

T/ZSIA

浙江省软件行业协会团体标准

T/ZSIA 0001—2025

面向元宇宙场景的分布式人工智能平台 技术要求

Technological Requirements for Distributed Artificial Intelligence Platforms for
Metaverse Applications

2025 - 12 - 17 发布

2025 - 12 - 17 实施

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 基本要求	1
5 功能要求	2
6 性能要求	3
7 接口要求	4
8 部署要求	4
9 运维要求	4

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这些专利的责任。

本文件由浙江省软件行业协会提出并归口。

本文件起草单位：浙江省软件行业协会、浙江大学、浙江华锐捷技术有限公司、杭州华橙软件技术有限公司、浙江省邮电工程建设公司、浙江和达科技股份有限公司、汇智智能科技有限公司、杭州小影创新科技股份有限公司、嘉兴职业技术学院、浙江湔瀚信息科技有限公司、杭州露电数字科技集团有限公司、创业慧康科技股份有限公司。

本文件主要起草人：吴超、王小号、王朝、缪其恒、丁林超、林秀闹、曹滢锋、陆军波、韩晟、娄在凤、王公韬、王辉、郭晨、李瑜娟。

本文件为首次发布。

面向元宇宙场景的分布式人工智能平台 技术要求

1 范围

本文件规范了面向元宇宙场景的分布式人工智能平台（以下简称“平台”）的基本要求、技术规范、性能要求、接口规范、部署要求和运维要求。

本文件适用于相关平台的规划、设计、研发和应用。涉及的场景要求，若无特殊说明，则为通用要求。

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 32399—2024 信息技术 云计算 参考架构
GB/T 41867—2022 信息技术 人工智能 术语
GB/T 44271—2024 信息技术 云计算 边缘云通用技术要求
T/AIIA 004—2023 人工智能企业等级评定规范

3 术语和定义

GB/T 45993—2025界定的及以下术语和定义适用于本文件。

3.1

元宇宙场景 metaverse scenario

为实现特定用户使用场景（例如社交、娱乐、办公、教育、消费等），依托元宇宙平台的技术支撑，通过组合数字资产，设定交互规则而构建的，具有具体时空范围和明确主题的沉浸式虚拟环境。

3.2

分布式人工智能平台 distributed artificial intelligence platform

基于分布式架构，由异构计算资源池、分布式任务调度系统、模型训练与推理引擎等核心组件构成的软硬件系统集成。该平台融合了人工智能能力，支持跨节点的模型训练、推理、资源管理与优化。

3.3

模型训练 model training

通过机器学习算法对模型参数进行迭代优化，以提升模型性能的过程。

3.4

模型推理 model inference

指利用经过训练的模型，对新输入的数据进行计算与分析，并输出预测或决策结果的过程。

3.5

边云协同 edge-cloud collaboration

一种云端与边缘端在资源、数据及服务上实现逻辑统一与深度融合的计算架构模式。

4 基本要求

4.1 应用场景

平台应具备适配多种元宇宙应用场景的能力。应能支持跨场景应用，可提供专属的优化和扩展。

4.2 平台架构

平台应基于分布式架构设计，支持微服务化和容器化部署。部署方式应支持中心化部署、边缘部署以及云边端协同部署，以适应不同元宇宙场景对计算、时延和带宽的需求。

4.3 兼容性与开放性

4.3.1 应支持主流人工智能框架的模型训练和推理。平台宜提供标准化接口，便于异构硬件和不同应用场景的适配和扩展。

4.3.2 异构资源适配：平台应具备异构资源的接入和统一管理的能力，应支持对国产异构计算设备、智能网卡和高速互联设备的统一调度与管理。

5 功能要求

5.1 数据要求

5.1.1 数据管理

数据管理应符合以下要求：

- a) 应支持对多模态数据的采集与接入，包括但不限于来自物理世界的传感器数据和来自虚拟世界的交互日志、仿真数据等；
- b) 应支持分布式数据存储，宜兼容结构化与非结构化数据。应提供数据全生命周期管理能力，包括数据版本控制、元数据管理等；
- c) 应提供数据清洗、标注、预处理和增强功能。宜支持通过生成式模型等技术进行数据增强，以扩充训练样本；
- d) 应提供数据分级分类管理机制。宜支持联邦学习、差分隐私等技术，在模型训练和推理过程中保护数据隐私。

