

众智FlagOS开源社区介绍

面向多种AI芯片的统一、开源系统软件栈



为什么统一、开源的系统软件至关重要？

• 当前AI创新面临的挑战

- 大模型算法验证成本极高：单次验证可能耗资数百万美元；
- 单一组织无法覆盖所有环节：数据、基准、工具链等准备成本过高；
- 计算资源昂贵且难以获取：多数研发团队难以承担高昂成本。

碎片化软件生态阻碍AI芯片大规模产业落地

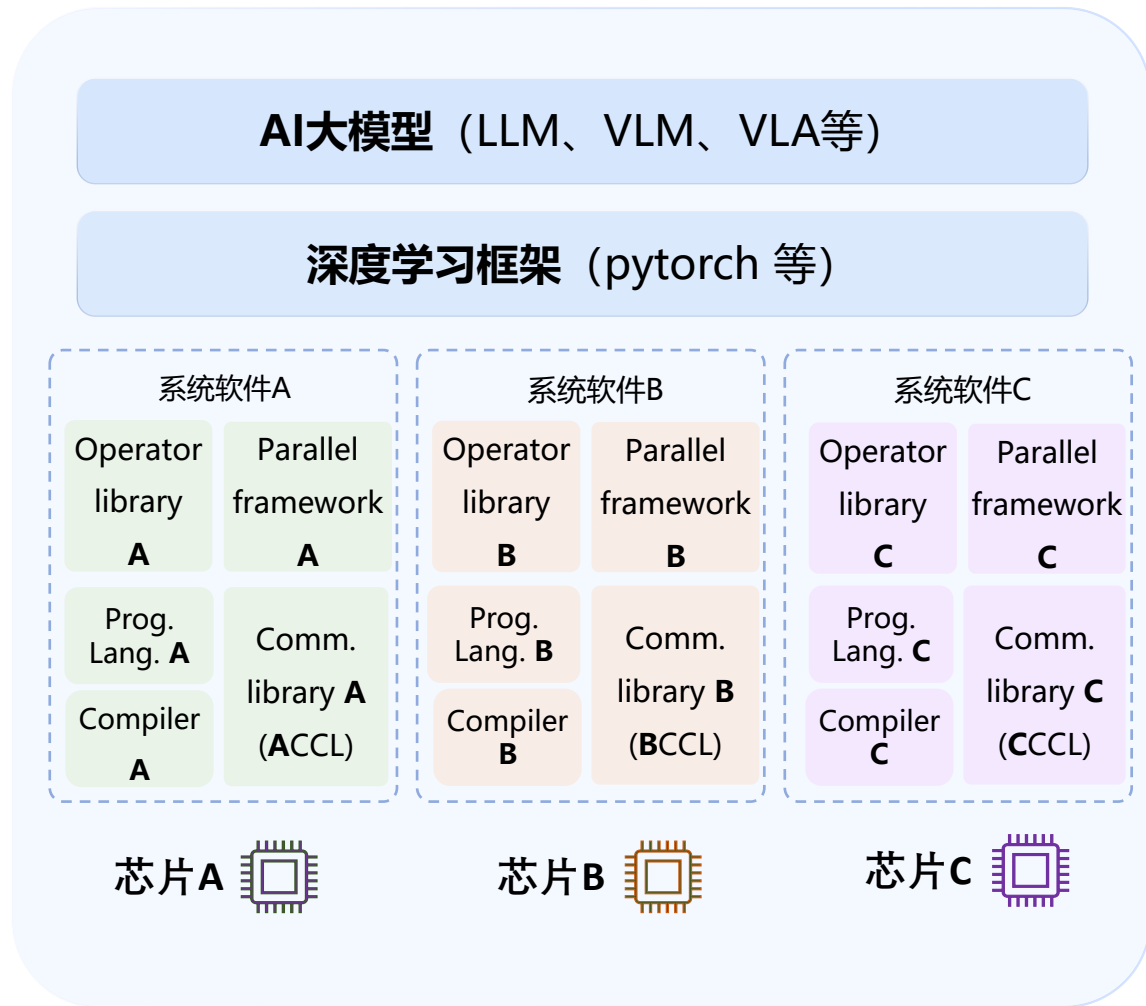
各厂商专有生态割裂：用户切换芯片需重写代码，迁移成本极高

- 芯片公司能力有限:

算子库支持不；
难以跟上模型和优化技术的高速迭代

- 芯片间无兼容性:

