

SimAI – High-Precision Simulator for AI Large-Scale Training

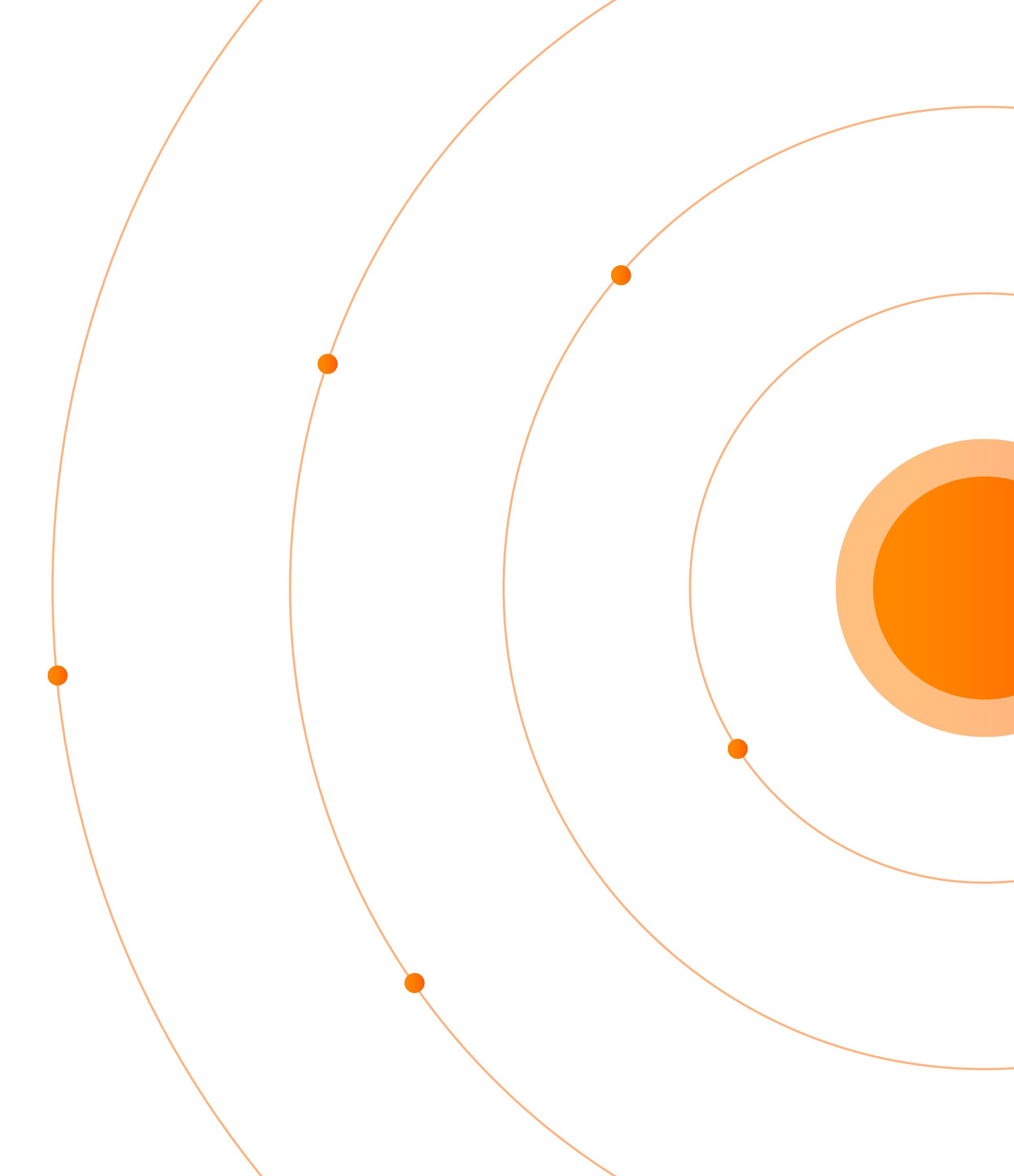
李庆旭

Alibaba Cloud

2024/11/26

Content

- Background
- SimAI Architecture
- User Manual
- **Future**



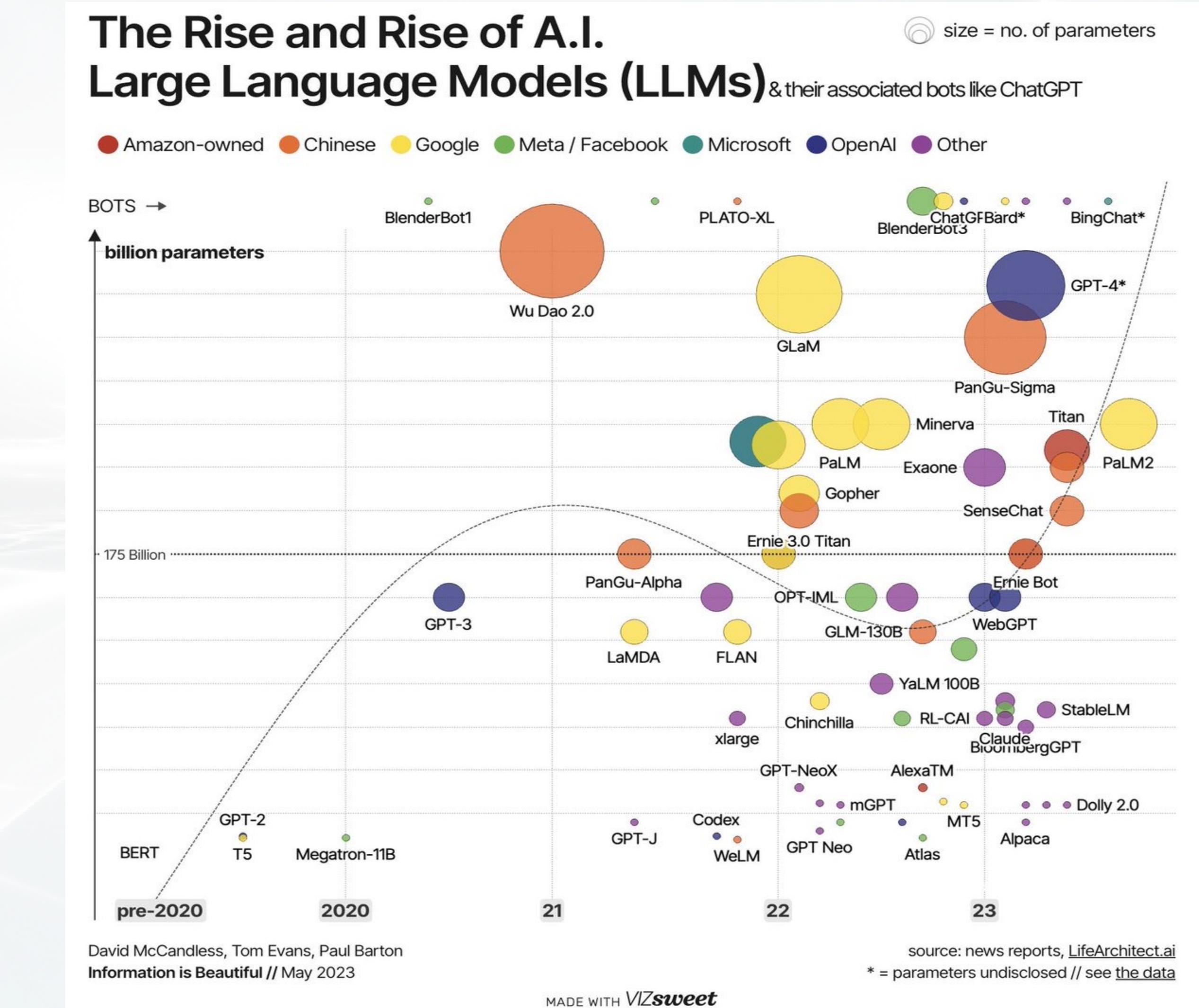
Background

需求

- LLM 领域发展迅速
- 框架迭代、并行训练算法迭代速度快
