

大模型智算软件栈OGAI

朱红

浪潮信息 AI应用架构师

大模型：生成式AI的核心技术

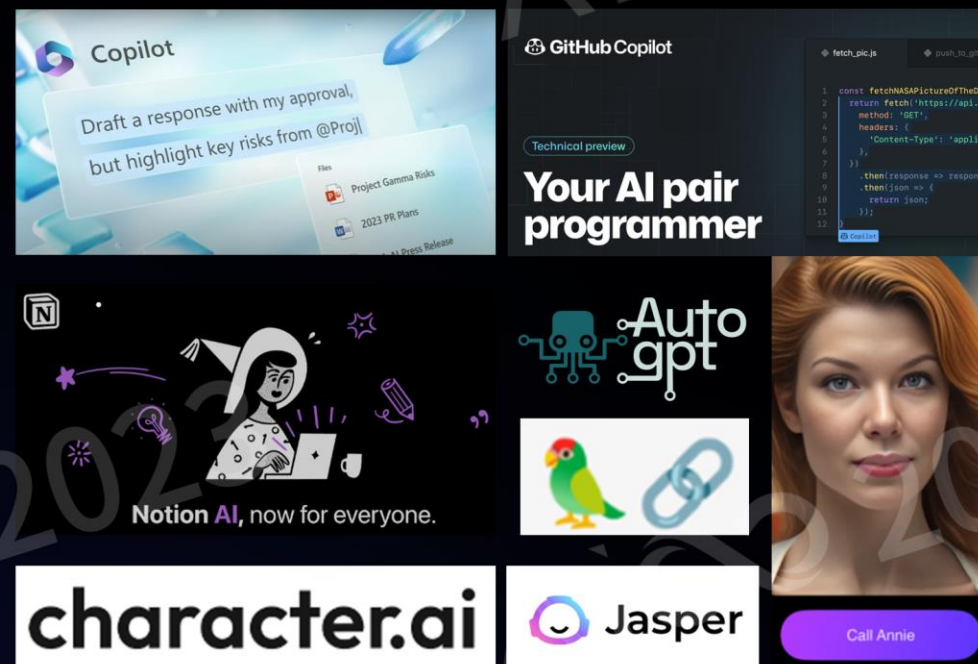
创新层出不穷

应用快速落地

Open Source



Closed Source



大模型：研发应用的需求与挑战

大模型研发

- 亟需好用的公共算力服务来迅速开展工作
- 数据怎么清洗的？怎么评估数据质量？
- 在大模型训练的过程中有哪些坑？如何避免？
- 怎么解决大模型分布式训练性能不佳的问题？
-

大模型应用

- 商业模型与开源模型能力表现各有专长，如何抉择？
- 从基础模型到领域模型的实践之路依然困难重重，行业know-how、领域知识库构建、私有化部署
-

我们的目标是确保年底前大模型在公司的业务中上线应用

-----某科技公司CTO

AI集群就像是一台动力强劲的法拉利，但目前为止我们还没有很好的掌握如何驾驭它。

-----某科技公司IT架构总监

无法连续训练超过1-2天...许多故障需要手动检测和修复，浪费计算资源和研究人员的时间

-----Meta OPT模型研发日志

迫切需要保障大模型生产力的AI Infra

源大模型的实践



- 数千颗GPU芯片智算集群部署
- PB级数据爬取、筛选、分类.....
- 千卡性能优化
- GPU失效、网卡失效管理
- 大规模计算输出不稳定、loss爆炸
-

服务客户的实践



- CUDA初始化失败、GPU掉卡
- NCCL通信性能低
- GPU direct RDMA 未使能
- RoCE网络用不起来、不稳定
- 分布式任务的环境配置复杂、易出错
- 集群性能上不去
-

- **系统全栈**：涉及多团队、多环节、多流程，需要系统性的架构和全栈的解决方案
- **兼容适配**：有很多的开源工具、组件、但存在很多的兼容和性能方面的问题
- **性能优化**：性能很重要，一个大的集群系统涉及的组件数以千计，如何优化性能很挑战

ChatBots

AI Copilot

Agents

.....

