负荷预测、调度决策和客户服务等业务场景开展规模化应用。然而，由于缺乏统一的评价尺度，不同大模型及大模型服务的业务适配能力难以横向比较，给电力企业的选型和落地带来不确定性。

为统一电力行业大模型业务适配能力的评估要求，规范评估指标、评估方法和评估流程，引导大模型在电力行业的科学选型、规范应用和持续优化，特制定本文件。

电力行业大模型业务适配能力评估技术规范

1 范围

本文件规定了电力行业大模型业务适配能力评估的指标体系、方法、流程和结果应用。

本文件适用于电力行业大模型的业务适配能力评估，涵盖基础大模型、行业大模型、场景大模型及基于大模型的智能应用。同时，本文件也适用于大模型的选型、验收和运营改进。

注：业务适配能力即大模型结合电力行业知识、业务数据与安全合规要求，满足特定电力业务任务需求、实现预期业务目标的能力。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 36572 电力监控系统网络安全防护导则

GB/T 41867-2022 信息技术 人工智能 术语

GB/T 45288.1-2025 人工智能 大模型 第1部分：通用要求

GB/T 45288.3 人工智能 大模型 第3部分：服务能力成熟度评估

3 术语和定义

GB/T 45288.1-2025界定的术语和定义适用于本文件。

4 总体要求

4.1 评估原则

开展电力行业大模型业务适配能力评估，应遵循下列原则：

- 科学性原则：评估指标、评估方法和评估流程应科学合理，评估依据充分、过程可追溯、结果可复现；
- 客观性原则：评估应采用可量化、可验证的客观数据和方法，不应受主观因素影响；
- 业务导向原则：评估应围绕电力业务实际需求，以业务任务完成质量为核心评价标准；
- 安全合规原则：评估应关注大模型应用中的数据安全、内容安全和网络安全，符合电力行业安全防护要求；
- 可操作性原则：评估指标应定义明确、数据可获得、操作可行；
- 动态性原则：评估应反映大模型技术和电力业务的动态发展，评估指标和评估方法应适时更新。

4.2 评估目标

电力行业大模型业务适配能力评估应实现下列目标：

- 客观反映大模型对电力业务任务的理解、处理与完成能力；
- 识别大模型在电力业务应用中存在的短板和风险；
- 为大模型的选型、验收、运营改进和迭代优化提供依据。

4.3 评估对象与范围

4.3.1 评估对象

评估对象包括电力行业应用的基础大模型、行业大模型、场景大模型及基于大模型的智能应用。评估对象应符合GB/T 45288.1-2025的规定。

4.3.2 评估范围

评估应涵盖发电、输电、变电、配电、用电和调度等电力业务领域，包括但不限于设备运维、故障诊断、负荷预测、调度决策、客户服务、安全生产监督等业务场景。

4.4 评估等级

电力行业大模型业务适配能力分为五个等级，由低到高依次为L1级、L2级、L3级、L4级、L5级。评估等级划分示例见附录A。

4.5 评估组织与人员

4.5.1 评估组织

评估应由具备相应能力的评估机构或专家组组织实施；评估组织应建立评估质量管理体系。

4.5.2 评估人员

评估人员应具备电力业务知识和人工智能相关专业知识。

4.6 评估条件

开展评估应具备下列条件：

- a) 评估对象已部署或可通过接口访问，具备开展测试和验证的条件；
- b) 评估数据应覆盖典型业务场景，真实准确且经脱敏处理；评估数据宜包含标准数据、业务数据和实际运行数据，覆盖业务知识、业务规范和典型用例。数据管理员应做好版本管理、标注与质量检查；
- c) 评估方案经相关方确认，评估目的、范围和等级明确。

5 评估指标体系

5.1 指标构成

5.1.1 业务适配能力评估指标体系由基础能力、电力业务知识能力、业务任务适配能力、性能与效率、安全与合规、可靠性与可维护性六类一级指标构成，每类一级指标下设若干二级指标。评估指标示例见附录A。

5.1.2 各类一级指标应根据评估目的和业务场景确定权重。业务任务适配能力和电力业务知识能力应作为核心指标；基础能力、性能与效率、安全与合规、可靠性与可维护性应作为基础性指标。

5.1.3 二级指标应根据评估场景进一步细化，形成可测量、可评价的具体指标。

5.2 基础能力

5.2.1 基础能力指标反映大模型的通用智能水平，应包括：

- a) 语义理解能力，包括对自然语言和专业文本的理解与归纳能力；
- b) 内容生成能力，包括文本、报告、方案等内容的生成质量；
- c) 逻辑推理能力，包括演绎推理、归纳推理和数值计算能力；
- d) 多轮交互能力，包括上下文理解与多轮对话的连贯性。