- 网络协议、拥塞控制、路由算法层出不穷
- 底层算力规格&网络架构多种多样

核心痛点：GPU 资源

- 绝大多数资源优先供应给业务
- 未知/新型 GPU , 只有样卡或者规格



Motivation

集群架构选型与论证

- GPU选型
- 机内Scale up网络架构
- 机间Scale out网络架构

.....

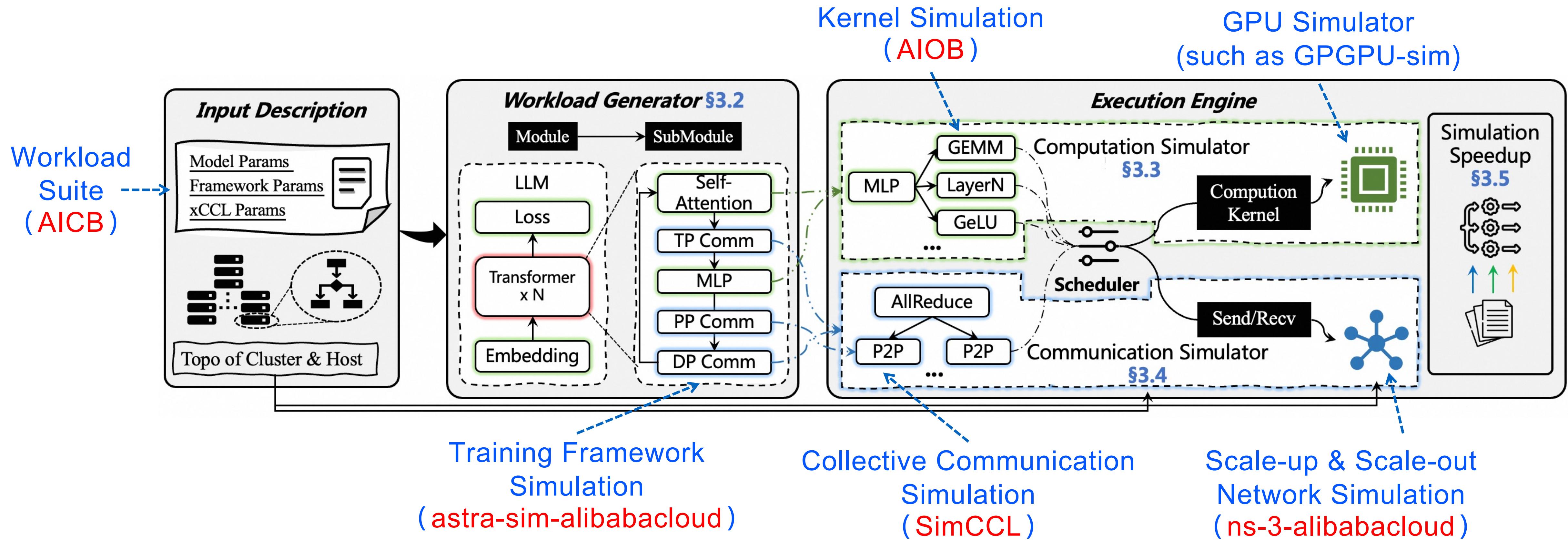
AI infra技术研发测试

- 框架&模型参数
- 通信库
- 拥塞控制
- 路由算法

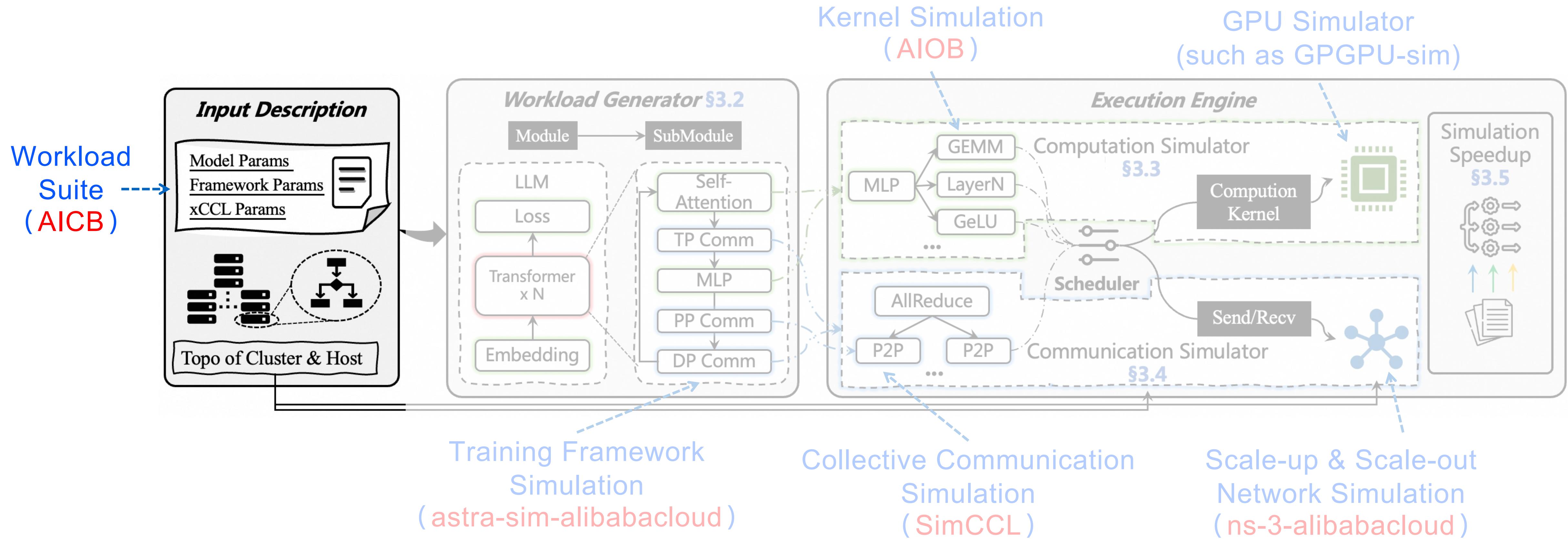
.....