迁移需投入大量研发资源



众智FlagOS: 一路向前，众智成城

众智FlagOS
Open Software for AI Systems

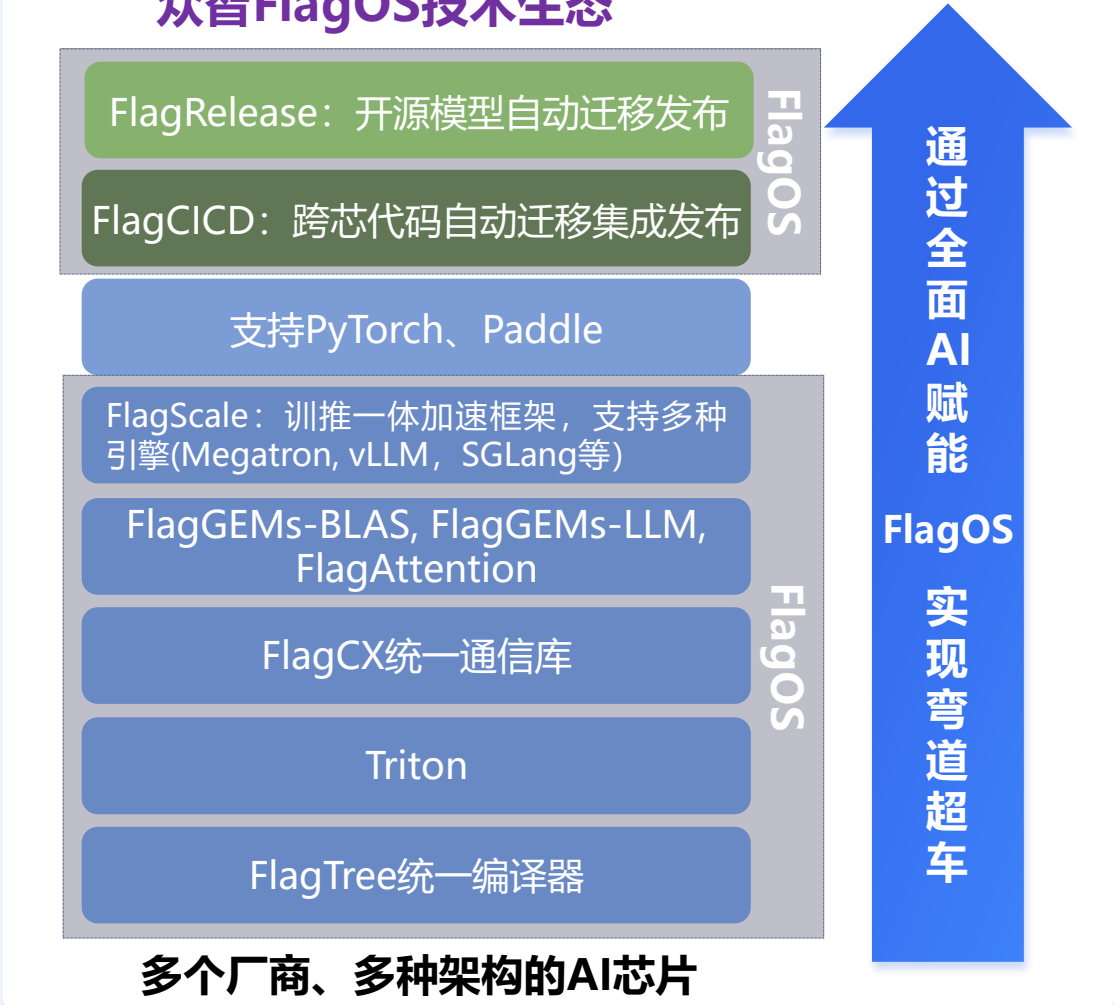
以技术攻关打破AI芯片生态壁垒，共筑开放计算的创新蓝图



CUDA技术生态



众智FlagOS技术生态



面向多种AI芯片的统一、开源系统软件栈

众智FlagOS v1.5 发布

支持从大模型到具身智能

全面支持更多芯片20+，支持PyTorch与Paddle两大框架

- FlagGEMs超过200个核心算子
- FlagTree统一编译器支持12家芯片
- FlagScale框架实现真正的跨框架、跨芯片、跨后端的统一能力
- FlagCX统一通信库支持五种通信协议

AI赋能全面提速自动化和开发效率

- 发布Triton-Copilot：实现端到端算子生成、验证及优化，支持多种芯片后端。
- FlagRelease支持了主流开源模型在11款芯片的持续发布，通过FlagOS+AI agent提升自动化迁移效率4倍。

性能全栈提升

- 端到端推理性能跃迁：在重点模型上获得比原生Triton的3倍提速。逼近CUDA最佳性能。
- FlagScale框架自动调优：训练可自动加速超过36%，推理加速20%。
- FlagCX通信库的pipeline优化提速2.3倍

新架构、新系统、新场景

- 支持更新模型架构：除支持Transformer，新增 RWKV 和 Diffusion 系列模型支持
- 支持新系统架构：支持浪潮、海光等超节点新架构
- 支持具身智能全链路：支持大脑、小脑（VLA）模型从预训练、后训练到量化推理全面覆盖；支持端云协同 + 工具检索

众智FlagOS v1.5：面向多种AI芯片的全栈开源技术

AI大模型
(语言大模型, 多模态大模型, MoE架构等)

深度学习框架
(PyTorch, PaddlePaddle, etc)

已支持大模型

语言模型

DeepSeek, Qwen,
Seed-oss, GPT-oss,
Step, Grok, Llama等

多模态模型

智源EMU, 面壁CPM,
Qwen-VL系列, ERNIE4.5,
Llava系列

具身智能模型

智源RoboBrain
Pai-0

众智FlagOS v1.5：面向多种AI芯片的系统软件栈

开源核心库

FlagGEMs:
通用大模型算子库

FlagScale:
训练推理并行框架

FlagTree:
统一编译器

FlagCX:
统一通信库

开源工具

Triton-Copilot
算子自动生成工具

FlagRelease
自动迁移和发版工具

FlagPerf
多芯片评测工具

各种大模型

各种深度学习框架

统一自主软件栈：统一支持

各种 AI 芯片

各种的智算集群

后端编译器 A
底层通信库 A

后端编译器 B
底层通信库 B

后端编译器 C
底层通信库 C

芯片 A



芯片 B



芯片 C



已支持的硬件架构: Nvidia, NPU, GPGPU, DSA, RISC-V AI, ARM

芯片企业：超过10家芯片企业，20多款不同芯片

数据中心
(train & Inference)

机器人
(cloud-edgecooperation)

边缘
(inference)

众智FlagOS v1.5

共享全球支持芯片厂商最多的AI系统生态

FlagGEMs算子库

华为、寒武纪、海光、
昆仑芯、摩尔线程、
清微智能、燧原、算
能、ARM周易、
ARM、英伟达等**16**
家厂商**25**款芯片。



FlagTree编译器

华为、寒武纪、海光、
昆仑芯、摩尔线程、
清微智能、算能、
ARM周易、英伟达
等**12**家厂商**20**款
芯片

FlagScale框架

华为、寒武纪、海光、
昆仑芯、摩尔线程、
清微智能、英伟达等
9家厂商多款芯片。

FlagCX通信库

华为、寒武纪、海光、
昆仑芯、摩尔线程、
英伟达等**8**家厂商多
款芯片。

FlagRelease自动 发版平台

华为、寒武纪、海光、
昆仑芯、ARM、英
伟达等**8**家厂商**12**
款芯片。

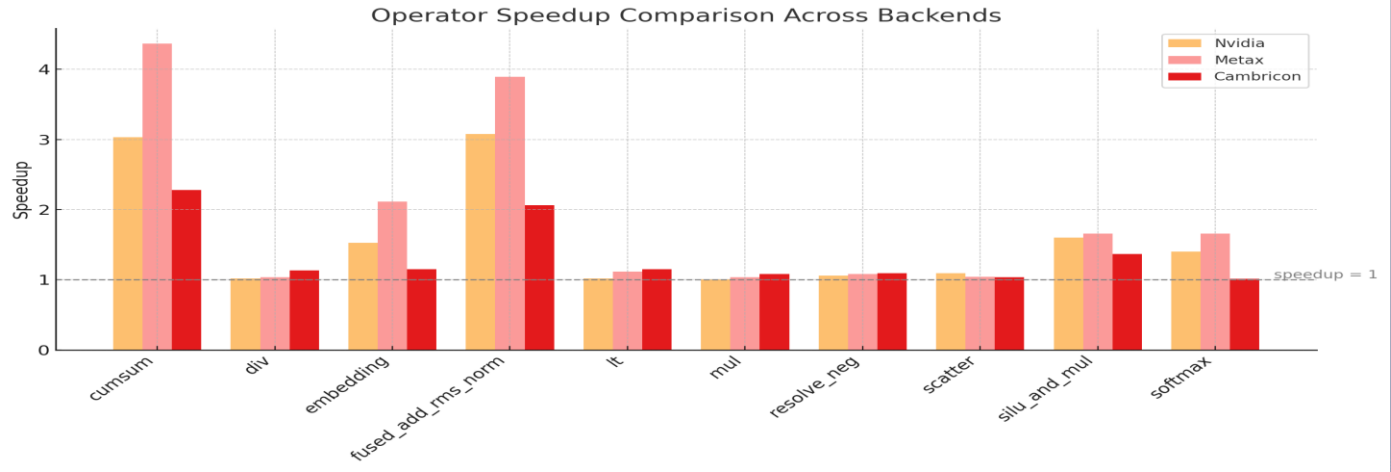


FlagOS 核心价值

➤ 提升推理场景覆盖度和性能：在英伟达芯片上，与原生Triton实现相比，性能有大幅提升，并逼近了CUDA性能

模型名称	覆盖度	性能	
		优化前	优化后
Qwen2.5-7B-Instruct	100%	35%	95%
Qwen3_30A_A3B	100%	25%	92%