大模型智算软件栈OGAI



系统环境部署



算力调度保障



模型开发管理

大模型算力系统

高性能AI服务器



高吞吐并行存储系统



低延迟RDMA网络



- **全栈全流程**: 提供从集群环境搭建到算力调度、大模型开发的全栈软件; 覆盖大模型从数据处理到预训练和微调到多模型管理的整个研发流程
- **算力充分释放**: 通过从BIOS选项优化到大规模算力调度策略优化的多尺度、多层次的性能优化实践, 充分释放AI集群性能
- **实战验证提炼**: 来自浪潮信息自身研发和服务客户的实战经验, 兼容适配业界广泛使用的主流开源系统、工具、框架和软件

L4 多模纳管

MModel

多模型管理与服务平台，帮助客户更好的管理和评估模型，加速模型的部署和应用

L3 模型工具

YLink

经过验证的数据治理、大模型预训练和微调开发工具链，降低大模型开发和落地的门槛

L2 调度平台

AIStation

商业化的面向大模型开发的成熟AI算力调度平台产品。具备大模型断点续训能力，保证长时间持续训练

L1 系统环境

PODsys

开源、高效、兼容、易用的智算集群系统环境部署方案，实现自动化部署和弹性扩展，提高系统可用性和扩展性

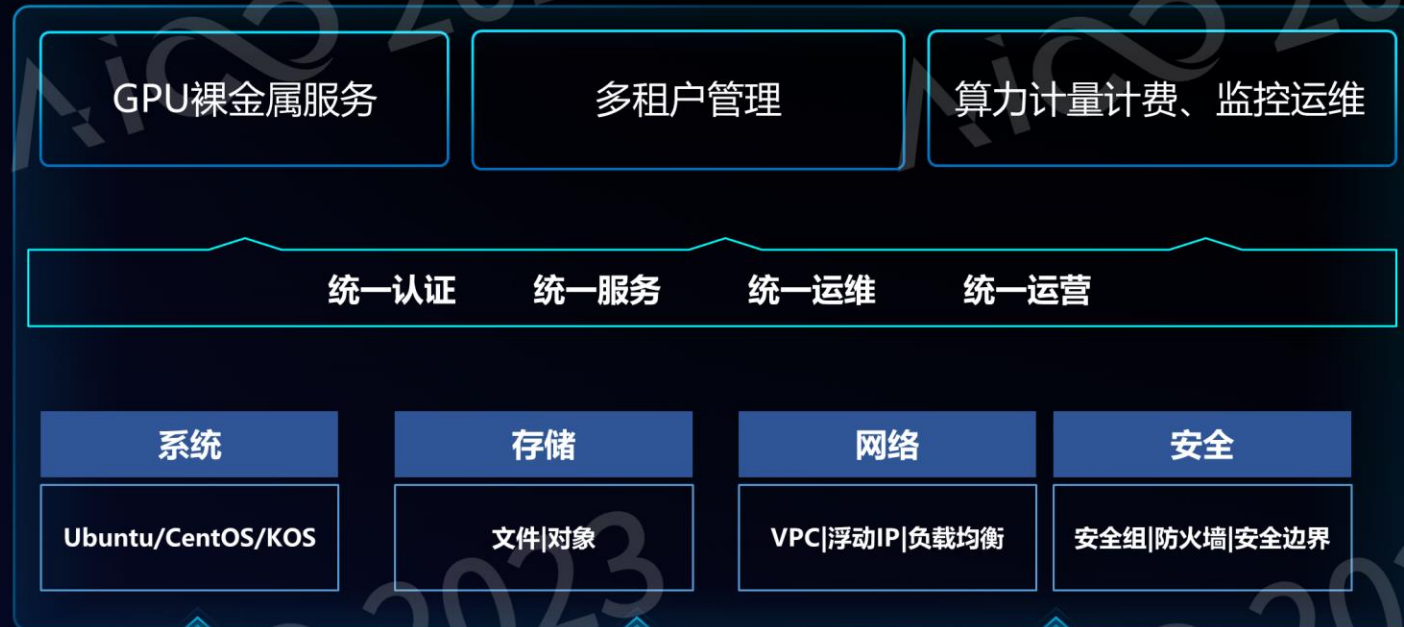
L0 基础设施

智算中心OS

提供多租户、裸金属的AI算力运营运维支撑平台

 高性能AI服务器 高吞吐并行存储系统 低延迟RDMA网络

面向大模型算力服务的智能算力运管平台，满足多租户以裸金属为主的弹性AI算力运管需求



裸金属服务

- 统一化算力平台，简便、灵活获取GPU裸金属服务
- 快速批量部署，分钟级部署上千规模裸金属服务，按需进行弹性扩展
- 异构计算芯片、IB、RoCE高速网络、高性能存储等高性能计算环境一键获取
- 计算、网络、数据隔离保障业务安全

多租户管理

- 多租户统一门户管理，简化管理，算力服务快速交付
- 细粒度租户算力配额管理，定制化算力供需

商业级运管能力

- 完善、准确的算力计量、计费
- 专业、全面的监控、运维
- 支持算力租赁服务

开源、高效、兼容、易用的智算集群系统环境部署方案

开放获取：访问 <http://podsys.ai> 下载安装包、脚本和使用文档

资源调度



Kubernetes

GPU operator

whereabouts operator

