

ICS

T/GXDSL

团 体 标 准

T/GXDSL 027—2025

人工智能算力芯片性能评估标准

Performance Evaluation Standards for Artificial Intelligence Computing Power Chips

意见征集稿

2025 - - 发布

2025 - - 实施

广西电子商务企业联合会 发布

目 次

前 言	III
一、引言	1
(一)背景	1
(二)目的	1
(三)适用范围	1
二、规范性引用文件	1
(一)国家标准	2
(二)行业标准	2
(三)法律法规	3
(四)国际标准	3
三、术语和定义	3
(一)人工智能算力芯片	3
(二)性能评估	4
(三)基准测试	4
四、评估指标体系	4
(一)计算能力	4
1. 浮点运算能力	4
2. 整数运算能力	5
(二)能效比	5
1. 功耗	5
2. 能效	5
(三)稳定性	6
1. 温度稳定性	6
2. 故障率	6
(四)兼容性	6
1. 软件兼容性	7
2. 硬件兼容性	7
五、评估方法	7
(一)基准测试	7

1. 测试环境	8
2. 测试程序	8
(二) 数据分析	8
1. 数据采集	8
2. 数据处理	9
(三) 评估报告	9
1. 报告内容	9
2. 报告格式	9
六、实施与监督	10
(一) 实施机构	10
(二) 监督机制	10
(三) 修订与更新	10
七、附则	10
(一) 标准解释	11
(二) 实施日期	11
(三) 其他	11

前 言

本文件依据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。

本文件由广西产学研科学研究院提出。

本文件由广西电子商务企业联合会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

人工智能算力芯片性能评估标准

一、引言

(一) 背景

人工智能（AI）技术的快速发展对计算能力提出了更高的要求，而算力芯片作为支撑 AI 应用的核心硬件，其性能直接影响 AI 系统的效率和效果。随着 AI 技术在图像识别、自然语言处理、自动驾驶等领域的广泛应用，算力芯片的需求日益增长。然而，由于市场上芯片种类繁多，性能差异较大，缺乏统一的评估标准，导致用户在选型和使用过程中面临诸多困难。因此，制定一套科学、全面、可操作的人工智能算力芯片性能评估标准势在必行。

(二) 目的

本标准旨在规范人工智能算力芯片的性能评估，提供一套科学、全面、可操作的评估方法，确保评估结果的公正性和可比性。通过明确评估指标、测试方法和实施流程，本标准将为芯片制造商、系统集成商和终端用户提供可靠的性能参考，促进技术进步和产业发展。

(三) 适用范围

本标准适用于各类人工智能算力芯片的性能评估，包括但不限于 GPU（图形处理单元）、TPU（张量处理单元）、FPGA（现场可编程门阵列）等。无论是用于数据中心的高性能计算芯片，还是用于边缘计算的低功耗芯片，均可参照本标准进行评估。

二、规范性引用文件

本标准的制定和实施参考了以下国家及行业标准、法律法规和技术规范。以下文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本标准。凡是不注日期的引用文件，其最新版本适用于本标准。