➤ 通用算子库FlagGEMs在不同架构的芯片上，都获得性能收益



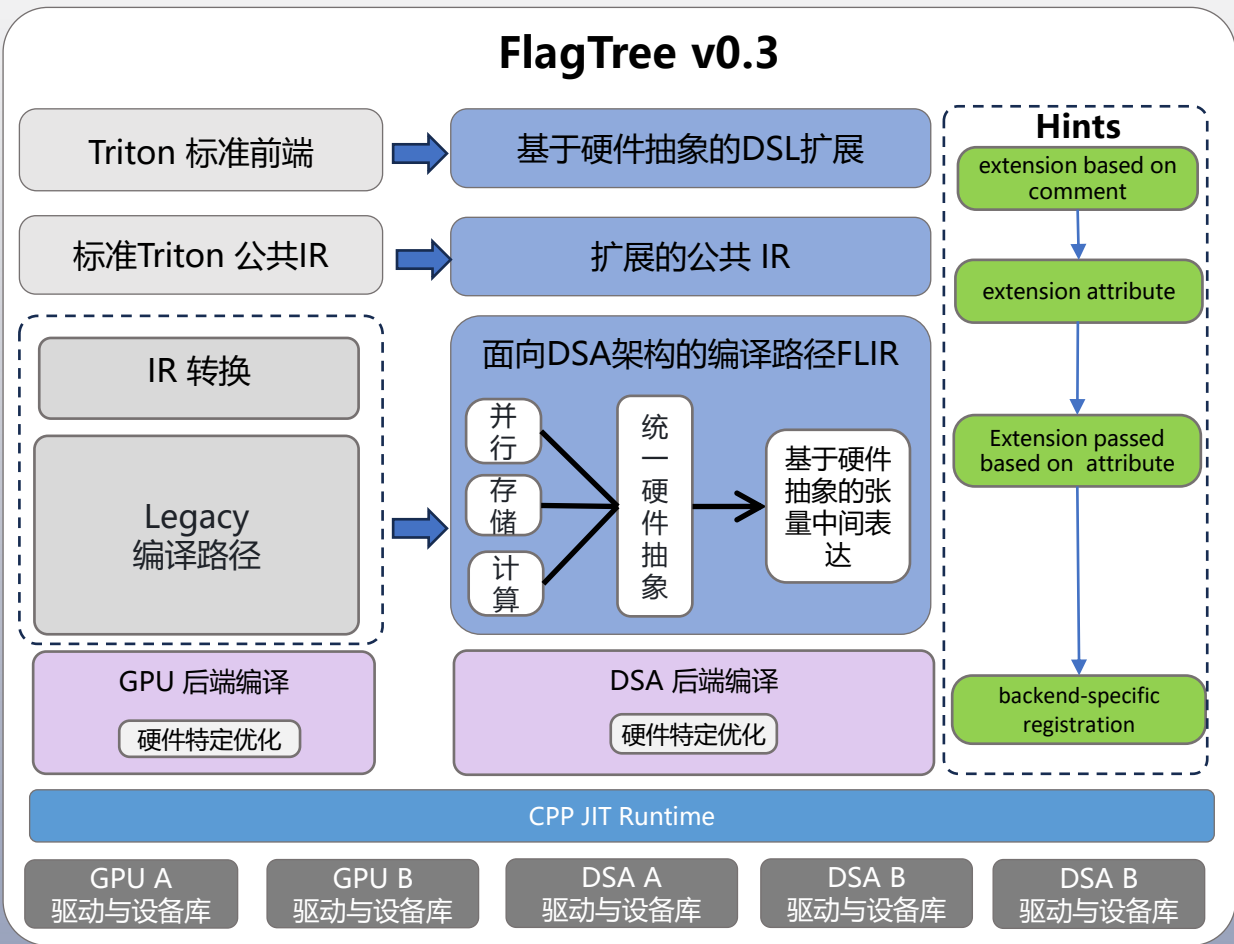
➤ PreTune 离线搜寻最优内核配置

问题：为了满足每次request的变长输入的优化，Triton使用online AutoTune，造成推理耗时陡增

解法：设计高效的离线搜寻机制，建立Shape与内核参数对应关系，online 直接查表获取内核参数，降低搜索耗时

收益案例：应用于 Qwen2.5-7B-Instruct，端到端推理性能提升**40%**。

FlagTree v0.3 正式发布：累计支持12家芯片厂商的20种芯片，通过研发通用优化技术，进一步提升性能



➤ 统一支持了多种芯片的 Triton编译，性能没有明显下降。相反，在DSA芯片上已经初现性能的提升。

表1 FlagTree与厂商原生 Triton编译器专用版本性能差异

DSA-1	DSA-2	GPU-1	GPU-2	GPU-3	GPU-4
2.77%	0.2%	0.55%	-0.62%	-1.01%	-2%

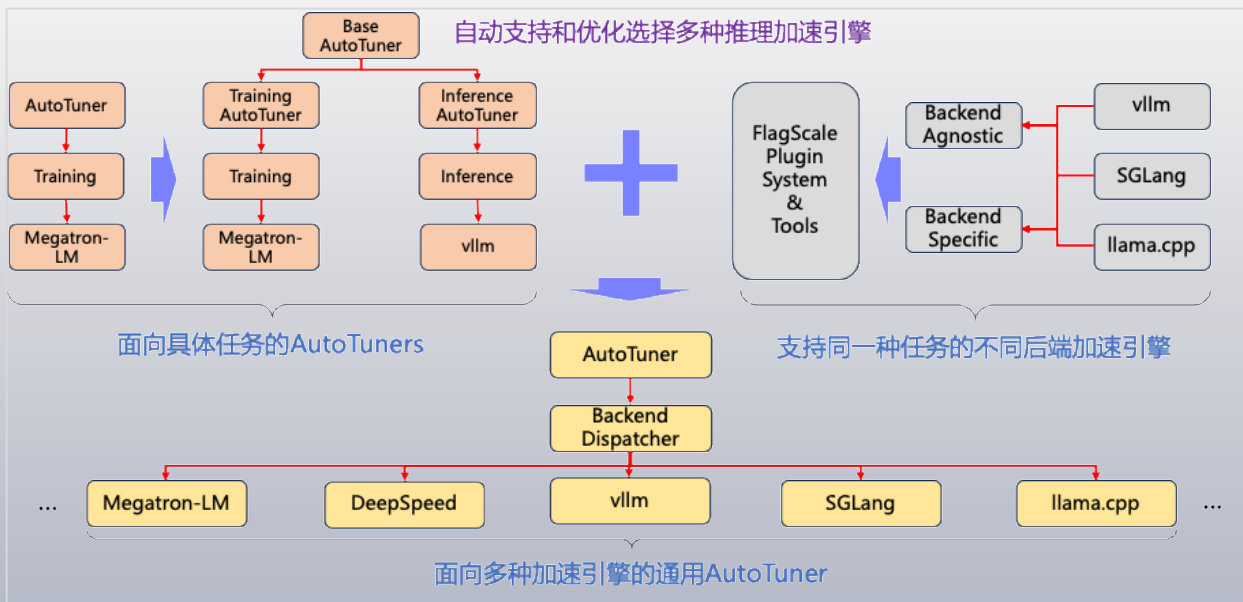
注：在超过130个Triton算子上进行测试

➤ Hints：基于硬件感知的编译优化技术
通过注释嵌入硬件优化提示，挖掘硬件潜力，提升算子性能，已在英伟达、华为昇腾和Arm China AIPU 构建编译链路，其中部分重点算子在华为昇腾提速10%以上