SR-IOV operator

Multus

系统监控



Prometheus

Cluster Nodes Network
CPU GPU Memory Disk

UI

Grafana

环境配置

File
accessNFSv4
Over
UDP/TCP/
RDMAParallel
File systemBeeGFS
Lustre
GPFSUser
management

NIS/LDAP

Automation
ToolsPDSH PassSSH
Puppet AnsibleGPU
ecosystemCUDA NCCL CUDNN OFED
NVIDIA Fabric-manager NGC
GDR/GDS moduleContainer
ToolkitDocker
NV-Docker

系统部署

Operating System

Ubuntu
22.04/20.04CentOS
8.6/7.9KeyarchOS
KOS 5.8

Driver

InfiniBand HCA 200G/400G

RoCE HCA NIC 200G/400G

Data Center/ Tesla GPU

H800/H100/H20/L40s/L20

全面覆盖：提供从OS、驱动到系统监控可视化、资源调度等集群系统环境要素全面覆盖

广泛兼容：选用最稳定和广泛兼容的软件版本

自动化部署：通过一系列脚本工具简化部署过程，缩短算力上线周期

专家服务：向企业用户提供实施安装服务和集群性能校准服务

传统方式部署AI集群

- 操作系统安装
- 安全性 & 访问控制
- GPU驱动 & IB驱动安装
- CUDA & Fabric Manager安装
- Docker & NVIDIA Container
- Toolkit安装
- 深度学习框架
- 资源调度
- 性能优化
- 管理软件和监控工具

- 数千颗GPU芯片智算集群部署
- 千卡性能优化
- GPU失效、网卡失效管理
- 分布式任务的环境配置复杂、易出错

- CUDA初始化失败、GPU掉卡
- NCCL通信性能低
- GPU direct RDMA 未使能
- RoCE网络用不起来、不稳定
- 集群性能上不去

PODsys方式部署AI集群

- ✓ PODsys OS 安装
- ✓ 监控及资源调度模块部署（可选）
- ✓ 性能基准测试

面向大模型开发的商业化人工智能算力调度平台

算法开发

LLaMA/SD/MPT...