需要模拟一个大规模训练集群

SimAI - Architecture



SimAI - Architecture



SimAI – AI Communication Benchmark

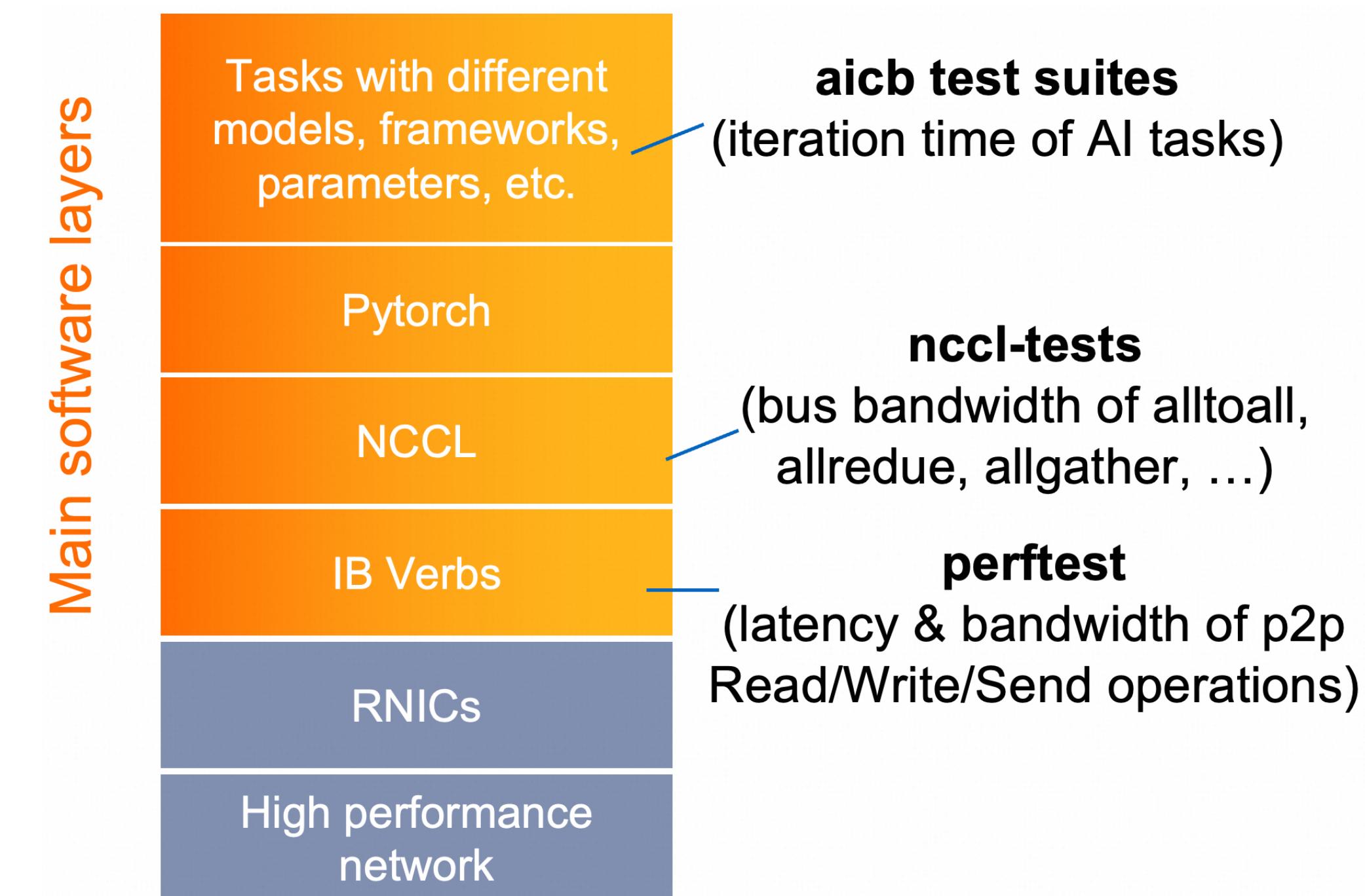
面向AI集群的高性能网络基准测试程序

1. 支持各规模、各模型、参数下通信负载的快速生成和分析 (workload generation)
2. 包含代表性的AI任务 (workload subsetting)
3. 模拟完整的软件栈以真实还原AI任务的通信负载 (workload replaying)