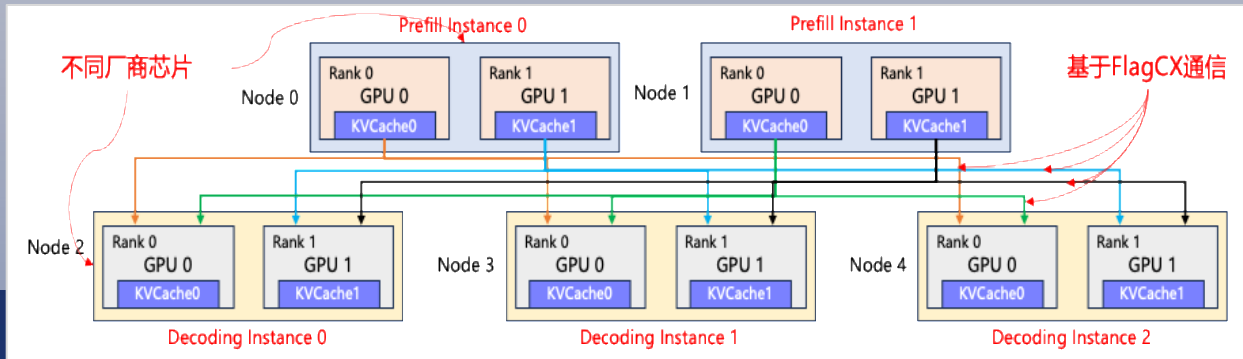
➤ CPP JIT Runtime提升算子速度
基于CPP语言的运行时包装机制，降低运行时开销，应用于20余个算子，平均性能提升20%以上；该技术已支持英伟达、天数和华为昇腾

训推框架FlagScale：自动获得训练、推理效率提升

- **全面自动调优**：支持多种芯片、多种后端，覆盖训练与推理的自动调优，支持同构集群、异构集群的训练和推理的自动调优。



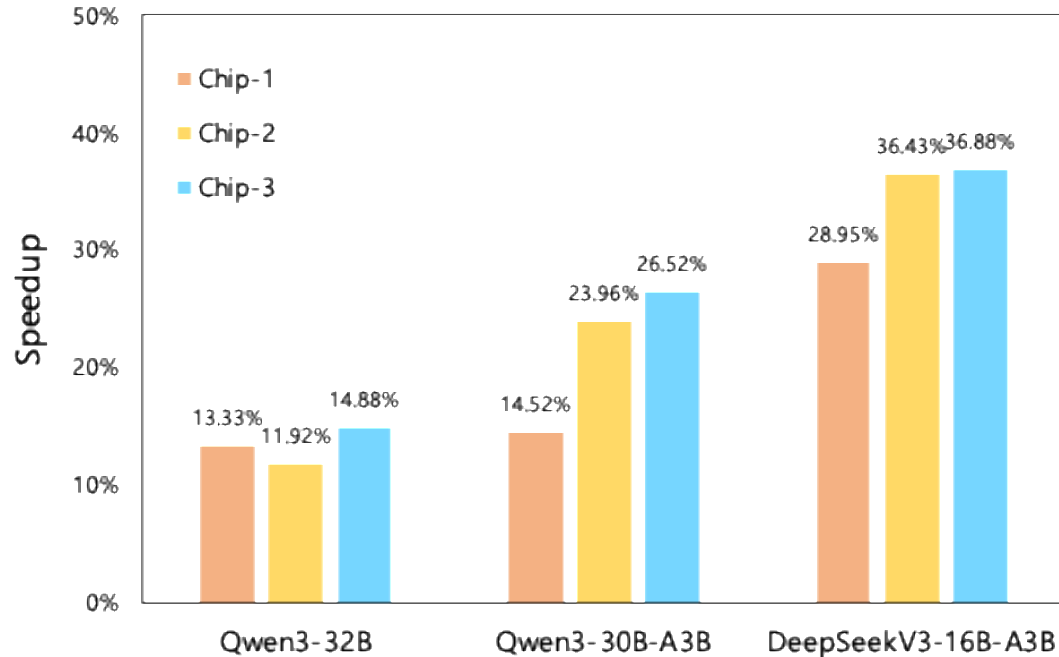
- **全面异构训练推理支持**：支持异构PD分离技术，提升集群推理效率



全面提升国产、混合集群的易用性和效率

- 通过**训练自动调优**最大加速比达到约**36.88%**（去年为23.08%），平均提升幅度约为**23%**（去年为11.3%）。
- 通过**推理自动调优**可以获得**2%–20%**的吞吐量提升，不同引擎在吞吐量、端到端时延、第一token响应上表现出不同。

不同芯片上训练自动调优结果



FlagRelease: 开源大模型跨芯自动迁移和版本发布

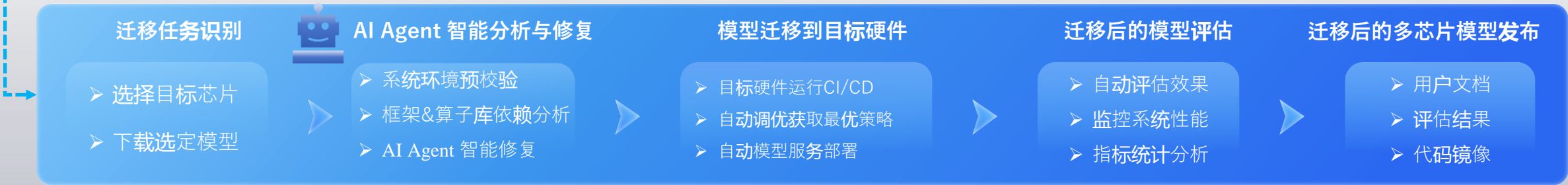
目标：依托FlagOS技术栈，构建了一套自动化迁移开源大模型至不同AI芯片的工具，并自动发布。通过FlagOS和智能体的相互支持，迁移和发布效率提升**4倍**，加速模型在多架构环境下的落地效率与生态成熟度。

源源不断自动产生多芯片模型，加快追赶生态差距

下载: 基于CUDA的原始模型

ModelScope & Hugging Face

发布: 迁移后的模型、代码及适配不同芯片的Docker 文件



<https://modelscope.cn/organization/FlagRelease>
<https://huggingface.co/FlagRelease/models?p=1>

硬件集群（搭载各种AI芯片）

一站式提供最新模型、多种芯片的统一版本，用户开箱即用，三步即可完成安装

开源模型	芯片
GLM4.5系列	Nvidia系列
Qwen2&3系列	Huawei
Kimi-K2、phi-4、grok-2、step3	ARM
MiniMax-M1-80k	Metax
MiniCPM-v4	Iluvatar
Seed-OSS-36B	Hygon
gpt-oss-120b	Cambricon
ERNIE-4.5-300B-A47B-PT	Kunlunxin
RoboBrain2.0-7B/32B	
MiniCPM-o-2_6-8B	
Deepseek-R1	