5.2.2 基础能力应通过通用评测集或标准测试集进行测试评价，作为电力业务适配能力评估的参考基准。机器学习、深度学习等人工智能相关术语应符合GB/T 41867的规定。

5.3 电力业务知识能力

5.3.1 电力业务知识能力指标反映大模型对电力行业知识的掌握与运用程度，应包括：

- a) 电力专业知识掌握程度，包括电力系统、设备、规程、标准等专业知识；
- b) 专业术语规范性，即使用电力行业规范术语的准确性；
- c) 业务场景理解能力，即对发电、输电、变电、配电、用电和调度等场景的理解能力；
- d) 制度规程熟悉程度，即对电力行业管理制度和运行规程的掌握程度。

5.3.2 电力业务知识能力是衡量大模型业务适配能力的基础，评估应重点关注专业知识的准确性、术语的规范性以及对电力安全规程的遵循情况。

5.4 业务任务适配能力

5.4.1 业务任务适配能力指标反映大模型完成特定电力业务任务的能力，应包括：

- a) 任务完成度，即业务任务结果与预期目标的符合程度；
- b) 任务准确率，即业务任务结果的正确比例；
- c) 微调适配效果，即通过微调提升业务任务效果的程度；
- d) 提示词适配效果，即通过提示词优化提升业务任务效果的程度。

5.4.2 业务任务适配能力是业务适配能力评估的核心，评估应结合典型业务任务的实际完成情况进行，突出业务导向。

5.5 性能与效率

5.5.1 性能与效率指标反映大模型在业务应用中的运行性能，应包括：

- a) 响应时间，即业务请求从发出到返回结果的时间；
- b) 吞吐量，即单位时间内处理的业务请求数量；
- c) 资源消耗，即模型运行所需的算力、存储等资源占用；
- d) 并发处理能力，即同时处理多路业务请求的能力。

5.5.2 性能与效率指标的判定应结合电力业务的实际应用要求。

5.6 安全与合规

5.6.1 安全与合规指标反映大模型应用的安全性及合规性，应包括：

- a) 数据安全，即业务数据在采集、传输、存储和使用过程中的安全防护能力，应符合GB/T 36572的规定；
- b) 内容安全，即生成内容不存在违法违规、虚假有害信息的保障能力；
- c) 网络安全，即大模型服务部署和运行过程中的网络安全防护能力；
- d) 合规性，即符合电力行业安全管理及数据管理要求的程度。

5.6.2 安全与合规指标应作为一票否决性指标，评估对象存在重大安全隐患的，不应评为L4级及以上等级。

5.7 可靠性与可维护性

5.7.1 可靠性与可维护性指标反映大模型应用的稳定运行与持续优化能力，应包括：

- a) 稳定性，即长时间连续运行的稳定程度；
- b) 鲁棒性，即面对异常输入、噪声数据时的容错能力；

- c) 可追溯性，即评估对象输出结果的可解释、可追溯程度；
- d) 可维护性，即大模型更新、升级和运维的便捷程度。

5.7.2 可靠性与可维护性应结合大模型的长期运行情况和运维记录进行评价。

6 评估方法

6.1 评估方式

6.1.1 业务适配能力评估应综合采用下列方式：

- a) 功能测试，通过标准化测试集验证基础能力和电力业务知识能力；
- b) 性能测试，通过模拟业务负载验证响应时间、吞吐量等性能指标；
- c) 安全性测试，验证数据安全、内容安全和网络安全防护能力；
- d) 专家评审，组织电力业务专家对大模型的专业性和准确性进行评审；
- e) 场景验证，在典型业务场景中对评估对象进行试运行验证。