Benchmark case选取：

- 线上任务 log
- 各论文中的典型数据
- 能覆盖各种集合通信pattern

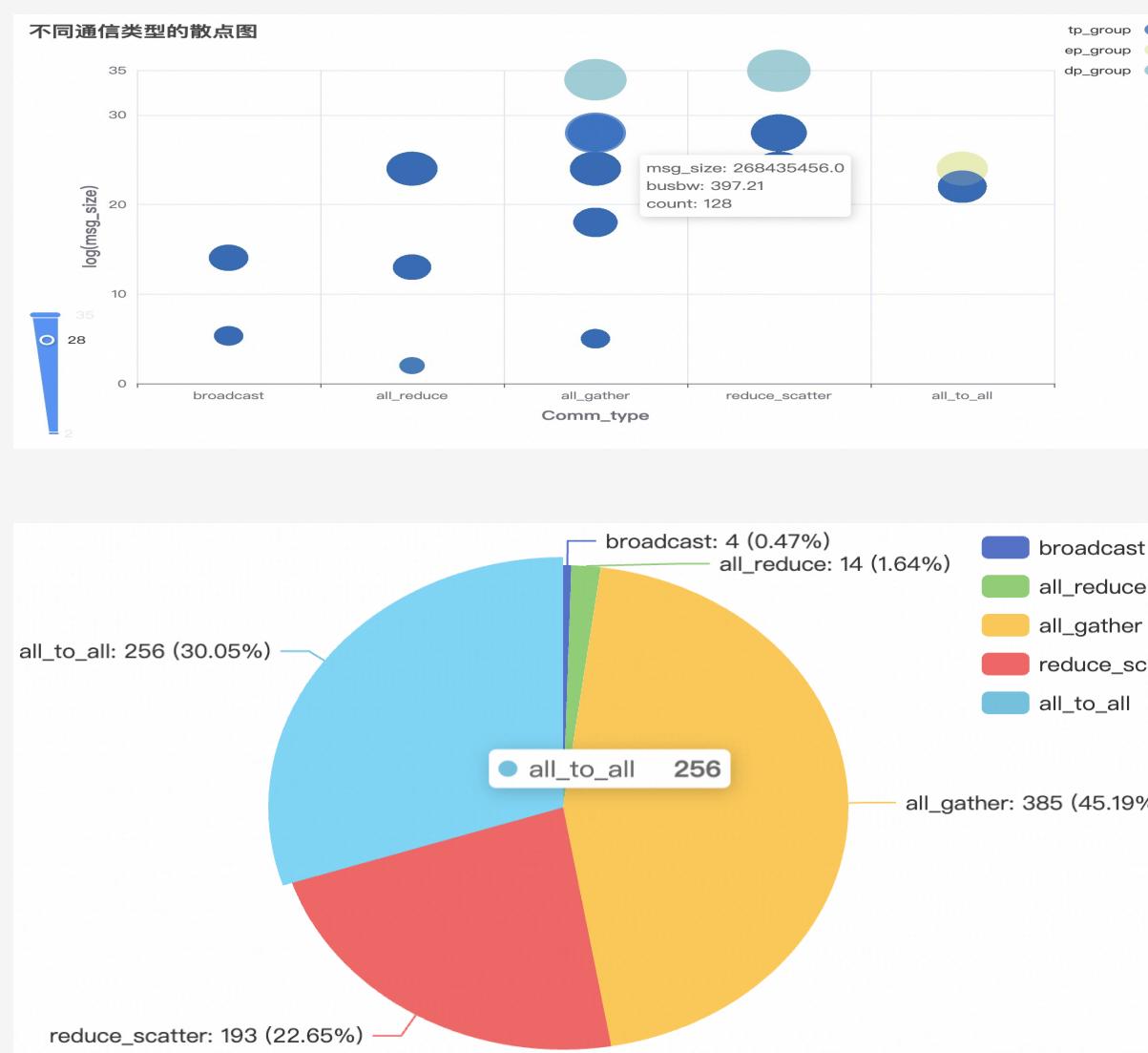
模型参数&并行参数



SimAI – AI Communication Benchmark

负载生成与分析

可根据规模、框架、模型等各类参数生成响应的负载，并给出通信模式的分析



AI集群的通信基准评测

运行AICB测试套件进行AI集群的通信性能的基准评测，测试套件包含了代表性的AI任务的工作负载

基准测试

高通量
以太网

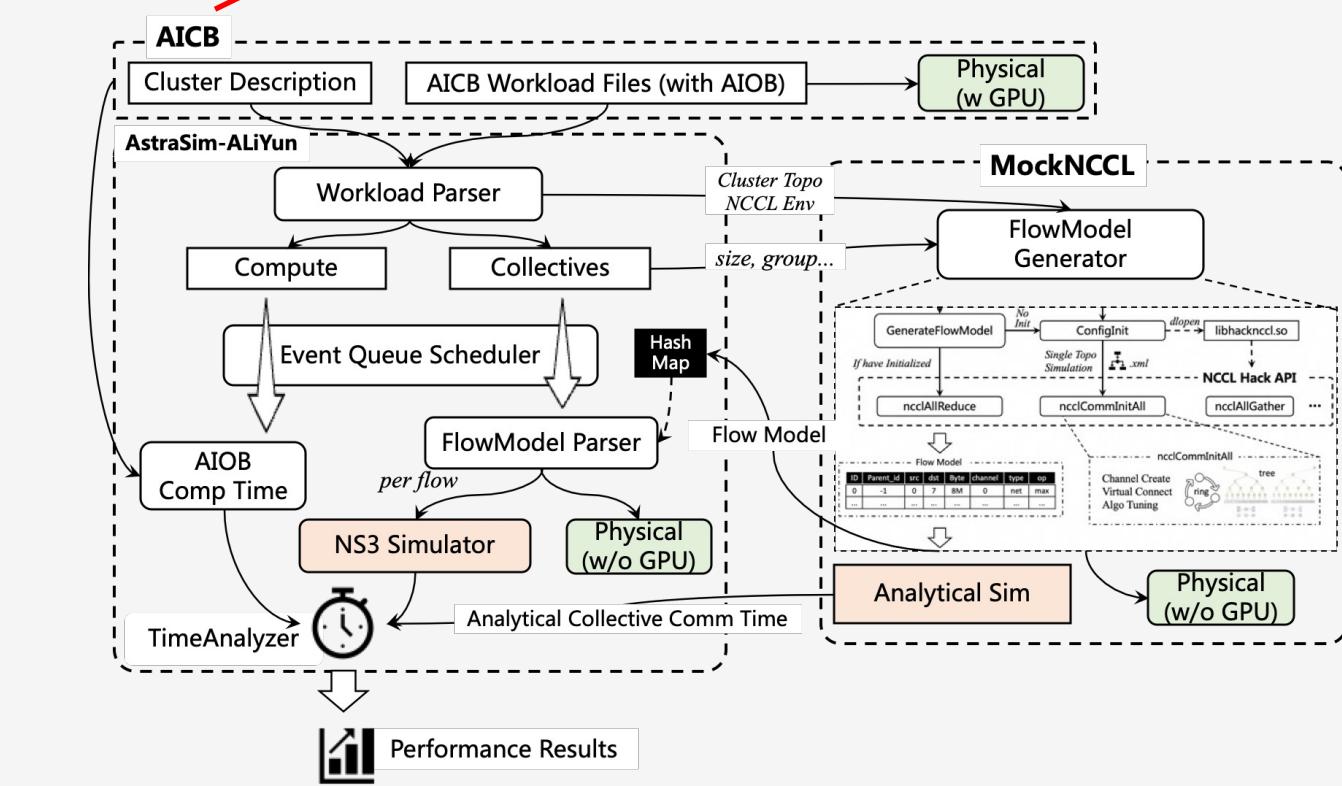
RoCE
网络

id	Name	Parameter_size	Hidden_size	Num_of_layers	Attention_heads	Sequence_length	FFN_hidden_size
1	GPT_7B	7B	4096	32	32	2048	16384
2	GPT_13B	13B	5120	40	32	2048	20480
3	GPT_22B	22B	6144	48	64	2048	24576
4	GPT_175B	175B	12288	96	96	2048	49152
5	GPT_13B	13B	5120	40	32	2048	20480
6	LLaMA_7B	7B	4096	32	32	4096	11008
7	LLaMA_7B	7B	4096	32	32	4096	11008
8	LLaMA_65B	65B	8192	80	64	4096	28672
9	LLaMA_65B	65B	8192	80	64	4096	28672
10	Mistral_8*7B	56B	4096	32	32	1024	14336