FlagRelease/Qwen3-32B-ascend-FlagOS Text Generation Safetensors PyTorch qwen3 FlagRelease 2025.09.23 ± 4	FlagRelease/Qwen3-32B-FlagOS Text Generation Safetensors PyTorch qwen3 FlagRelease 2025.09.18 ± 1	FlagRelease/GLM-4.5-FlagOS Text Generation Safetensors PyTorch FlagRelease 2025.08.14 ± 4 1	FlagRelease/MiniCPM-V-4-FlagOS Safetensors FlagRelease 2025.08.14 ± 6
FlagRelease/Qwen3-Next-80B-A3B-Instr... Text Generation Safetensors PyTorch FlagRelease 2025.09.18 ± 2	FlagRelease/Seed-OSS-36B-Instruct-Fl... Text Generation Safetensors PyTorch FlagRelease 2025.09.12 ± 4	FlagRelease/Qwen2.5-VL-32B-Instruct-Flag... Image-Text-to-Text Safetensors PyTorch FlagRelease 2025.08.13 ± 95	FlagRelease/MiniMax-M1-80k-FlagOS Text Generation Safetensors PyTorch FlagRelease 2025.08.11 ± 2.1k
FlagRelease/Qwen3-4B-FlagOS-Nvidia Text Generation Safetensors PyTorch qwen3 FlagRelease 2025.09.11 ± 6	FlagRelease/grok-2-FlagOS Text Generation Safetensors PyTorch git FlagRelease 2025.09.11 ± 75	FlagRelease/Kimi-K2-Instruct-FlagOS Text Generation Safetensors PyTorch FlagRelease 2025.08.07 ± 20	FlagRelease/Qwen3-235B-A22B-Instruct-250... Text Generation Safetensors PyTorch FlagRelease 2025.07.29 ± 34
FlagRelease/phi-4-hygon-FlagOS Text Generation Safetensors PyTorch phi3 FlagRelease 2025.09.01 ± 1	FlagRelease/gpt-oss-120b-FlagOS Text Generation Safetensors PyTorch FlagRelease 2025.08.29 ± 9	FlagRelease/ERNIE-4.5-300B-A47B-PT-FlagOS Text Generation Safetensors PyTorch FlagRelease 2025.07.23 ± 8	FlagRelease/DeepSeek-R1-FlagOS-Cambricon... Safetensors PyTorch FlagRelease 2025.07.23 ± 14
FlagRelease/step3-FlagOS Image-Text-to-Text Safetensors PyTorch FlagRelease 2025.08.28 ± 93	FlagRelease/phi-4-FlagOS Text Generation Safetensors PyTorch phi3 FlagRelease 2025.08.27 ± 1	FlagRelease/Qwen3-30B-A3B-Iluvatar-FlagOS Text Generation Safetensors PyTorch FlagRelease 2025.07.23 ± 16	FlagRelease/DeepSeek-R1-FlagOS-Kunlunxin... Text Generation Safetensors PyTorch FlagRelease 2025.07.23 ± 15

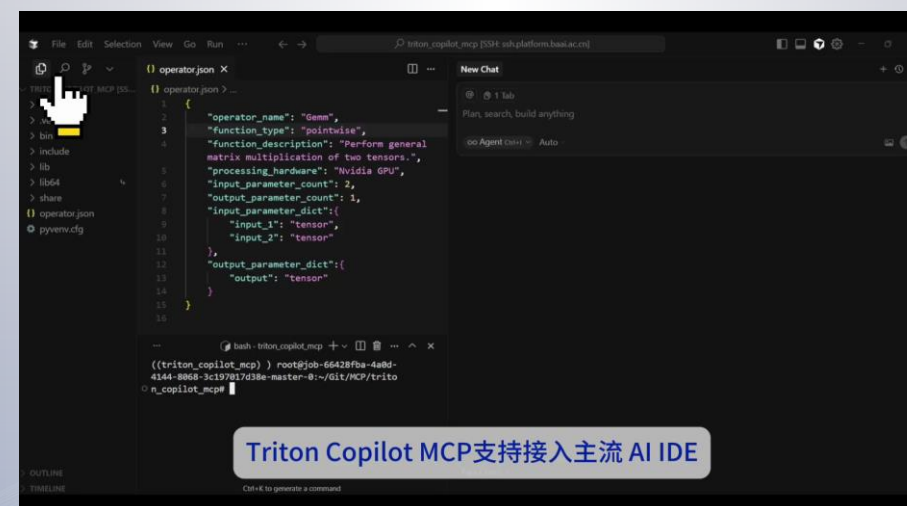
基于FlagOS的Triton-Copilot 算子自动生成工具

- **端到端算子生成及优化**：实现从算子需求解析、代码生成到测试验证的全流程闭环迭代。
- **正在支持多种硬件后端**：加快 Triton kernel 的实现与跨芯片适配
- **极低算子开发门槛**：开箱即用，0基础也能快速上手。相比起专家需要1~2天才能完成的单个算子开发，使用智源 Triton-Copilot，哪怕是无经验的开发者，也可以缩短到1~2小时。
- **性能不断迭代**：多数算子生成性能达到CUDA的50%~80%，个别算子超过CUDA算子性能
- **多种服务模式**：提供平台服务和MCP服务两种模式。
- **平台地址**：<https://triton-copilot.baai.ac.cn/>

Triton-Copilot的平台算子生成服务



基于Triton-Copilot的MCP服务



上线 FlagOS 知识库、面向开发者的一站式“自学”工具

众智 **FlagOS**
Open Software for AI Systems

FlagOS

囊括FlagOS核心模块的知识库

欢迎使用FlagOS-Wiki

为代码仓库提供AI驱动的文档

按项目名称、所有者或仓库搜索

FlagTree/flir

githubzh

FLIR 是从 microsoft/triton-shared 分支而...

处理时间: 9/25/2025

FlagTree/flagtree

githubzh

FlagTree 是一个面向多种 AI 芯片的统一开...

处理时间: 9/25/2025

FlagOpen/FlagScale

githubzh

FlagScale 是北京智源人工智能研究院 (BA...

处理时间: 9/25/2025

FlagOpen/FlagPerf

githubzh

FlagPerf 是一体化 AI 硬件基准测试平台, 支...

处理时间: 9/25/2025

FlagOpen/FlagGems

githubzh

FlagGems 是基于 OpenAI Triton 的高性能...

处理时间: 9/25/2025

FlagOpen/FlagCX

githubzh

FlagCX 是由北京智源人工智能研究院 (BAA...

处理时间: 9/25/2025

FlagOS

FlagTree 项目文档 Wiki

FlagTree 是一个面向多种 AI 芯片的统一开源编译器项目, 致力于打造多样化 AI 芯片编译器及相关工具平台, 推动 Triton 生态上下游发展。

FlagTree/flagtree

项目简介

核心特性

系统架构

后端系统

项目简介

Relevant source files

项目简介

FlagTree 是一个面向多种 AI 芯片的开源统一编译平台, 致力于多元化的 AI 硬件生态提供统一的编译、适配和优化能力。项目以 Triton 生态为基础, 兼容现有主流 AI 芯片后端, 统一代码仓库, 并快速实现单仓库多后端支持, 极大便利了上游模型开发者与下游芯片厂商的协作与创新。
Sources: [README_cn.md:1-26](#), [README.md:1-50](#)

FlagTree 当前正处于快速发展阶段, 持续集成更多后端, 拓展对 Triton 及其各版本的支持, 并提供硬件感知的优化能力。其目标是构建一个面向 AI 编译领域的共建平台, 实现从模型到多芯片的高效编译与验证流程。
Sources: [reports/v0.2.0/release_notes_v0.2.0_cn.md:1-28](#), [reports/v0.1.0/release_notes_v0.1.0_cn.md:1-26](#)

项目架构与核心能力

FlagTree 采用插件化、多后端架构, 围绕 Triton 编译链路, 兼容并扩展多种 AI 芯片的编译需求。平台支持两条主流编译路径 (TritonGPU Dialect 与 Linalg Dialect), 并通过高差异度模块插件机制, 允许芯片厂商按需扩展和维护专属后端。
Sources: [reports/v0.1.0/release_notes_v0.1.0_cn.md:7-20](#), [README_cn.md:8-18](#)

FlagTree和Triton-lang的主要差异?