6.1.2 功能测试、性能测试和安全性测试应结合评估对象的实际部署方式和业务场景设计测试方案，专家评审和场景验证应覆盖典型业务任务和关键业务场景。

6.2 指标计算与评分

6.2.1 各评估指标应采用百分制或等级制进行量化评分。

6.2.2 综合评分应采用加权求和方式，权重可采用层次分析法、专家咨询法等确定；综合评分的计算口径和评分规则宜在评估方案中明确。

6.2.3 各指标的评分方法应在评估方案中明确。

6.2.4 评分过程中宜记录评估数据、评分依据和计算过程。

6.3 等级评定

6.3.1 等级评定宜参照GB/T 45288.3的规定，并结合电力业务适配的实际情况确定。

6.3.2 评估等级应依据综合评分结果确定。

6.3.3 等级评定宜参照GB/T 45288.3的规定，结合电力业务特点，由评估机构或专家组对综合评分与等级进行认定。

6.3.4 评估等级可作为大模型选型和验收的重要参考依据。

6.4 评估工具

6.4.1 评估宜采用自动化测试工具与人工评审相结合的方式开展。

6.4.2 功能测试、性能测试和安全性测试应采用标准化的测试工具和测试平台，测试工具应具备测试用例管理、结果记录和报告生成等功能。

7 评估流程

7.1 评估准备

评估准备应包括下列内容：

- a) 明确评估目的、评估对象、评估范围和评估等级；
- b) 制定评估方案，确定评估指标体系、指标权重和评分方法；
- c) 准备评估数据、测试环境和评估工具；
- d) 组建评估组，明确人员分工和职责。

7.2 评估实施

7.2.1 评估实施应包括下列内容：

- a) 按照评估方案开展功能测试、性能测试和安全性测试；
- b) 组织开展专家评审和场景验证；
- c) 记录评估过程数据，保存评估证据；
- d) 对异常情况和偏差进行分析与复核。

7.2.2 评估实施过程中应对测试结果进行复核，对存疑结果应重新测试或由专家组评议确认。

7.2.3 评估实施应在受控环境下进行且不应受外部因素影响。

7.3 评估结果与报告

评估结束后应形成评估报告。评估报告应包括评估对象、评估范围、评估指标体系、评估方法、评估过程、评估结果、等级评定结论及改进建议等内容。

7.4 评估复评与改进

当大模型发生重大版本更新、微调或业务场景发生重大变化时，应重新开展评估。评估对象应根据评估结论进行针对性改进，并可通过复评验证改进效果。

7.5 评估质量保证

评估机构应建立评估质量保证机制，包括测试工具的校准、评估人员的培训、评估结果的复核和评估档案的管理等。

8 评估结果与应用

8.1 评估结果

8.1.1 评估结果应以评估报告形式呈现，并附评估指标得分表、等级评定结论及相关证据材料（评估报告应经评估机构审核后发布）。

8.1.2 评估结果应客观、真实、完整。

8.2 结果应用

评估结果宜应用于下列方面：

- a) 大模型选型，为电力企业选择适用的大模型及大模型服务提供依据；
- b) 验收评价，作为大模型项目验收和评价的依据；
- c) 运营改进，指导大模型的迭代优化和业务适配能力提升；
- d) 监督管理，为电力行业大模型应用的监督管理提供参考。

8.3 评估档案

评估过程中形成的评估方案、评估数据、测试记录、评分记录和评估报告等资料，应按有关规定归档保存。

附 录 A
(规范性)
业务适配能力评估指标示例

A.1 电力行业大模型业务适配能力评估指标示例见表A.1。

表 A.1 业务适配能力评估指标示例

一级指标	二级指标	指标说明
基础能力	语义理解、内容生成、逻辑推理、多轮交互	反映大模型对自然语言和专业文本的理解、生成与推理能力
电力业务知识能力	专业知识掌握、专业术语规范、业务场景理解、制度规程熟悉	反映大模型对电力专业知识、术语和业务场景的掌握程度
业务任务适配能力	任务完成度、任务准确率、微调适配效果、提示词适配效果	反映大模型完成特定电力业务任务的能力
性能与效率	响应时间、吞吐量、资源消耗、并发处理能力	反映大模型在业务应用中的运行性能
安全与合规	数据安全、内容安全、网络安全、合规性	反映大模型应用的安全性与合规性
可靠性与可维护性	稳定性、鲁棒性、可追溯性、可维护性	反映大模型应用的稳定运行与持续优化能力

A.2 电力行业大模型业务适配能力评估等级划分示例见表A.2。

表 A.2 评估等级划分示例

等级	等级名称	等级说明
L1级	初步适配	具备基本能力，业务任务完成度较低，尚不能满足电力业务应用要求
L2级	基础适配	能满足基础业务任务，在知识、性能或安全等方面存在较多短板
L3级	良好适配	能满足主要业务任务，性能和安全基本达到电力业务应用要求
L4级	优秀适配	能高质量完成业务任务，具备较强的安全性与可靠性
L5级	卓越适配	全面满足电力业务需求，业务适配能力达到行业领先水平

A.3 本附录所列指标和等级为示例，评估机构应结合具体评估目的和业务场景，对评估指标、指标权重和等级划分进行适当调整。

参 考 文 献

- [1] GB/T 22239 信息安全技术 网络安全等级保护基本要求
 - [2] GB/T 35273 信息安全技术 个人信息安全规范
-