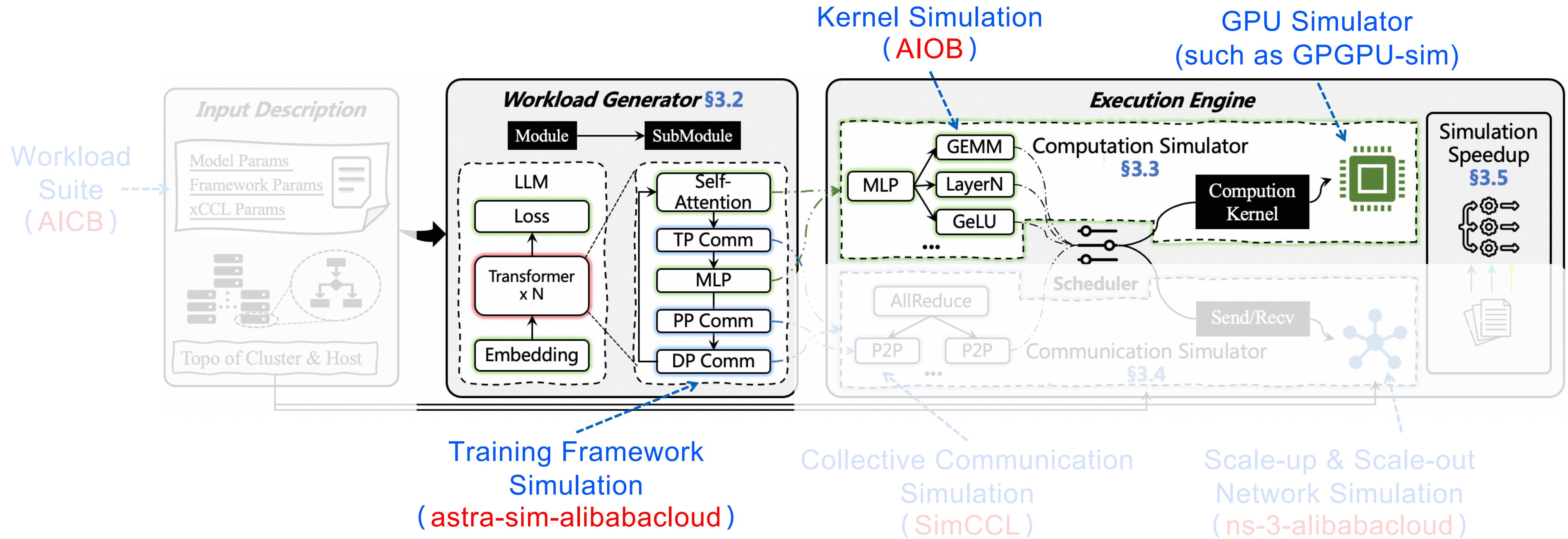
通信负载结构化输出

提供结构化的通信负载描述文件，可供模拟器（如大型模型模拟器SimAI）、集合通信负载重放工具（如SimCCL）等使用

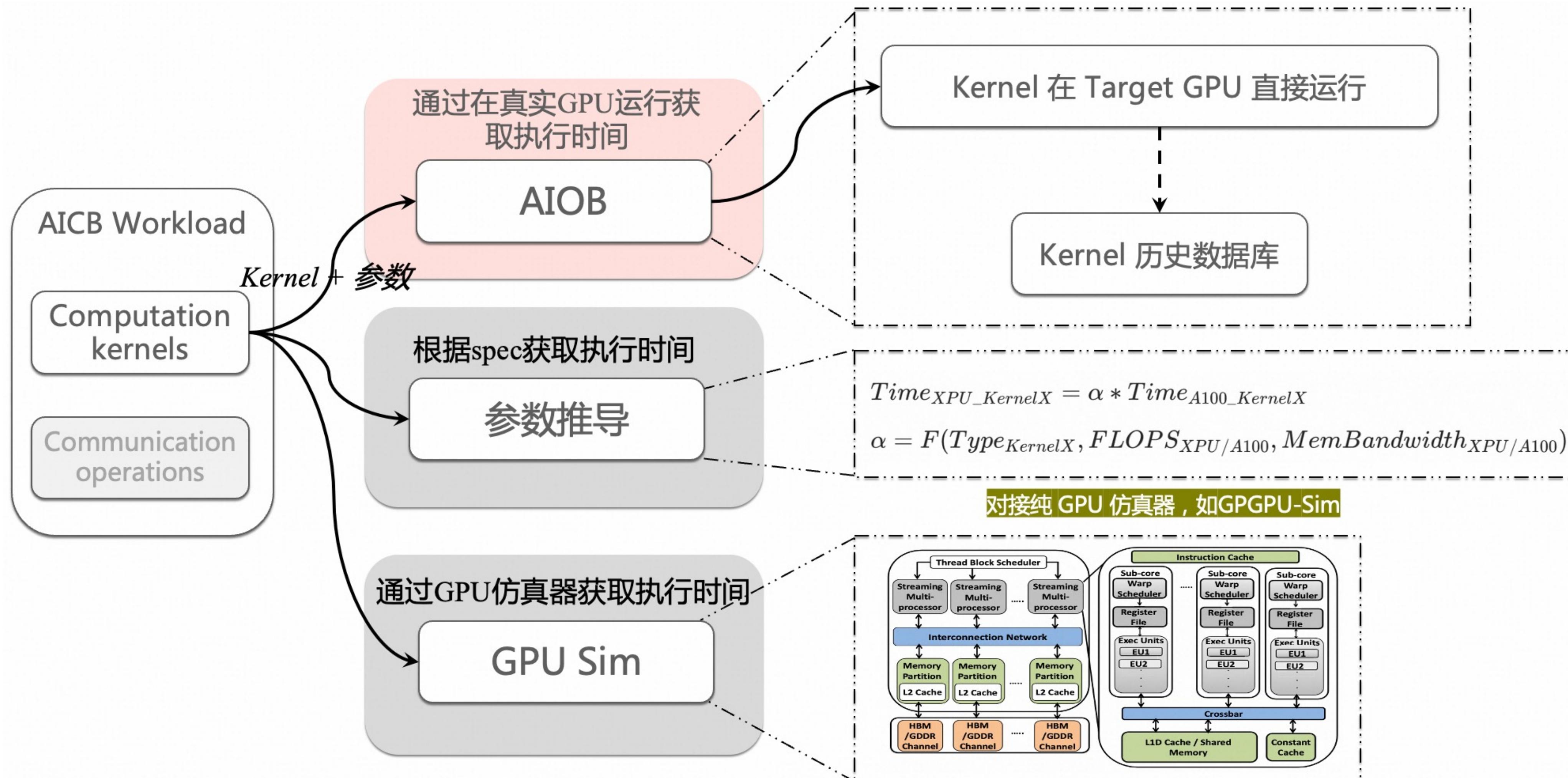
作为模拟器的负载输入



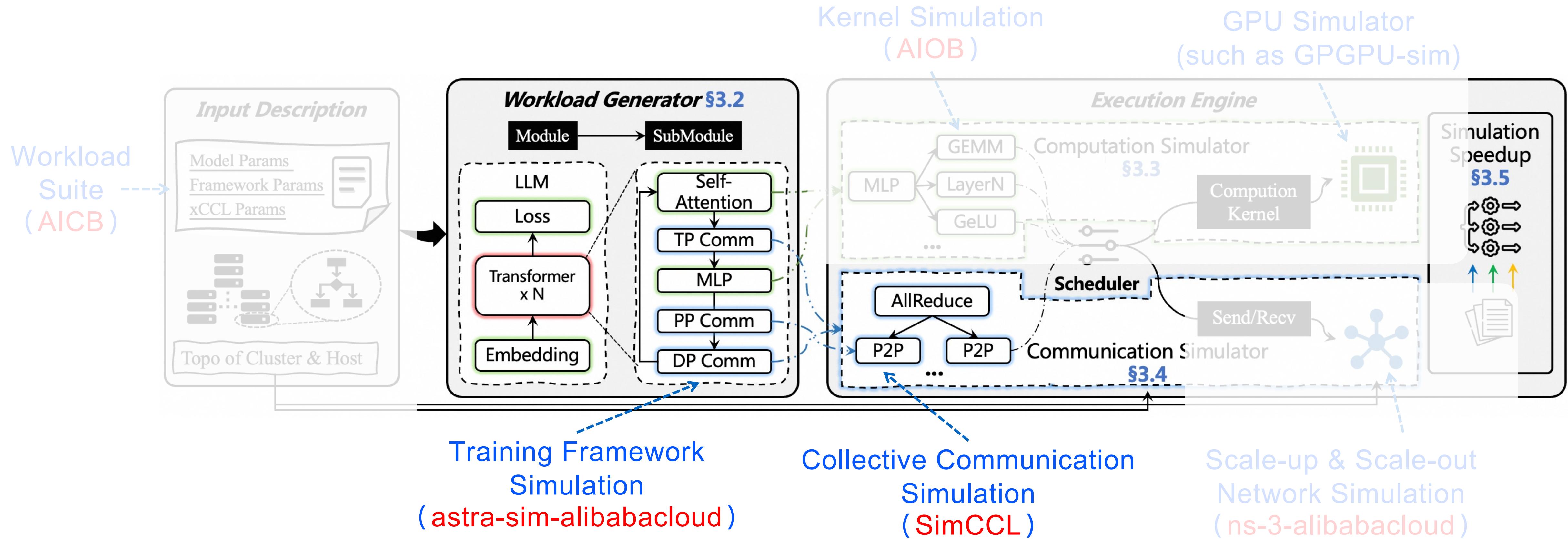
SimAI - Architecture