Jupyter
WebShell

常规训练

TF/PyTorch

大规模预训练

DeepSpeed
Megatron-LM

自动故障定位

断点自动续训

算力调度

IB、RoCE适配

数据缓存

弹性训练

批调度

Gang调度

紧急队列

高性能AI服务器



高吞吐并行存储系统



低延迟RDMA网络



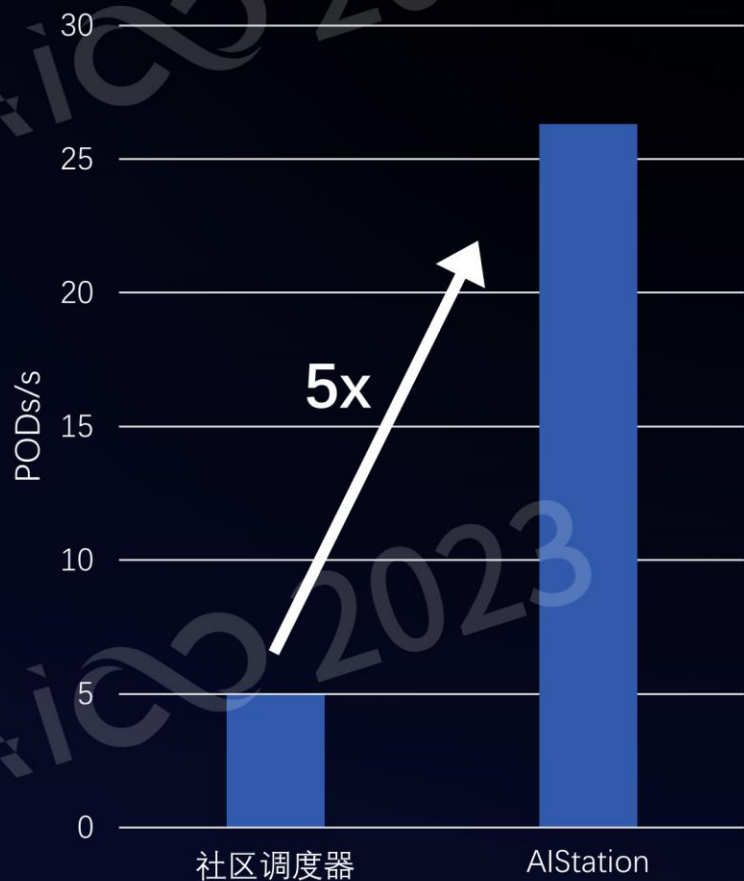
训练异常快速定位，断点自动续训

- 快速定位芯片、网卡、通讯设备异常或故障
- 全局训练暂停保持，热备算力自动弹性替换，健康节点快速Checkpoint读取，断点自动续训
- 训练全生命周期监管和异常全自动化处理，实现无人自动化