(一) 国家标准

GB/T 22239-2019 《信息安全技术 网络安全等级保护基本要求》

该标准规定了网络安全等级保护的基本要求，适用于人工智能算力芯片在网络安全环境中的性能评估和安全保障。

GB/T 25069-2010 《信息技术 安全技术 信息安全风险管理》

该标准提供了信息安全风险管理的框架和方法，适用于人工智能算力芯片在设计和应用中的风险评估和控制。

GB/T 35273-2020 《信息安全技术 个人信息安全规范》

该标准明确了个人信息处理的安全要求，适用于人工智能算力芯片在处理用户数据时的隐私保护。

GB/T 36342-2018 《信息技术 云计算 云服务级别协议规范》

该标准规定了云服务级别协议的基本内容和要求，适用于人工智能算力芯片在云服务环境中的性能评估。

GB/T 37732-2019 《信息技术 云计算 云服务计量指标》

该标准定义了云服务的计量指标和方法，适用于人工智能算力芯片在云计算环境中的性能量化评估。

(二) 行业标准

YD/T 2542-2019 《云计算服务协议参考框架》

该标准提供了云计算服务协议的参考框架，适用于人工智能算力芯片在云计算服务中的协议制定和执行。

YD/T 3756-2020 《人工智能云计算平台技术要求》

该标准规定了人工智能云计算平台的技术要求，适用于人工智能算力芯片在云计算平台中的技术实现和性能评估。

(三) 法律法规

1. 《中华人民共和国网络安全法》

该法律明确了网络运营者的安全责任和义务,适用于人工智能算力芯片在网络安全环境中的性能评估和安全保障。

2. 《中华人民共和国数据安全法》

该法律规定了数据处理活动的安全要求,适用于人工智能算力芯片在处理用户数据时的安全保障。

3. 《中华人民共和国个人信息保护法》

该法律明确了个人信息处理的法律要求,适用于人工智能算力芯片在处理用户数据时的隐私保护。

(四) 国际标准

ISO/IEC 27001:2013 《信息技术 安全技术 信息安全管理体系要求》

该标准提供了信息安全管理体系的框架,适用于人工智能算力芯片在信息安全管理中的性能评估。

ISO/IEC 27018:2019 《信息技术 安全技术 保护公有云中个人信息的操作规范》

该标准规定了公有云中个人信息保护的操作规范,适用于人工智能算力芯片在公有云环境中的隐私保护。

IEEE 754-2008 《浮点算术标准》

该标准定义了浮点数的表示和运算规则,适用于人工智能算力芯片浮点运算性能的评估。

三、术语和定义

(一) 人工智能算力芯片

指专门用于加速人工智能计算任务的集成电路芯片。这类芯片通常具有高并行计算能力和优化的内存架构,能够高效处理深度学习、机器学习等 AI 算法。

(二) 性能评估

指通过一系列测试和分析，对算力芯片的计算能力、能效比、稳定性等指标进行量化评价。性能评估的目的是为用户提供客观、准确的性能数据，帮助其选择适合的芯片产品。

(三) 基准测试

指使用标准化的测试程序和方法，对算力芯片的性能进行测量和比较。基准测试通常包括计算能力测试、功耗测试、温度测试等，以确保测试结果的可靠性和可比性。

四、评估指标体系

(一) 计算能力

计算能力是衡量算力芯片性能的核心指标，主要包括浮点运算能力和整数运算能力。

1. 浮点运算能力

定义：浮点运算能力是指芯片在单位时间内完成的浮点运算次数，通常用于衡量芯片在处理科学计算、图形渲染等任务时的性能。

评估方法：使用标准浮点运算测试程序（如 LINPACK）进行测试。LINPACK 是一种广泛使用的基准测试程序，能够测量芯片在解决线性方程组时的浮点运算性能。

指标：FLOPS (Floating Point Operations Per Second)，即每秒浮点运算次数。例如，某芯片的浮点运算能力为 10 TFLOPS，表示其每秒可完成 10 万亿次浮点运算。

2. 整数运算能力

定义：整数运算能力是指芯片在单位时间内完成的整数运算次数，通常用于衡量芯片在处理逻辑运算、数据压缩等任务时的性能。

评估方法：使用标准整数运算测试程序（如 Dhrystone）进行测试。Dhrystone 是一种经典的基准测试程序，能够测量芯片在整数运算任务中的性能。

指标：IOPS（Integer Operations Per Second），即每秒整数运算次数。例如，某芯片的整数运算能力为 5 GIOPS，表示其每秒可完成 50 亿次整数运算。