5.1.2 空间与资产数据支持

空间与资产数据支持应符合以下要求：

- a) 应支持稀疏或稠密点云、环境深度图、空间锚点及 6DoF 位姿等空间数据；
- b) 应支持主流 3D 格式的解析、版本控制与轻量化处理。

5.2 模型要求

5.2.1 模型训练

模型训练应符合以下要求：

- a) 应支持数据并行、模型并行等多种模型训练策略，兼容自动混合精度训练、梯度压缩等优化技术；
- b) 应支持同步与异步训练模式，具备负载均衡与容错能力；
- c) 宜支持边云协同环境下的训练任务自动迁移与资源智能调度。

5.2.2 模型压缩与优化

模型压缩与优化应符合以下要求：

- a) 应提供模型量化和结构剪枝能力，以减小模型体积和计算复杂度。宜支持自动化策略搜索，在精度损失可控的前提下最大化压缩率；
- b) 宜支持知识蒸馏技术，允许将大型、复杂模型的知识迁移至小型模型，以适应边缘和终端设备的部署需求；
- c) 宜支持计算图优化、算子融合等技术，优化模型推理性能。

5.2.3 元宇宙专用算法与模型

元宇宙专用算法与模型应符合以下要求：

- a) 应支持 3D 生成式模型训练与推理能力，支持通过文本或图像描述自动生成 3D 模型、材质贴图及场景布局；
- b) 宜支持神经辐射场、3D 高斯泼溅等神经渲染模型；

- c) 应支持多模态驱动模型；
- d) 应支持 3D 生成式模型训练与推理能力，支持通过文本或图像描述自动生成 3D 模型、材质贴图及场景布局；
- e) 宜支持神经辐射场、3D 高斯泼溅等神经渲染模型；
- f) 应支持多模态驱动模型，在实时推理条件下生成数字人的口型、面部表情及肢体动作。

5.3 应用要求

5.3.1 边云协同

边云协同应符合以下要求：

- a) 应支持对计算、存储、网络等资源的动态调度与弹性扩展能力；
- b) 应实现边缘与云端的资源统一管理，支持任务迁移、负载均衡和故障恢复，支持平台的高可用性和弹性。

5.3.2 元宇宙场景适应性

元宇宙场景适应性应符合以下要求：

- a) 应提供虚实世界数据的双向映射和转换机制。应保证虚实场景在关键时空节点的一致性对齐；
- b) 应支持虚实环境下异构计算资源的统一调度。应为高实时性要求的虚实交互任务提供服务质量（QoS）保障机制；
- c) 宜支持跨场景的数据共享和知识迁移，实现模型在不同但相关的元宇宙应用场景间的快速适应；
- d) 平台应具备计算资源与渲染资源的协同调度能力。应支持异构任务（如模型推理与图形渲染）的并发处理与高效数据交互，以满足元宇宙场景的实时性需求；
- e) 在边云协同场景下，平台宜具备基于用户感知（如视域或感兴趣区域）的智能流送能力。应支持根据用户交互状态动态调整数据传输策略与内容精度，优化网络带宽利用率。

5.4 迭代要求

平台迭代应符合以下要求：

- a) 应对已部署模型的性能、精度、资源消耗等进行持续监控，并建立评估指标体系；
- b) 宜支持模型的增量学习和在线学习，利用新数据对模型进行快速迭代优化，以适应元宇宙场景的动态变化。

6 性能要求

6.1 并发处理和响应延迟

并发处理和响应延迟应符合以下要求：

- a) 单台边缘计算设备应支持不少于 12 个模型的并发推理；
- b) 在云边协同场景下，根据任务交互实时性要求区分延迟规范，应符合以下要求：
 - 1) 实时交互任务（例如实时渲染协同、虚实同步）：核心业务的数据传输平均延迟应 $\leq 50\text{ms}$ ，端到端平均响应时间应 $\leq 100\text{ms}$ ；
 - 2) 非实时任务（例如离线推理、背景内容生成）：端到端平均响应时间应 $\leq 500\text{ms}$ 或支持根据业务需求进行自定义配置。