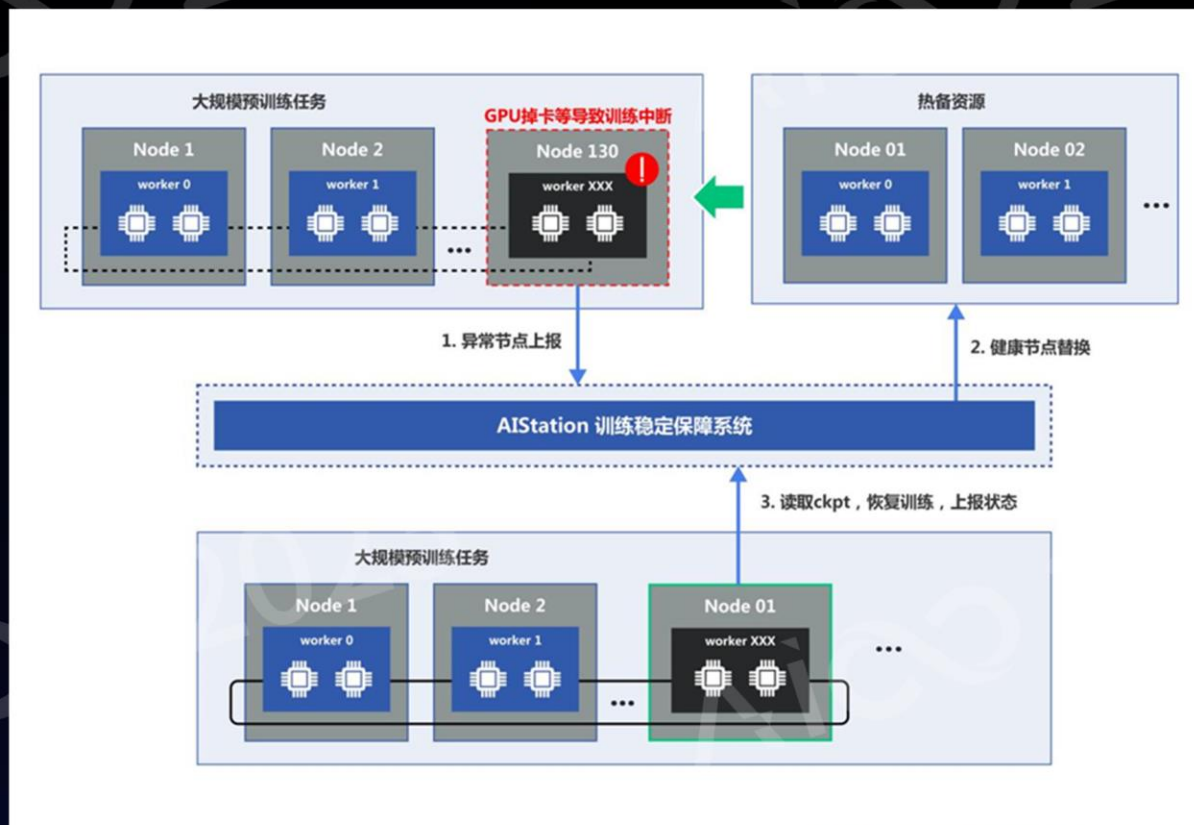
简化复杂网络适配，灵活高效使用

- 兼容IB、RoCE等复杂集群组网环境，解决开源调度版本对RoCE网络无损透传、灵活资源配比的使用难题
- 针对大规模训练场景，设备故障自动容错，保证大模型训练长时间高效、稳定运行

大规模POD调度任务性能对比



大规模预训练任务的异常处理和断点续训



面向大模型数据治理、预训练、微调的高效工具链

大模型微调



数据处理工具包 Y-DataKit

- DataGen: 数据生成, 基于主题与大语言模型APIs调用, 生成高质量微调数据集
- FileQA: 数据抽取, 基于预设问题抽取文档信息, 构建用于模型微调的问答数据

- Gather: 数据处理, 从Web、知识文档等3种安全数据源进行数据采集
- Transform: 格式转换, 对PDF、Docx等8种格式数据及数学公式进行转换
- Purify: 数据清洗, 提供内容主题、标点、分词、段落等28种数据内容清洗能力

