

OUYANG YANJIA

HPC-AI Technology – Business Manager



HPC-AI.COM

Democratizing Efficient Large
AI Models for Everyone





Scan me

CONTENTS

- Company Overview
- Product Demo
- Technical Highlights
- Applications in Practice





Scan me

Company Overview





Scan me

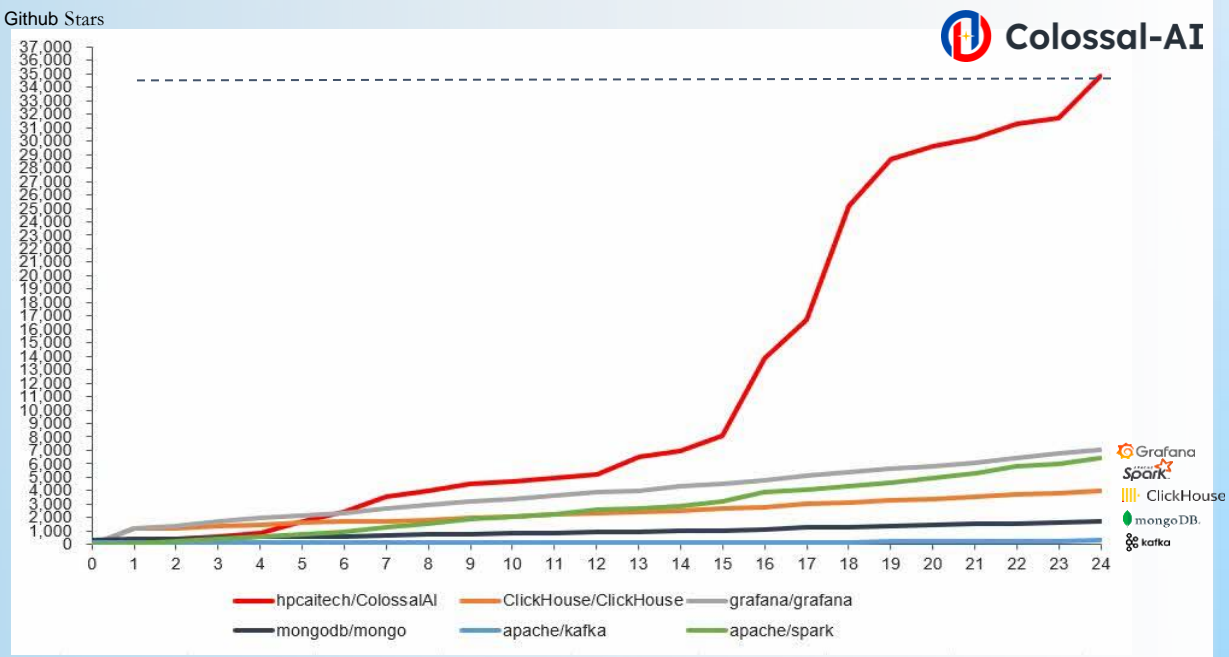
HPC-AI Team: The Most Powerful Team in Computing Power Optimization

**First
in the World**



Helped Google speed up the world's best large model from 3 days to 76 minutes on the first thousand-card AI-specific cluster in 2018

Top One



Colossal-AI leads globally in the AI system track within the open-source community.

