

# 国家标准《人工智能 服务器系统性能测试规范》

## （征求意见稿）编制说明

### 一、工作简况

根据国家标准化管理委员会 2022 年下达的国家标准制修订计划，全国信息技术标准化技术委员会启动了《人工智能 服务器系统性能测试规范》（立项号：20221794-T-469）的制定工作。本标准为自主制定标准，起草单位为中国电子技术标准化研究院、华为技术有限公司、中国科学院软件研究所、北京航空航天大学、浪潮电子信息产业股份有限公司等，归口单位为全国信息技术标准化技术委员会（SAC/TC28）。

标准计划下达后，标准起草组进行了在全国信标委组织领导下，按计划开展本标准制定工作，具体内容如下：

2023 年 2 月，标准工作组正式成立，并开展了广泛调研和资料收集工作。

2023 年 4 月 26 日，完成开题。

2023 年 6 月，中国电子技术标准化研究院组织召开第一次标准研讨会，标准起草组对标准现有草案的编制思路和具体内容提出了修改意见，同时吸纳更多单位参与到标准编制工作中。会后按照分工编制形成工作组讨论稿 V1.0。

2023 年 9 月，面向标准工作组征求意见。形成标准讨论稿 V1.1，增加了大模型相关测试负载的测试方法及要求。

2023 年 11 月，进一步完善标准内容，针对大模型测试负载进行优化，针对通用模型进行更新及优化。

2023 年 12 月，标准进行内部审查，专家建议修改完善引言等内容，同时建议修改附录 E 为“测试系统示例”，条款 6.3 中添加主要性能参数列表，优化条款层级，完善测试要求等。

### 二、标准编制原则和确定主要内容的论据及解决的主要问题

[如技术指标、参数、公式、性能要求、试验方法、检验规则等的论据，包括试验、统计数据，解决的主要问题。修订标准时应列出与原标准的主要差异和水平对比]

#### 1. 标准编制原则

该标准的编制原则包含两个方面：

（1）该标准涉及相关方众多，鼓励人工智能相关单位参与，以确保本标准

规定的内容科学合理，具有普适性。

(2) 该标准是《国家新一代人工智能标准体系建设指南》中的总体基础性标准,对人工智能规划与建设具有重要的基础支撑作用。该标准应充分借鉴国际、国内相关先进研究成果，与国家人工智能相关政策导向相一致。

## 2. 标准主要内容

该标准规定了人工智能服务器系统（含 AI 服务器、AI 服务器集群、AI HPC 计算设施等）的性能测试方法。该标准适用于 AI 服务器系统（含 AI 服务器、AI 服务器集群、AI HPC 计算设施等）的性能评估。

内容框架如下：

### 1. 测试类型

- a) 开放测试
- b) 封闭测试

### 2. 训练

#### a) 训练过程：

b) 测试控制要素（含准确率门限、机器学习框架、混合精度训练、训练结果精度要求、数据读入要求、数据预处理过程、数据遍历要求、目标模型要求、超参选取规则、优化算法技术要求、损失函数技术要求、操作系统及环境要求等）；

c) 结果符合性（含元数据、结果模型一致性、训练代码要求、日志要求等）

d) 场景（含通用及行业专用场景）；

e) 指标及获取方法（含通用指标、测试学相关指标、行业专用指标等）；

f) 测试工具要求（构架、功能要求、公平性保障要求）；

### 3. 推理

#### a) 推理过程：

b) 测试控制要素（含作业到达控制、准确率、机器学习框架、量化、重训练、数据预处理、数据后处理、数据遍历、数据缓存、作业延误及丢失控制、操作系统及环境要求等）；

- c) 结果(含元数据、模型一致性要求、代码技术要求、日志要求等);
  - d) 场景(含通用及行业专用场景);
  - e) 指标及获取方法(含通用指标、测试学相关指标、行业专用指标);
- 测试工具要求(构架、功能要求、公平性保障要求等)。