- 多后端支持: FlagTree 在 Triton-lang 基础上, 集成了多种 AI 芯片后端 (如 iluvatar, xpu, mthreads, cambricon, aiipu, ascend, tsingmicro 等), 实现单仓库多后端统一编译, 而 Triton-lang 仅官方支持 nvidia/amd/cpu。
- 插件化与高差异度模块: FlagTree 支持后端厂商以插件形式集成高差异度定制模块, 便于芯片特性扩展和维护。
- 双编译路径: FlagTree 支持 TritonGPU dialect 和 Linalg dialect 两种编译路径, 兼容更多硬件和接入范式。
- 跨平台编译与验证: FlagTree 支持交叉编译和多平台快速验证, 提升开发与适配效率。
- CI/CD与质量管理: FlagTree 构建了多后端自动化 CI/CD、CLA 签署、安全合规等质量管理体系。
- Triton 版本扩展: FlagTree 支持多版本 Triton (3.0.x/3.1.x/3.2.x/3.3.x), 并为后端适配提供升级指南。
- 与量子库协同: FlagTree 联合 FlagGems 等量子库适配和扩展。

询问关于此仓库的问题...

FlagOS

详细介绍FlagOS每个开源项目

项目简介

Relevant source files

项目简介

FlagTree 是一个面向多种 AI 芯片的开源统一编译平台, 致力于多元化的 AI 硬件生态提供统一的编译、适配和优化能力。项目以 Triton 生态为基础, 兼容现有主流 AI 芯片后端, 统一代码仓库, 并快速实现单仓库多后端支持, 极大便利了上游模型开发者与下游芯片厂商的协作与创新。
Sources: [README_cn.md:1-26](#), [README.md:1-50](#)

FlagTree 当前正处于快速发展阶段, 持续集成更多后端, 拓展对 Triton 及其各版本的支持, 并提供硬件感知的优化能力。其目标是构建一个面向 AI 编译领域的共建平台, 实现从模型到多芯片的高效编译与验证流程。
Sources: [reports/v0.2.0/release_notes_v0.2.0_cn.md:1-28](#), [reports/v0.1.0/release_notes_v0.1.0_cn.md:1-26](#)

项目架构与核心能力

FlagTree 采用插件化、多后端架构, 围绕 Triton 编译链路, 兼容并扩展多种 AI 芯片的编译需求。平台支持两条主流编译路径 (TritonGPU Dialect 与 Linalg Dialect), 并通过高差异度模块插件机制, 允许芯片厂商按需扩展和维护专属后端。
Sources: [reports/v0.1.0/release_notes_v0.1.0_cn.md:7-20](#), [README_cn.md:8-18](#)

架构总览

下面用 Mermaid 图展示 FlagTree 的主要架构组件及其关系:

```
graph TD; A[模型开发者] --> B[FlagTree 编译器];
```

FlagOS

FlagTree 项目文档 Wiki

FlagTree 是一个面向多种 AI 芯片的统一开源编译器项目, 致力于打造多样化 AI 芯片编译器及相关工具平台, 推动 Triton 生态上下游发展。

FlagTree/flagtree

项目简介

核心特性

系统架构

后端系统

项目简介

Relevant source files

项目简介

FlagTree 是一个面向多种 AI 芯片的开源统一编译平台, 致力于多元化的 AI 硬件生态提供统一的编译、适配和优化能力。项目以 Triton 生态为基础, 兼容现有主流 AI 芯片后端, 统一代码仓库, 并快速实现单仓库多后端支持, 极大便利了上游模型开发者与下游芯片厂商的协作与创新。
Sources: [README_cn.md:1-26](#), [README.md:1-50](#)

FlagTree 当前正处于快速发展阶段, 持续集成更多后端, 拓展对 Triton 及其各版本的支持, 并提供硬件感知的优化能力。其目标是构建一个面向 AI 编译领域的共建平台, 实现从模型到多芯片的高效编译与验证流程。
Sources: [reports/v0.2.0/release_notes_v0.2.0_cn.md:1-28](#), [reports/v0.1.0/release_notes_v0.1.0_cn.md:1-26](#)

项目架构与核心能力

FlagTree 采用插件化、多后端架构, 围绕 Triton 编译链路, 兼容并扩展多种 AI 芯片的编译需求。平台支持两条主流编译路径 (TritonGPU Dialect 与 Linalg Dialect), 并通过高差异度模块插件机制, 允许芯片厂商按需扩展和维护专属后端。
Sources: [reports/v0.1.0/release_notes_v0.1.0_cn.md:7-20](#), [README_cn.md:8-18](#)

如何用FlagScale训练一个多模态7B模型

- 数据准备
 - 准备多模态数据集, 包括文本和图片 (或其他模态), 确保格式符合 FlagScale 的数据加载要求。
 - 按照 FlagScale 支持的数据格式 (如 JSONL, Parquet 等) 组织数据, 常见字段包括 input_ids, attention_mask, image 等。
- 模型配置
 - 选择或自定义 7B 参数数量的多模态模型结构 (如 LLaVA, Qwen-VL 等在 FlagScale 生态中有支持)。
 - 配置模型结构文件 (如 config.json), 包括模型参数、模型融合方式、视觉编码器等信息。
- 环境与依赖
 - 安装 FlagScale 及其依赖: `pip install flagscale`

询问关于此仓库的问题...

进行知识讲解：如FlagTree与Trion-lang的差异

动手教学：如何用FlagScale训练一个多模态7B模型

1. 通过全面AI赋能FlagOS，实现迁移的全面自动化、算子的100%覆盖

真正做到赋能AI，AI赋能

算子生成与调优

kernel generation &
tuning

架构抽象与编译

architecture abstraction
& compilation

适配与迁移

Adaptation & migration

工程集成与CICD

Engineering & CICD

Phase 1: AI 辅助开发
(AI-augmented development)

Phase 2: 自动化子系统
(Sub-sys automation)

Phase 3: 端到端自主适配
(End-to-end automatic adaptation)