SimAI – Computation Simulator



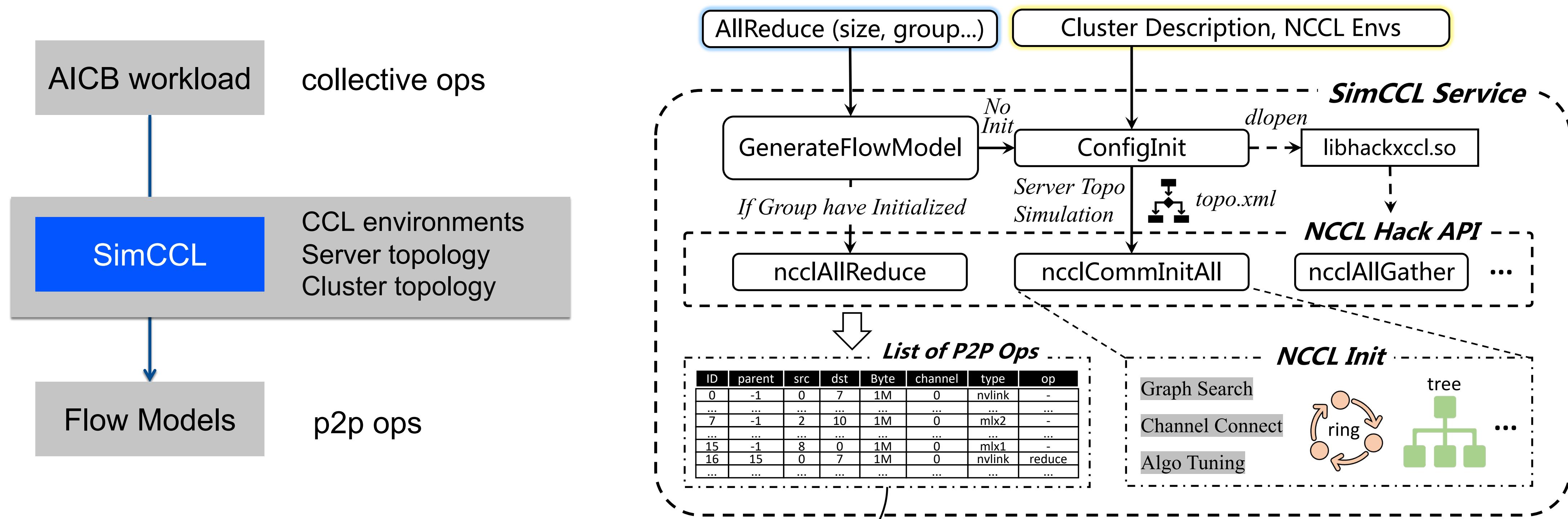
SimAI - Architecture



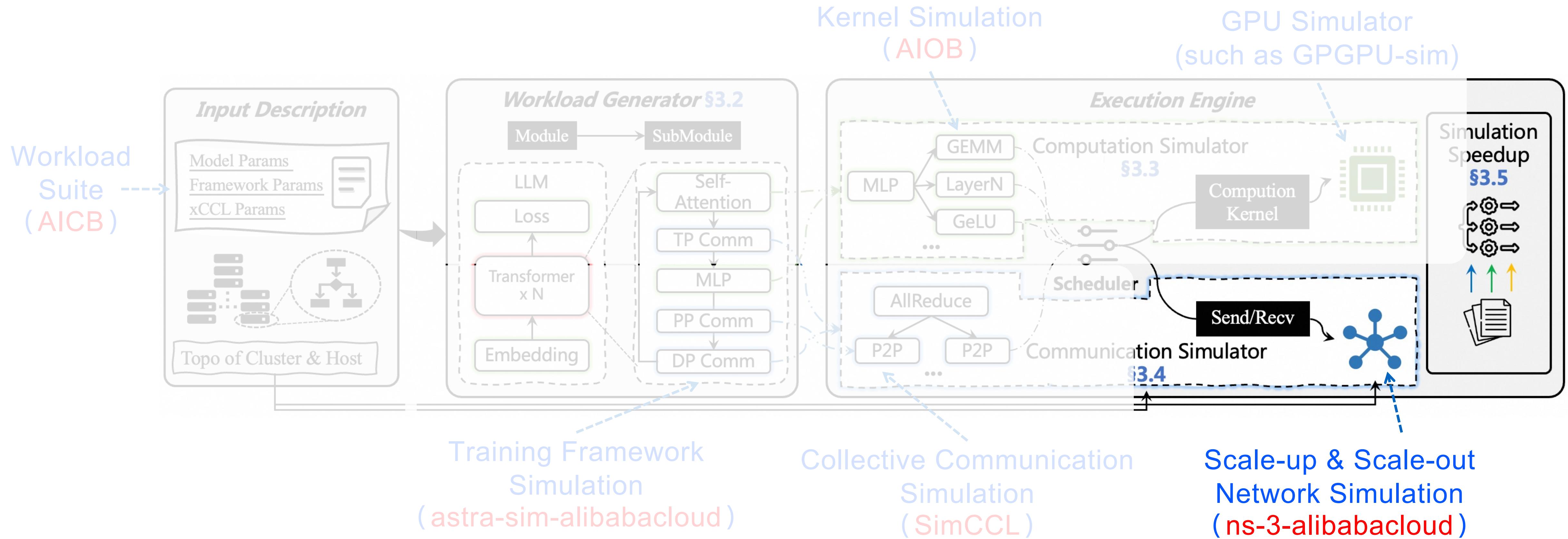
SimAI - Communication Simulator

SimCCL service实时转换AICB workload中的集合通信

1. 保留CCL的通信pattern的真实性（支持原生的集合通信算法，Env，protocol优化）