(二)能效比

能效比是衡量算力芯片在单位功耗下完成计算任务的能力，主要包括功耗和能效两个指标。

1. 功耗

定义：功耗是指芯片在满负荷运行时的功耗，通常用于衡量芯片的能源消耗情况。

评估方法：使用功率计测量芯片的功耗。测试时，芯片需在标准测试平台上运行满负荷任务，功率计实时记录功耗数据。

指标：瓦特（W）。例如，某芯片的功耗为 200W，表示其在满负荷运行时的功耗为 200 瓦。

2. 能效

定义：能效是指芯片每瓦特功耗所能完成的运算次数，通常用于衡量芯片的能源利用效率。

评估方法：计算 FLOPS 或 IOPS 与功耗的比值。例如，某芯片的浮点运算能力为 10 TFLOPS，功耗为 200W，则其能效为 50 GFLOPS/W。

指标：FLOPS/W 或 IOPS/W。例如，某芯片的能效为 50 GFLOPS/W，表示其每瓦特功耗可完成 500 亿次浮点运算。

(三) 稳定性

稳定性是衡量算力芯片在长时间高负荷运行下的性能表现，主要包括温度稳定性和故障率两个指标。

1. 温度稳定性

定义：温度稳定性是指芯片在长时间高负荷运行下的温度变化情况，通常用于衡量芯片的散热性能和可靠性。

评估方法：使用温度传感器监测芯片温度。测试时，芯片需在标准测试平台上运行满负荷任务，温度传感器实时记录温度数据。

指标：摄氏度（℃）。例如，某芯片在满负荷运行时的最高温度为 85℃，表示其在长时间高负荷运行下的温度稳定性较好。

2. 故障率

定义：故障率是指芯片在测试期间发生故障的频率，通常用于衡量芯片的可靠性。

评估方法：记录芯片在标准测试期间的故障次数。测试时，芯片需在标准测试平台上运行满负荷任务，记录故障发生的次数和时间。

指标：故障次数/小时。例如，某芯片在测试期间的故障率为 0.01 次/小时，表示其每小时发生故障的概率为 1%。

(四) 兼容性

兼容性是衡量算力芯片与软硬件平台的适配性，主要包括软件兼容性和硬件兼容性两个指标。

1. 软件兼容性

定义：软件兼容性是指芯片支持的主流 AI 框架和库，通常用于衡量芯片的软件生态适配性。

评估方法：测试芯片与 TensorFlow、PyTorch 等主流框架的兼容性。测试时，需在芯片上运行标准 AI 模型，记录运行结果和性能数据。

指标：支持框架数量。例如，某芯片支持 TensorFlow、PyTorch、Caffe 等主流框架，表示其软件兼容性较好。

2. 硬件兼容性

定义：硬件兼容性是指芯片与主流硬件平台的兼容性，通常用于衡量芯片的硬件适配性。

评估方法：测试芯片与不同主板、内存等硬件的兼容性。测试时，需将芯片安装在不同硬件平台上，记录运行结果和性能数据。

指标：兼容硬件平台数量。例如，某芯片兼容 Intel、AMD、ARM 等主流硬件平台，表示其硬件兼容性较好。

五、评估方法

(一) 基准测试

基准测试是评估算力芯片性能的核心方法，主要包括测试环境和测试程序两个方面。

1. 测试环境

硬件环境：标准测试平台，包括主板、内存、电源等。测试平台需符合国家相关标准，确保测试结果的可靠性和可比性。

软件环境：标准操作系统和测试软件。操作系统需为最新版本，测试软件需为标准化版本，确保测试结果的一致性。

2. 测试程序

浮点运算测试：使用 LINPACK 进行测试。LINPACK 是一种广泛使用的基准测试程序，能够测量芯片在解决线性方程组时的浮点运算性能。

整数运算测试：使用 Dhrystone 进行测试。Dhrystone 是一种经典的基准测试程序，能够测量芯片在整数运算任务中的性能。

功耗测试：使用功率计进行测量。功率计需符合国家相关标准，确保测量结果的准确性。

温度测试：使用温度传感器进行监测。温度传感器需符合国家相关标准，确保监测结果的可靠性。

(二) 数据分析

数据分析是评估算力芯片性能的重要环节，主要包括数据采集和数据处理两个方面。

1. 数据采集

采集方法：使用自动化测试工具采集数据。自动化测试工具需符合国家相关标准，确保数据采集的准确性和一致性。

采集频率：每 5 分钟采集一次数据。采集频率需根据测试任务的特点进行调整，确保数据的全面性和代表性。

2. 数据处理

数据清洗：去除异常值和噪声数据。数据清洗需采用科学的方法，确保数据的准确性和可靠性。

数据分析：计算各项指标的平均值、最大值、最小值等。数据分析需采用科学的统计方法，确保分析结果的客观性和准确性。

(三) 评估报告

评估报告是评估算力芯片性能的最终成果，主要包括报告内容和报告格式两个方面。

1. 报告内容

测试概述：测试目的、测试环境、测试程序等。测试概述需简明扼要，确保读者能够快速了解测试的基本情况。

测试结果：各项指标的测试数据和图表。测试结果需详细全面，确保读者能够全面了解芯片的性能表现。

结论与建议：对芯片性能的综合评价和改进建议。结论与建议需客观公正，确保读者能够根据报告内容做出科学决策。

2. 报告格式

格式要求：使用标准报告模板，确保格式统一。报告模板需符合国家相关标准，确保报告的专业性和规范性。

提交方式：电子版和纸质版同时提交。提交方式需符合国家相关规定，确保报告的及时性和有效性。

六、实施与监督

(一) 实施机构

广西电子商务企业联合会：负责标准的制定、修订和实施。广西电子商务企业联合会需组织专家团队，确保标准的科学性和权威性。

第三方检测机构：负责具体的测试和评估工作。第三方检测机构需具备国家认可的资质，确保测试结果的公正性和可靠性。

(二) 监督机制

定期检查：每年对实施情况进行一次全面检查。定期检查需由广西电子商务企业联合会组织，确保标准的有效实施。

投诉处理：设立投诉渠道，及时处理用户投诉。投诉处理需由广西电子商务企业联合会负责，确保用户的合法权益。

(三) 修订与更新

修订周期：每三年对标准进行一次修订。修订周期需根据技术发展和行业需求进行调整，确保标准的时效性和适用性。

更新程序：根据技术发展和行业需求，及时更新标准内容。更新程序需由广西电子商务企业联合会组织，确保标准的科学性和权威性。

七、附则

(一) 标准解释

本标准由归口广西电子商务企业联合会。解释内容需符合国家相关规定，确保标准的权威性和规范性。

(二) 实施日期

本标准自发布之日起实施。实施日期需根据国家相关规定进行调整，确保标准的及时性和有效性。

(三) 其他

本标准未尽事宜，参照国家相关标准和规定执行。其他内容需由广西电子商务企业联合会负责解释，确保标准的全面性和适用性。