Case Studies



*Utilizing Colossal-AI's Technologies in
Large-Scale Training*



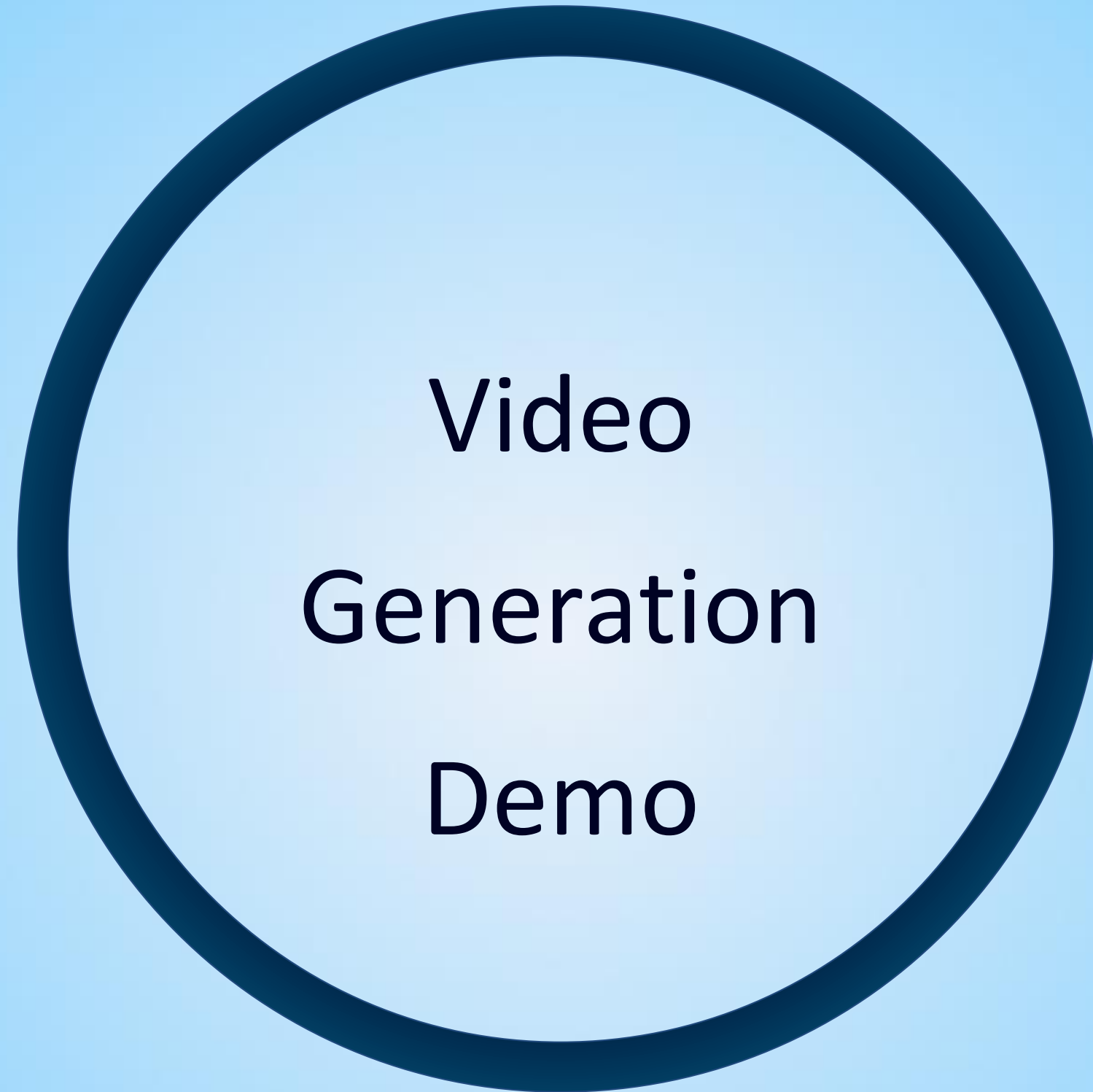
*Accelerating Large-Scale Training with
Colossal-AI's Technology on a Ten-
Thousand- Card Cluster*



*NVIDIA's Products Rely on Colossal-AI's
Sequential Parallel Technology*



Scan me



Video Generation Demo



Scan me



All videos on this page were generated directly by Video Ocean without modification.