 2. 不依赖GPU服务器或服务器规模

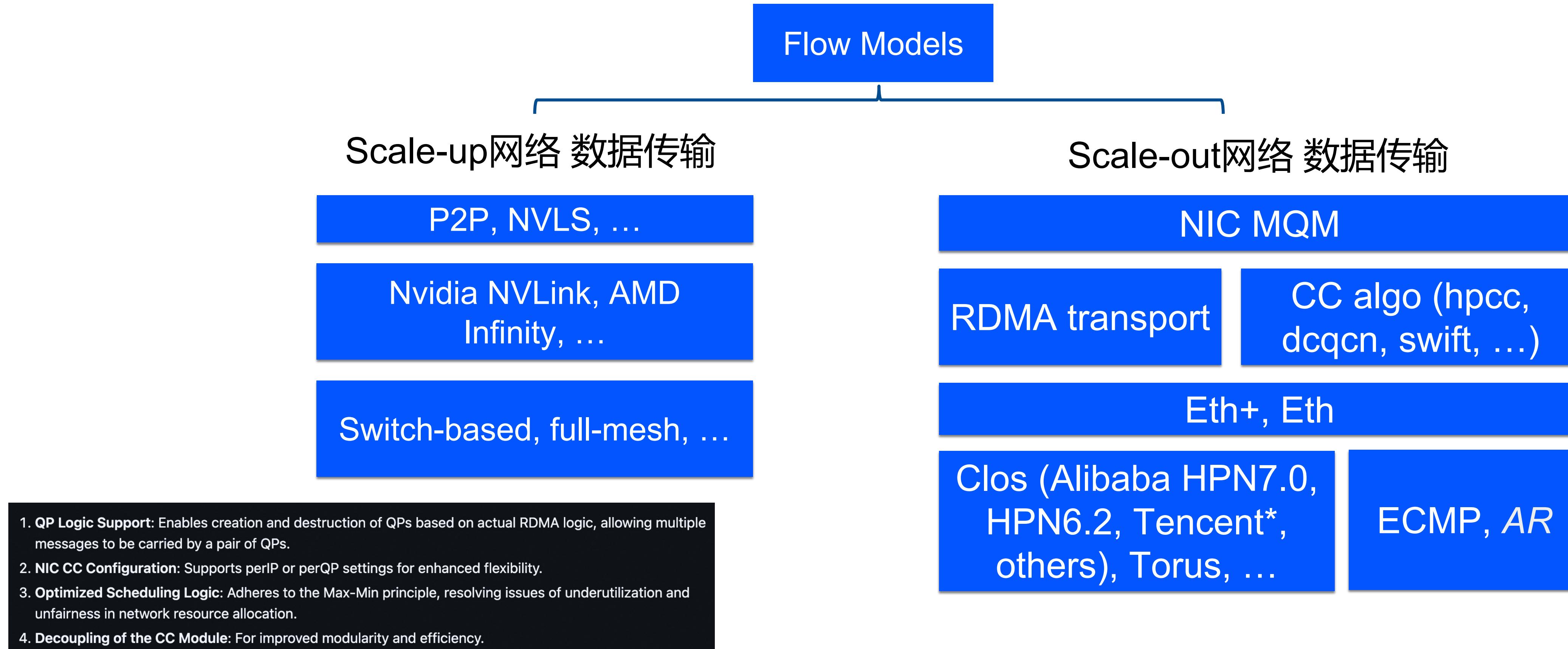


SimAI - Architecture



SimAI – Communication Simulator

基于NS-3的Scale-up&Scale-out网络模拟器

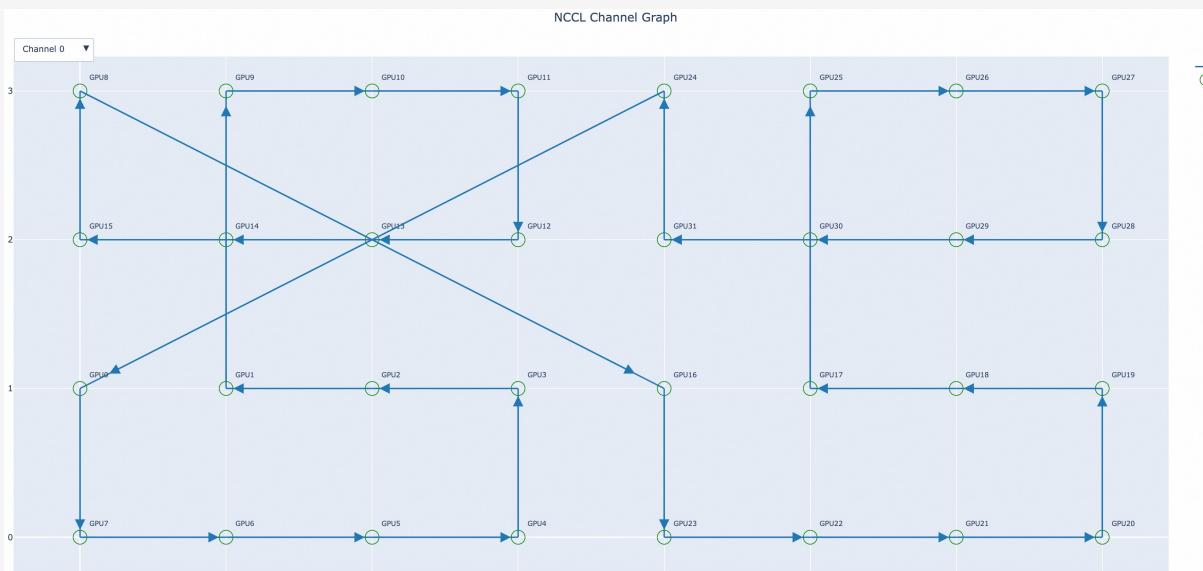


SimAI – 使用场景

SimCCL (physical)