### 3. 标准解决的问题

当前具有代表性的通用 AI 测试基准、HPC 性能测试基准或服务器技术规范,如 MLPerf、AI Benchmark、benchcouncil、AI-HPL、Linpack、DAWNBENCH、T/CESA 1043-2019《面向深度学习的服务器规范》、GB/T 9813.3《计算机通用规范 第3部分:服务器》《人工智能芯片 面向云侧的深度学习芯片测试指标与测试方法》以及 AIIA DNN benchmark 等,在 AI 服务器系统性能测试方面,仍存在一些未解决的问题:

- 1) 通用服务器技术规范对 AI 服务器系统的性能测试的规定不深入,而且存在同质化倾向。在测试指标方面,仅端到端运行时间、能耗等不能准确反映 AI 服务器系统。系统内部运行时效能方面,尚不存在标准化的测试方法。
- 2) 通用 AI 性能测试,使用公众可获得的模型、数据集。它们与行业(如金融、医疗)实际使用的模型、数据有较大差异。通用测试的结果,对行业指导意义不强,行业业务在 AI 服务器系统计算设施上的运行效果,无直接评判依据。
- 3) 当前的测试基准对测试科学(test technology)理论的实践相对较薄弱。这些测试基准对性能的理解、解释及测试方法仅限于稳定状态运行时间,而未考虑真实运行环境及系统自身现实状态。

针对以上问题,该标准拟结合测试技术,在通用及行业应用两方面,研究并标准化测试方法、用例,达到较为全面、准确的测试效果。

### 三、主要试验[或验证]情况分析

目前参与编写的单位中,已有较多基于人工智能服务器系统产品的商用或研发,本标准所规定的内容经过与各大厂商的对接和沟通,已被证明确实可行。

### 四、知识产权情况说明

本标准不涉及专利和知识产权问题。

### 五、产业化情况、推广应用论证和预期达到的经济效果

当前，测试科学发展已较完善，在软件、硬件、系统性能测试方面，其内涵和测试手段、指标多样。而当前的测试基准或多或少地存在：1) 与测试科学联系相对不紧密；2) 测试指标不全面；3) 非专门针对 AI 服务器系统；4) 行业/领域适用性、指导性不强；5) 同质化等问题。测试科学的发展与当前测试基准的结合有很大空间。从技术上讲，测试技术理论趋于成熟，但未深入应用到 AI 服务器系统的实际测试中。二者的融合，将是未来 AI 服务器系统性能测试，及其它 AI 性能测试的主要发展方向之一。本标准将在这方面进行探索和实践，为未来面向行业用户的、全面科学的 AI 服务器系统性能测试奠定基础。

## 六、采用国际标准和国外先进标准情况

本标准为自主研发的国家标准，未采用国际标准和国外先进标准。

## 七、与现行相关法律、法规、规章及相关标准的协调性

本标准符合我国有关的现行法律、法规和相关强制性标准的相关规定，并与相关标准保持协调一致。

## 八、标准性质的建议

鉴于本标准的内容，建议该标准作为推荐性国家标准发布。

## 九、贯彻标准的要求和措施建议

为更好地发挥标准实施效力，建议采取以下措施推动本标准的贯彻实施：

1、政府积极推动：标准发布后，由国家市场监管总局、国家工信部指导，由全国信息技术标准化技术委员会牵头，组织召开标准发布通报会，并通过官方媒体在相关产业链和应用领域加强宣传，提升标准的宣传权威性和受众针对性；

2、制定配套实施政策：政府通过制定配套实施政策，如在招投标、市场准入、认证服务、企业对标达标等方面出台相关政策，加强标准的实施力度，引导和提升相关企业和单位应用标准积极性；

3、加强标准检测认证一体化：鼓励第三方依据和采用标准进行产品试验、检测和认证，开展产品国家标准符合性测试，同时推动建立政府和市场的多方标准质量采信机制，不断提升标准实施的效力。

## 十、替代或废止现行相关标准的建议

无。

## 十一、其它应予说明的事项

无。

国家标准《人工智能 服务器系统性能测试规范》起草工作组

2024年11月16日