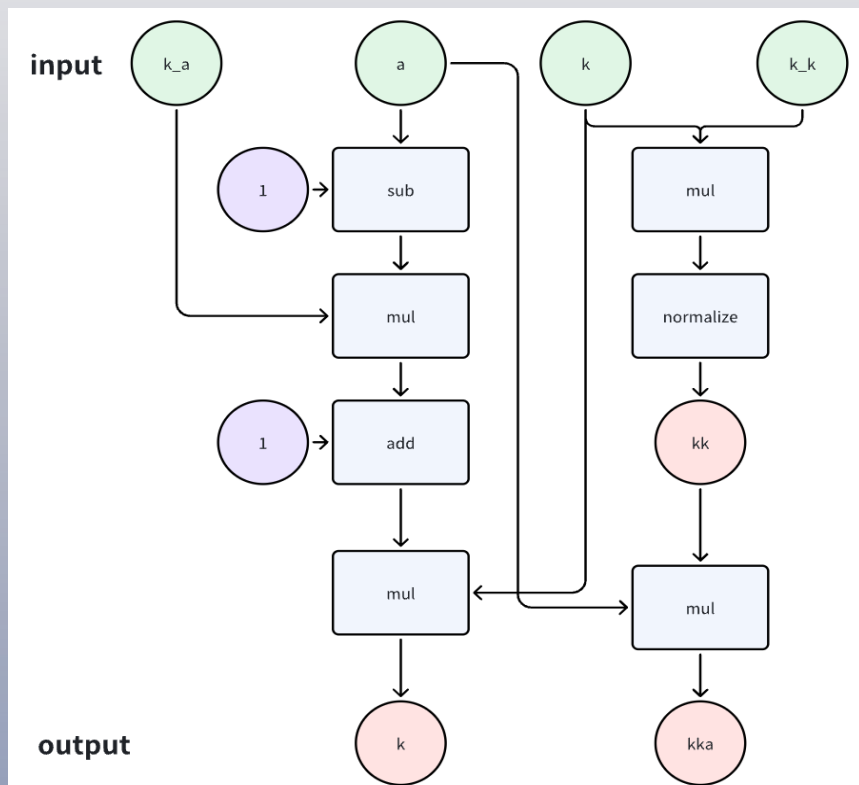
鼓励多个厂商共同参与和推动

- 对开源软件中要实现多芯片后端的支持方法进行标准化，避免不同AI开源项目、不同芯片厂商的实现方法过于分裂。

FlagOS 支持模型架构创新、效率提升

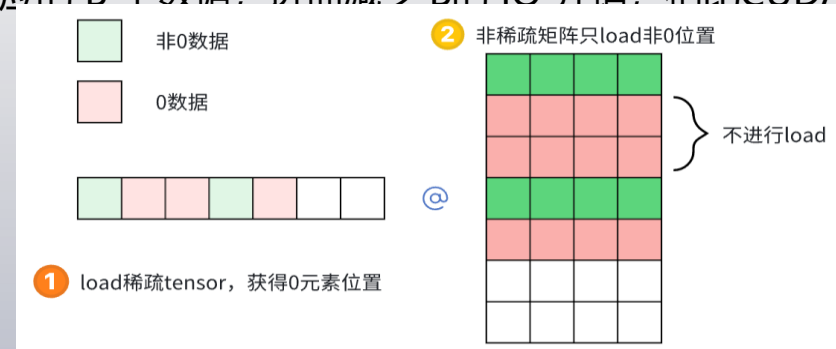
① KA-fusion算子：基于FlagGEMs的优化

将normalize、mul等多步操作进行融合，减少内存访问与调度开销，相比CUDA加速比 **1.8x**



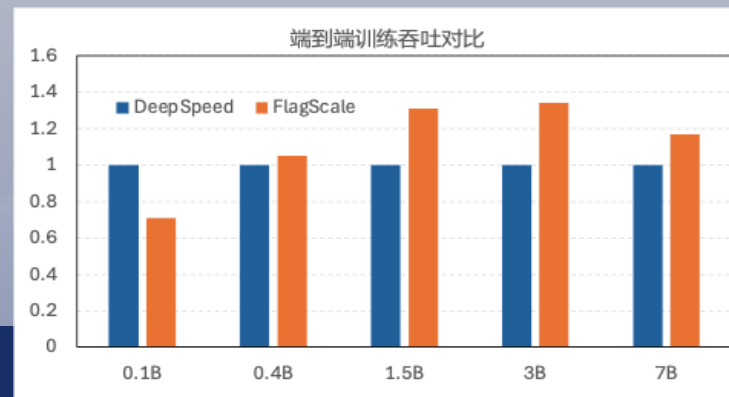
② MM-Sparsity算子：基于FlagGEMs的优化

针对MM操作 A (一维向量) @ B (二维矩阵)。当 A 含90%以上0元素，不load A 中0对应的 B 中数据，进而减少 B 的IO开销，相比CUDA加速比 **1.53x**

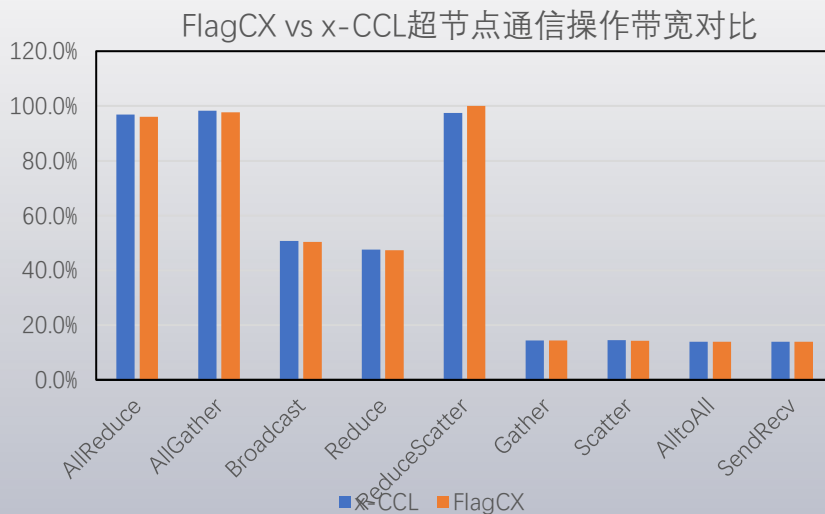


③ 基于FlagScale实现RWKV端到端高效训练

基于FlagScale，相比官方DeepSpeed版本，端到端训练最高加速**34%**，平均加速11.6%。

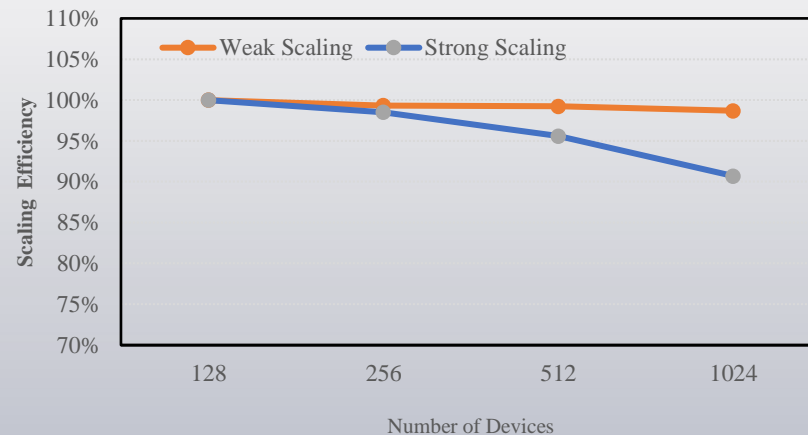


浪潮信息元脑 SD200 超节点支持



- 浪潮信息元脑SD200已适配FlagCX，平均带宽达到原生通信库99.8%。
- 浪潮信息元脑SD200已适配FlagScale，支持预训练、微调、推理等功能，64卡扩展效率近线性。