- 生成集合通信的 FlowModel、Graph信息
- 非 GPU 集群 verbs 打流

ID	parent	src	dst	Byte	channel	type	op
0	-1	0	7	1M	0	nvlink	-
...
7	-1	2	10	1M	0	mlx2	-
...
15	-1	8	0	1M	0	mlx1	-
16	15	0	7	1M	0	nvlink	reduce
...

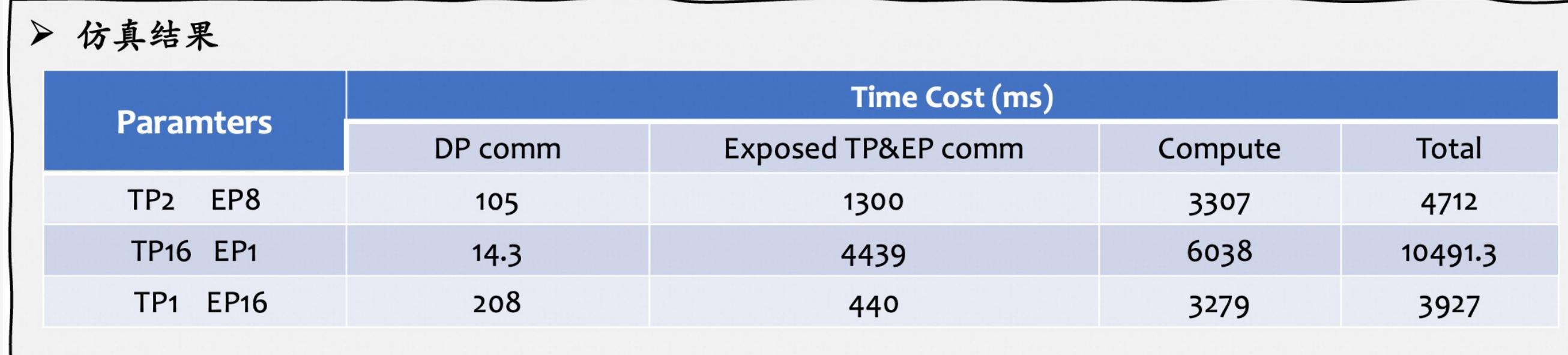


Analytical

- 屏蔽底层通信细节
- 快速仿真验证
- Parameter Tuning

All Simulation

- 集成 ns3 的网络全仿真
- 更多关注网络细节



SimAI – 全栈工具集

SimAI核心组件

AICB
SimCCL
Astra-sim-AlibabaCloud
NS-3-Alibaba Cloud

SimAI套件支持的主要场景

使用场景	组件构成
为模拟器提供test suite 或自定义的workload	AICB workload description files
运行基础test suite 或 自定义workload ; 分析典型负载集合通信特征	AICB test suites for physical GPU clusters
分析集合通信的RDMA流量特征	Flow model generated by SimCCL
在非GPU的集群进行网络测试分析	AICB + SimCCL + Astra-sim-Alibabacloud (Verbs)
在任意Server进行AICB workload的重放时间的分析演算	AICB + Astra-sim-Alibabacloud (Analytical)
在任意Server进行AICB workload的重放时间的全流程仿真	AICB + SimCCL + Astra-sim-Alibabacloud (Simulation) + NS-3-Alibabacloud*

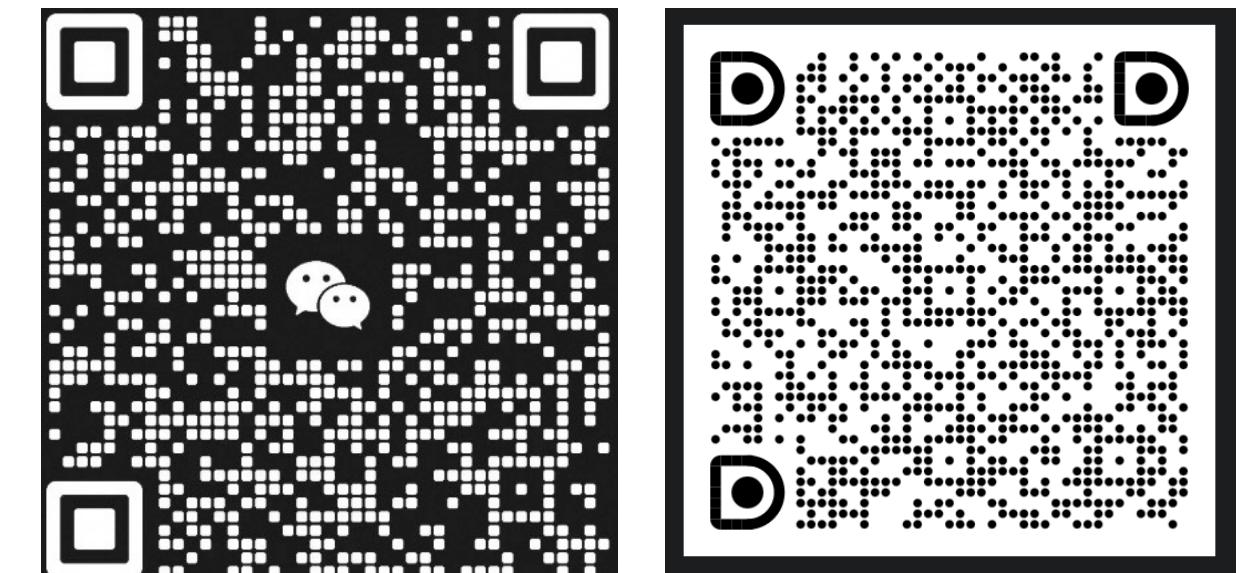
<https://github.com/aliyun/aicb>
<https://github.com/aliyun/SimCCL>

<https://github.com/aliyun/SimAI>
<https://github.com/aliyun/ns-3-alibabacloud>

SimAI – 未来展望



诚邀各位技术同仁参与社区合作开发！



Thank you
欢迎私戳