注：单台边缘计算设备的典型配置可参考如下指标：

- 1) 中央处理器：架构不限 x86 或 ARM，配置不低于 8 核，主频 $\geq 2.0\text{ GHz}$ ；
- 2) AI 加速单元：具备独立的 AI 计算加速能力，单卡显存或专用内存 $\geq 16\text{ GB}$ ，INT8 算力 $\geq 70\text{ TOPS}$ 或 FP16 算力 $\geq 15\text{ TFLOPS}$

6.2 系统可用性

平台可用性应符合以下要求：

- a) 平台核心服务的年均可用率应 $\geq 99.9\%$ ；

b) 平台应支持至少 100 个边缘节点的接入管理。

注：系统可用率=（年度总时长-年度不可用时长）÷年度总时长×100%

7 接口要求

7.1 API 接口

API接口应符合以下要求：

- a) 应提供模型训练、推理全流程的标准化 API，包括任务提交、参数配置、过程监控和结果管理；
- b) 应支持高性能模型推理服务接口，具备同步或异步推理、模型版本管理与负载均衡能力。

7.2 应用扩展接口

应用扩展接口应符合以下要求：

- a) 应提供标准化的插件机制和开发套件，允许开发者扩展平台功能，并提供插件的生命周期管理能力；
- b) 宜支持用户自定义算子的开发、注册和优化接口，以适配特定的硬件或算法；
- c) 应提供模型转换接口，支持不同人工智能框架间的模型格式转换。

7.3 权限与安全接口

应具备完善的用户认证、权限管理、审计和安全监控接口，支持多租户资源隔离与计费管理。

8 部署要求

8.1 硬件环境应符合以下要求：

- a) 平台应支持异构计算，包括高性能张量处理器、图形处理器、神经网络处理器、智能处理器、智能网卡等多种加速硬件；
- b) 应具备大内存、高带宽低时延互联的服务器与分布式存储设施，支持多样化边缘设备部署。

8.2 软件环境应符合以下要求：

- a) 应兼容主流操作系统，支持各类依赖组件与驱动的版本管理；
- b) 应具备数据加密、防护、入侵检测等安全机制。

8.3 应提供自动化的模型部署流水线，支持从模型仓库到目标环境（云、边、端）的一键式部署。

8.4 应支持多种部署策略，如蓝绿部署、金丝雀发布等，以保证服务更新的平滑过渡和稳定性。

8.5 应提供统一的配置管理能力，支持对不同环境下模型服务的参数进行集中管理和动态更新。

8.6 应支持容器化部署，并兼容主流的容器编排系统，实现资源的弹性伸缩和高效管理。

9 运维要求

9.1 日常运维

日常运维应符合以下要求：

- a) 应提供对计算、存储、网络等关键资源的全面监控功能，并具备灵活的告警配置能力；
- b) 应提供标准化的日志收集、检索和分析功能，宜支持日志可视化；
- c) 应提供性能分析工具，辅助运维人员进行瓶颈检测和资源优化；
- d) 宜提供智能化的故障诊断能力和自动化的应急响应预案，应支持数据备份和恢复机制。

9.2 安全运维

安全运维应符合以下要求：

- a) 应实现完善的访问控制和用户认证体系，支持基于角色的权限管理；
- b) 应提供全方位的数据安全保护机制，包括传输加密、存储加密和密钥管理等；
- c) 应支持对关键操作和系统活动的日志记录与审计，确保安全事件的可追溯性；
- d) 应建立漏洞管理流程，支持定期的漏洞扫描、风险评估和补丁管理；

- e) 应具备内容安全与溯源机制，支持对敏感数据及模型生成内容添加数字水印，并具备水印提取与验证能力，实现数据流向监控与内容的全链路溯源。

全国团体标准信息平台