海光 Nebula 超节点支持



- 海光超节点适配FlagScale训练自动调优，十分钟内搜索出较优配置。
- 在32B+32K序列长度下，千卡弱扩展效率超过98%，强扩展效率超过90%。

FlagOS已经支持从边缘到数据中心的多种系统硬件形态：

端侧计算模组



AI PC



各种AI服务器

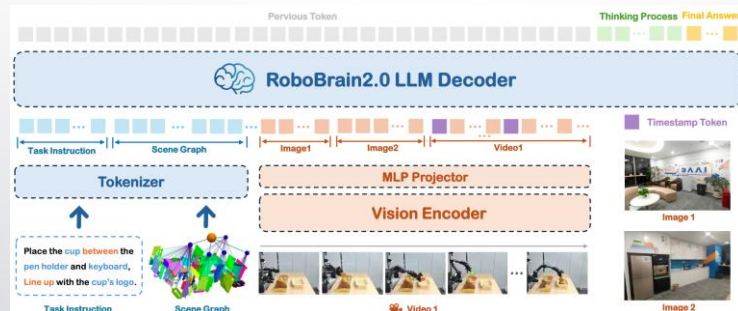


超节点系统



FlagOS支持具身智能全链路训练推理：实现跨芯片、提效

VLM大脑模型



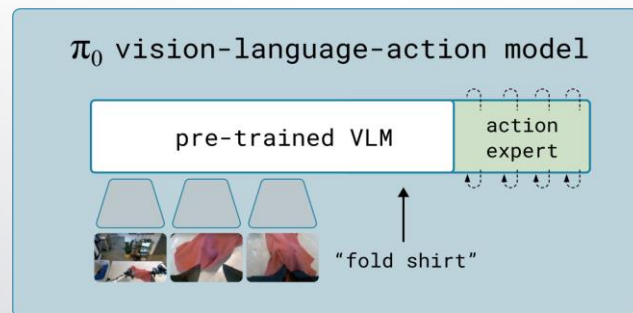
分布式训练：

- 通过非均匀流水线并行实现高效并行
- 基于Energon 实现高效分布式数据加载
- 利用显存分配预处理消除碎片
- 精细化重计算显著降低显存占用
- 端到端性能相比Llama-factory **提升154.81%**

高效推理：

- 多后端支持
- 部署参数自动调优
- 模型量化（WA）加速
- 端到端推理性能 **提升22%**

VLA端到端模型



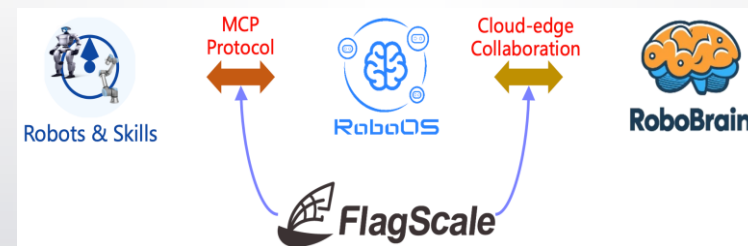
分布式训练：

- 预训练/后训练数据加载统一
- 后训练通过冻结VLM支持
- 基于Energon 实现高效分布式数据加载

高效推理：

- 多后端支持
- 部署参数自动调优
- 模型量化（W8A16）加速
- 端到端推理性能 **提升23%**
- 端侧部署支持
- 真机系统代理

高效端云协同



机器人技能注册标准化：

- 典型场景输入 token **减少 65%**
- 自动技能注册和技能商店构建

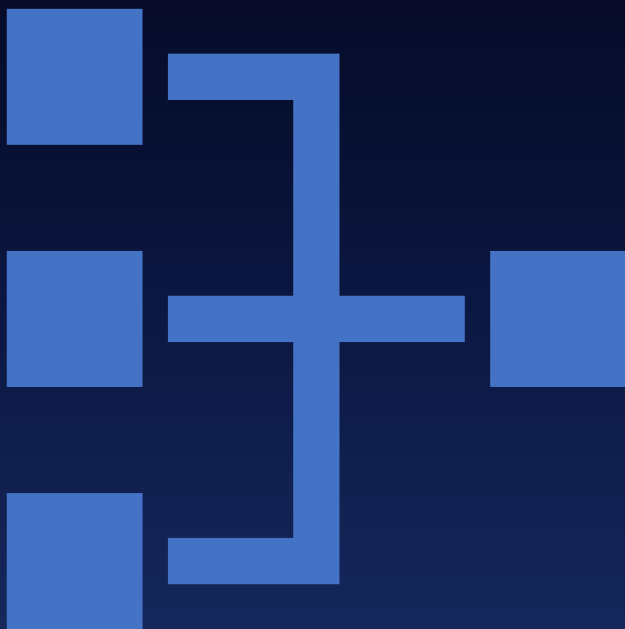
端云协同的快速通信方案：

- 平均延迟降低至 **<3ms**
- 支持全流程历史数据随机访问

技能检索功能：

- 总 Token 用量 **降低 29.8%**

FlagOS支持“2025中关村具身智能大赛”，为选手提供跨平台的训练、推理能力。



打造FlagOS全球化社区

聚拢生态合作企业和机构

- 芯片企业：寒武纪、华为等17家
- 服务器企业：浪潮、新华三等5家
- AI软件企业：硅基流动、中科加禾等7家
- 高校和科研机构：清华、北大、计算所、先进编译实验室等5家
- 集成和应用企业：中国移动、中国联通、天翼云等16家
- 大模型企业：百度、科大讯飞、面壁智能等3家
- 操作系统：麒麟软件、OpenCloudOS等3家
- 行业和开源组织：Circle Linux、CSDN等4家

生态合作平台

北京人工智能公共算力平台



Hugging Face



Modelers



ModelScope

CSDN

全球合作，拓展国际技术生态

- 通用算子库成为Pytorch基金会正式生态合作项目
- 统一通信库正式递交ITU国际标准

PyTorch



高校人才培养

- 基于FlagOS，中国科技大学、中科院计算所、北航、北邮等高校开展课程培训，累计覆盖487+学生
- 启动“自主软硬件生态技术 高校人才培养计划”，邀请北京大学、清华大学、中科院计算所、中科院软件所、北京航空航天大学、北京邮电大学、湖南大学、郑州大学等首批重点大学共同启动。

FlagOS 全球开发者吸引

- 当前全球触达开发者近2万人，开源贡献者404人。
- 启动“FlagOS开放计算全球大赛”，从算子、到大模型和具身智能四大赛道

这是一个全球都关注的问题：如何推动开放计算，让AI普惠？

通过“开放计算”，来自不同区域、不同AI硬件的计算资源都可以轻松被发展中地区使用，降低算力成本和门槛。

2025年5月Open-source strategy Forum@巴黎



2025年7月全球数字合作峰会@日内瓦

Global Speakers and Collaborators

Co-organized by: ITU, iSpirt, Generative AI Commons

Featuring speakers from:
ITU, iSpirt, BAAI, OFE, Rise, IEEE, World Economic Forum, Open Wallet Foundation, The First Person Project, Digital Governance Institute, ForceCo.io, Redhat, Generative AI Commons, LF AI & Data



2025年9月与联合国共同举办的Open for SDG大会@杭州



FlagOS高校计划：全景赋能体系

众智FlagOS
Open Software for AI Systems



FlagOS开放计算全球大赛

主办单位：FlagOS社区、智源研究院

合作单位：CCF开源发展技术委员会、CSDN、魔乐社区、魔搭社区、中关村科学城等

合作平台：国际赛道 - Kaggle平台，国内赛道 - 天池平台

总奖金池价值50万人民币（分设一等奖、二等奖、三等奖）

比赛时间

组队报名
2025年11月

开发作品
2025年12月中旬

提交作品
2026年2月

作品评审&颁奖
2026年3月

谢谢！

社区秘书长：朱其罡
电话/微信：13701234598