Scan me

Technical Highlights





Scan me

Low Cost Model Training Scheme



Fig.1 Sora-like model training scheme

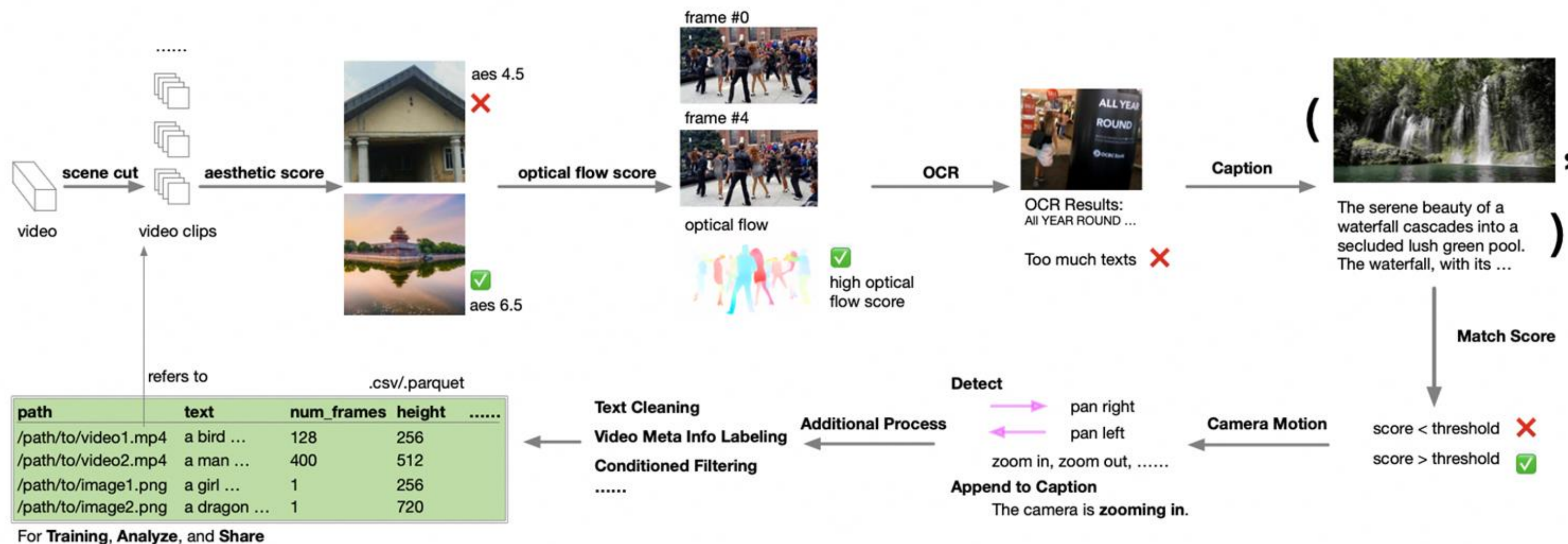
Our scheme: controlling the cost within 10K USD

- Phase 1: Large-scale image pre-training;
 - We just use PixArt- α by Huawei
- Phase 2: Large-scale video pre-training;
- Phase 3: High-quality video data fine-tuning.



Data Collection Pipeline

Open-Sora Data Processing Pipeline v1.1



Scan me

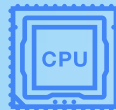
Fig.2 Open-Sora Data Processing Pipeline