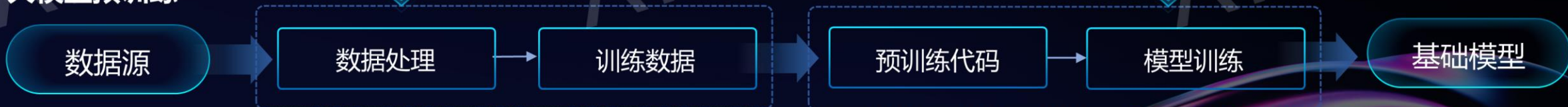
大模型微调工具包 Y-FTKit

- LMFlow*: 模型微调工具, 兼容主流模型, 提供参数高效微调工作流程
- Prompter*: 提示词优化工具, 为大模型提供自动化的提示词工程

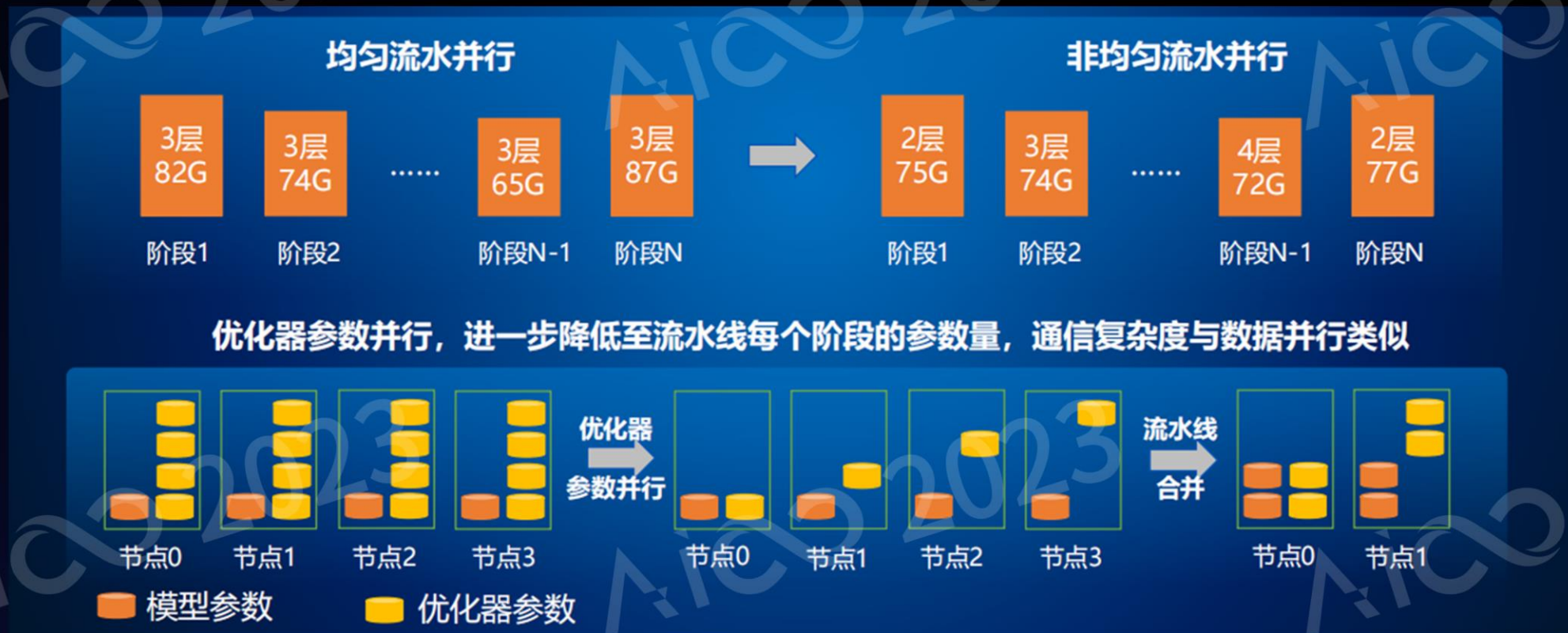
大模型训练工具包 Y-TrainKit

- Megatron*: 分布式训练框架, 用于训练Transformer架构的大语言模型
- DeepSpeed*: 分布式训练框架, 提高大规模模型训练的效率和可扩展性

大模型预训练



优化Megatron-LM的3D并行策略，源2.0算力效率47%@BF16



构建非均匀流水策略，有效缓解流水线头部与尾部的内存瓶颈，让模型在流水并行各阶段的内存占用量分布更均衡；结合优化器参数并行，降低内存需求，进而降低流水线路数。

提供多模型接入、服务、评测等功能的纳管平台

可视化

低代码/无代码

数据集

模型纳管

评测



数据管理

支持数据上传与分享，内置多维度模型评估数据



数据安全

支持数据权限隔离，私域数据集加密



数据接入

自定义接入任务与基础评测数据集与测试方法



模型接入

提供Hugging Face模型下载与适配功能



模型服务

支持多种主流开源模型与闭源模型APIs服务



模型管理

支持多类型、多版本大模型适配与纳管



基准评测

支持语言、知识、推理、多学科等模型评测方案



性能评测

支持推理延迟、稳定性测试，支持测试报表生成



自定义测试

支持多路模型面向任务与性能的横向测评

全面的测试数据支撑

丰富的主流模型支持

多维度模型能力评测

The logo for Aico 2023, featuring the word "Aico" in a white sans-serif font followed by a stylized infinity symbol in blue and white. The background is dark blue with abstract, glowing purple and blue wave-like patterns in the corners.

Aico

感谢聆听