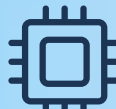
Colossal-AI: Fast and Cost-Effective





Scan me

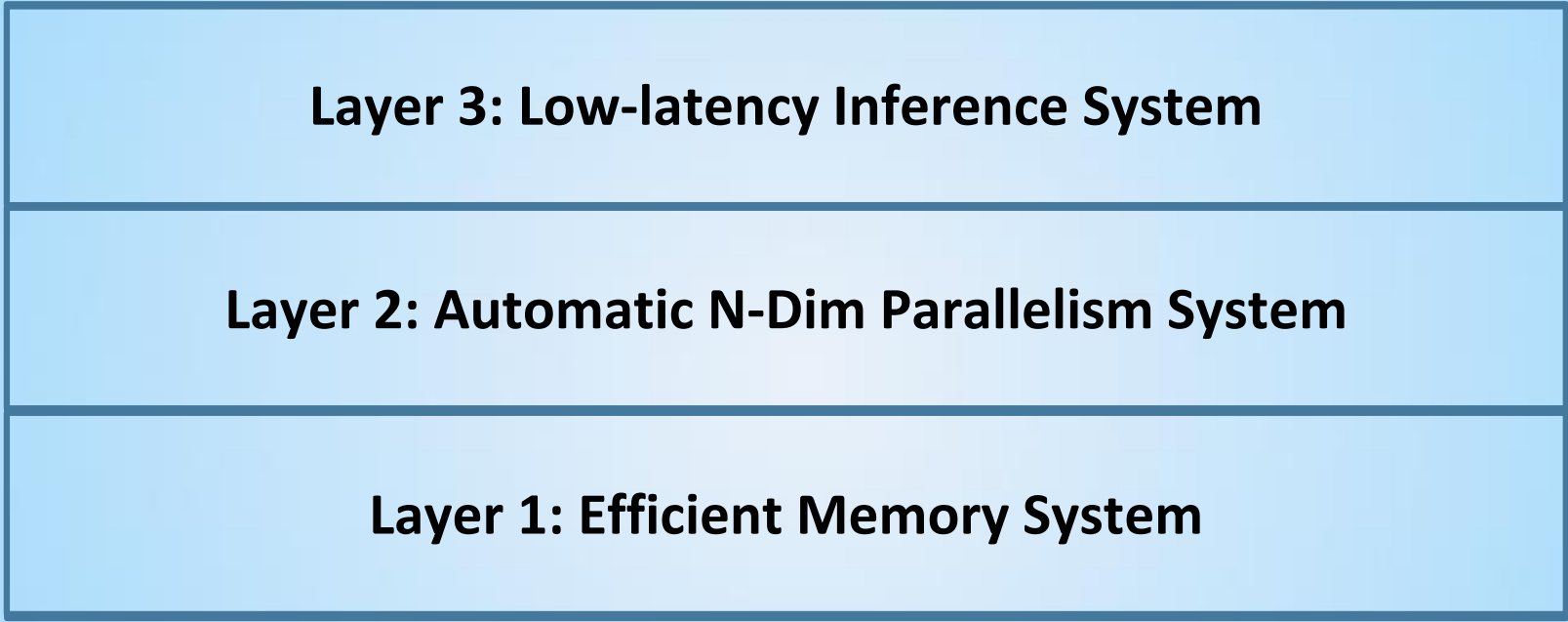
Hardware ➡ HPC-AI Tech ➡ framework

 CPU

 GPU

 TPU

 FPGA









 Hugging Face

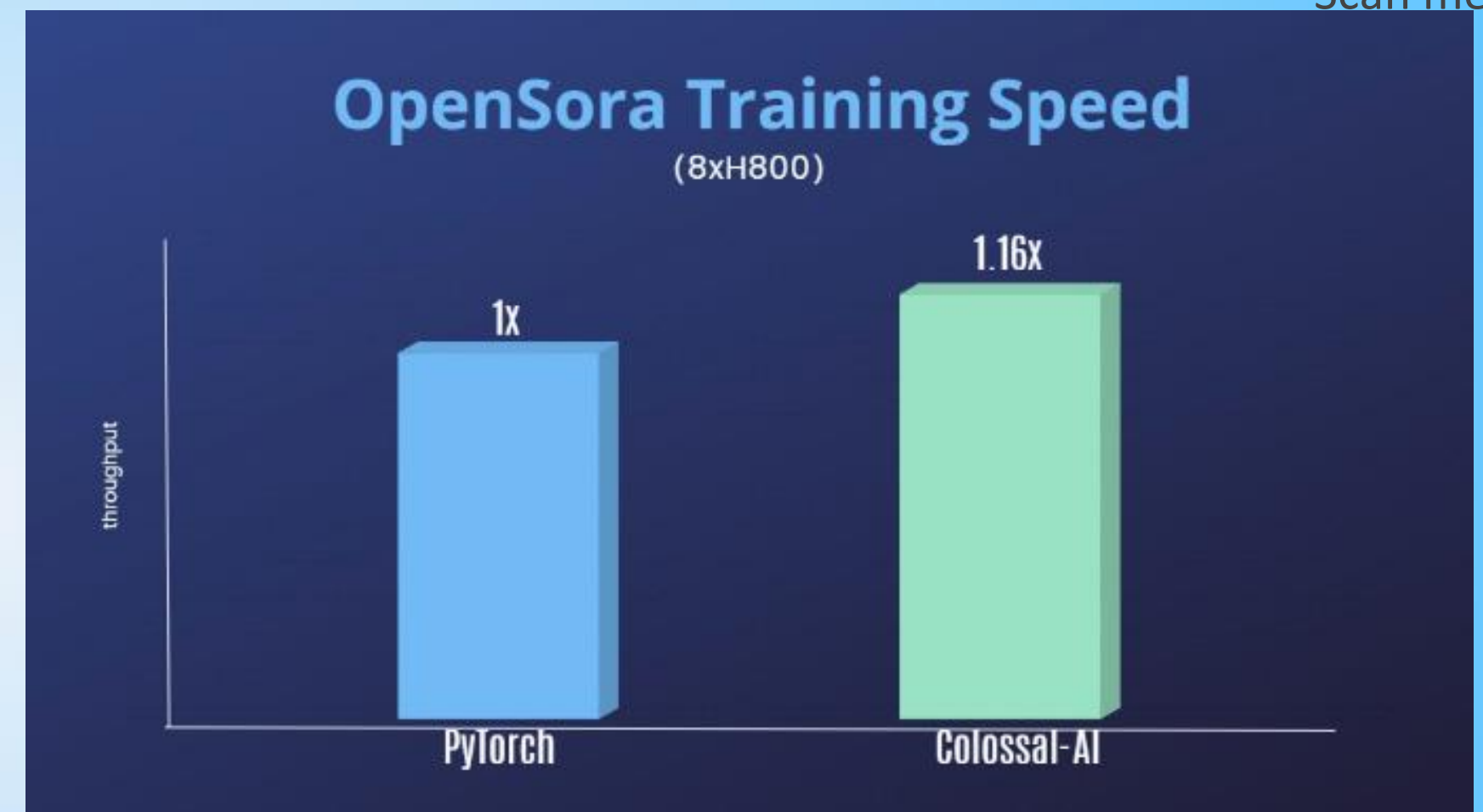
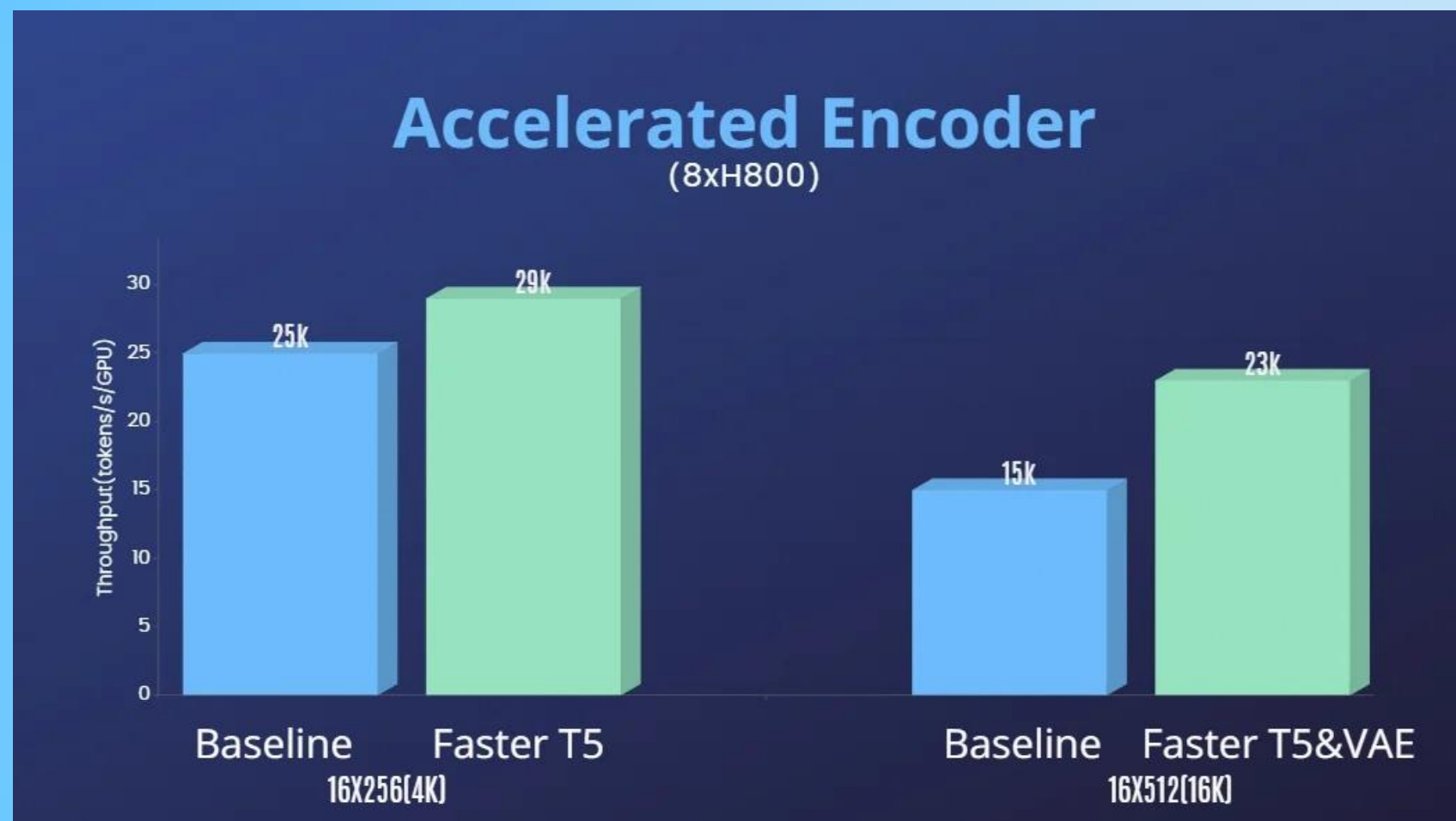


- Minimize Deployment Cost
- Maximize Computational Efficiency

Open-Sora: Efficient Training Strategies



Scan me



- T5's share of the computational load decreased from **13% to 2%** after acceleration.
- Achieves 1.55x training acceleration for 64-frame, 512x512 resolution videos.
- Seamless training of 1-minute 1080p HD video task on a single server (8*H800).
- Colossal-AI provides 16% acceleration on a single server (8*H800).

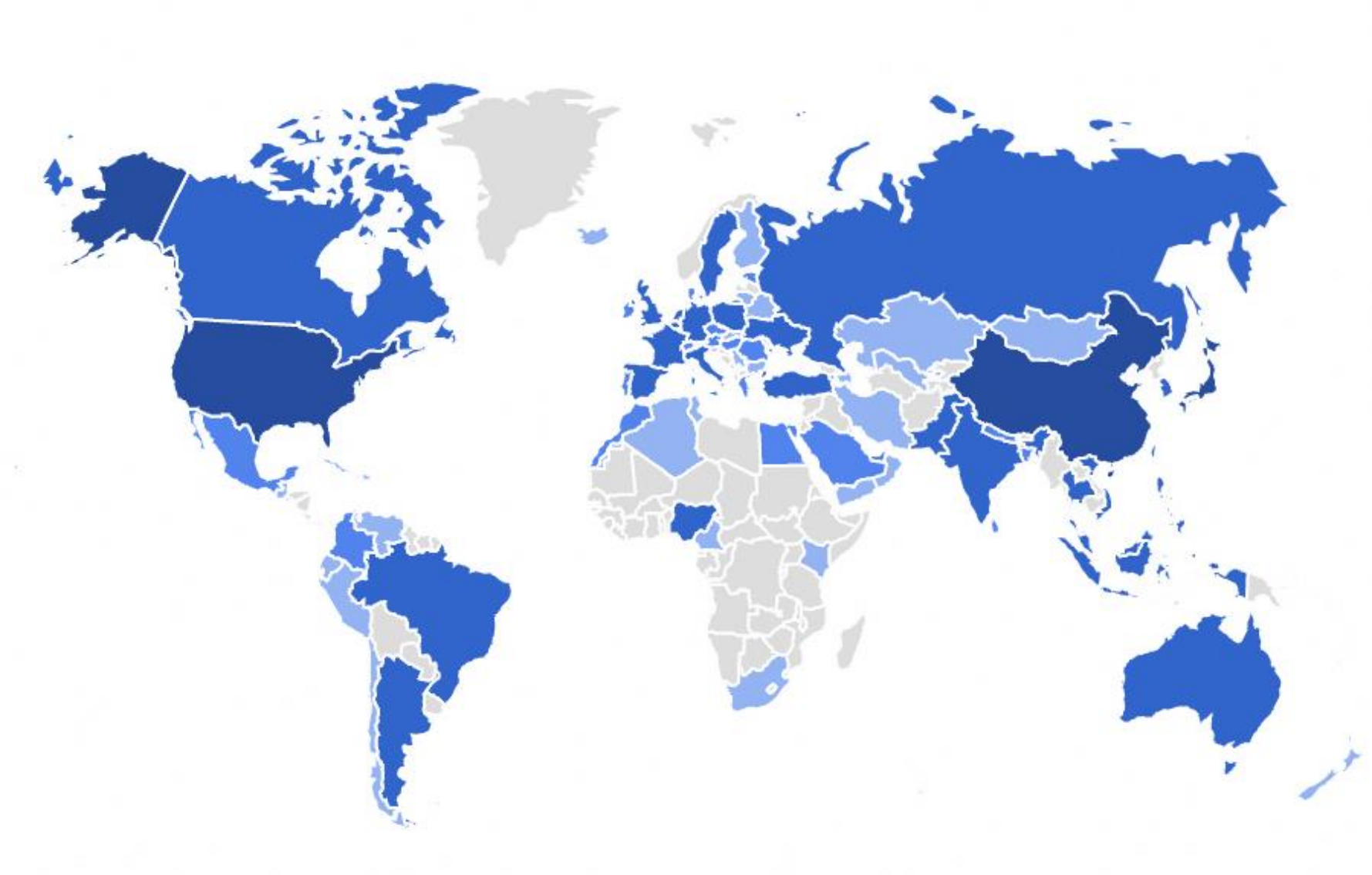


Scan me

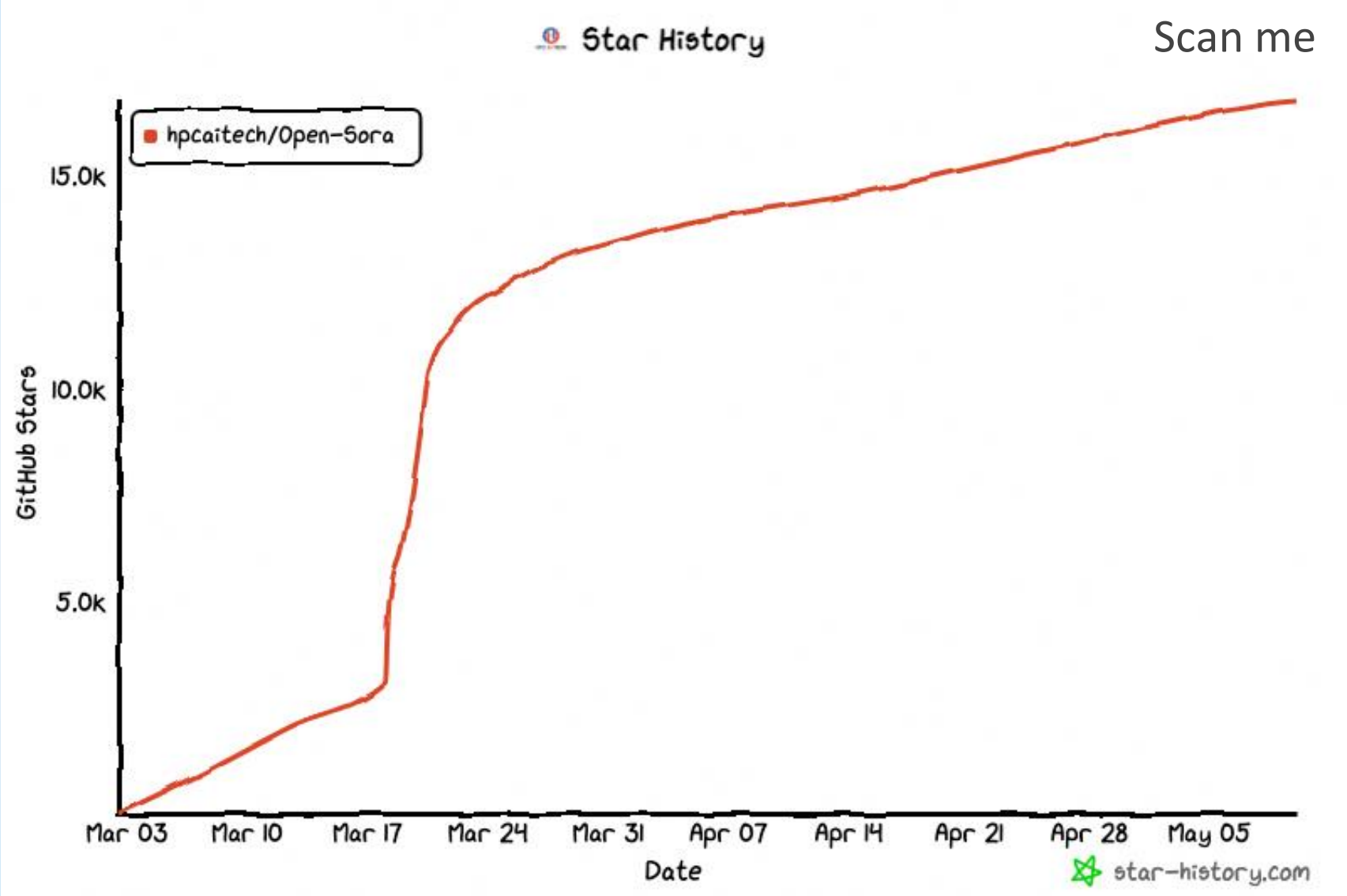
Applications in Practice



Our Model: Over 20K Stars



Global Followers Everywhere





GitHub Stars Surging Rapidly


A Cool Work from LambdaLabs Built Upon Our Model




Scan me

 README

 License






Text2Bricks: Fine-tune Open-Sora For Generating Brick Style Animations.

wandb report Huggingface datasets Huggingface model gradio demo

Welcome to the tutorial for this brick video project! This project is about the make of a customized [Open-Sora](#) model that generates brick style stop animations from text inputs.

You can play with the videos created by the models in this [game](#).

Below are some examples generated by the model, trained as described in this tutorial.



A young man walks alone by the seaside.

HPC-AI.COM: Uber 2.0 Compute



Scan me

Uber

HPC-AI.COM: Uber 2.0 Compute

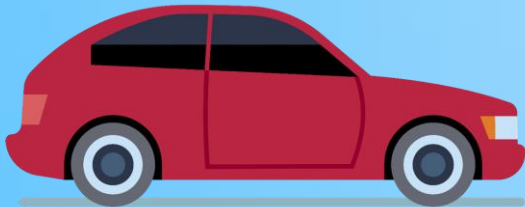


Regular Car



Regular Software + GPU = Efficient Compute

Simple, User-Friendly: Rapid Large Model Training & Inference



Premium Car



High-End Software + GPU = Premium Compute

Low/No-Code Training and Deployment of Large Model Applications
High Performance, Efficiency, Abundant Model Options

Build Your Own Sora-like Model with HPC-AI.COM



- More customizable, scenario-based, and business-oriented video models
- Ready-to-use OpenSora model images
- High-performance compute clusters accelerated by Colossal-AI



HPC-AI.COM

PREMIUM GPU, ANYTIME

USE 1 H100 SERVER FOR JUST \$1



Promotion :

- ✓ H100, H200 GPU RENTAL starting from 1h
- ✓ Refer a friend who successfully adds funds and **earn equivalent credits** or **20% cash**.
- ✓ Optimized for AI & ML at Every Stage
- ✓ Cost-Effective & Reliable

[Register Now!](#)



Thank you

Slack Channel



Twitter



Q&A



**Scan to input
your
questions for
Q&A**

