# Watch the NVIDIA Keynote at CES

# NVIDIA CES 2025 总结&洞察





ZOMI

**Powering Advanced Al** 



# 视频目录大纲

- 1. NVIDIA CES 2025 硬件篇
- 2. NVIDIA CES 2025 软件篇
- 3. NVIDIA CES 2025 总结与趋势





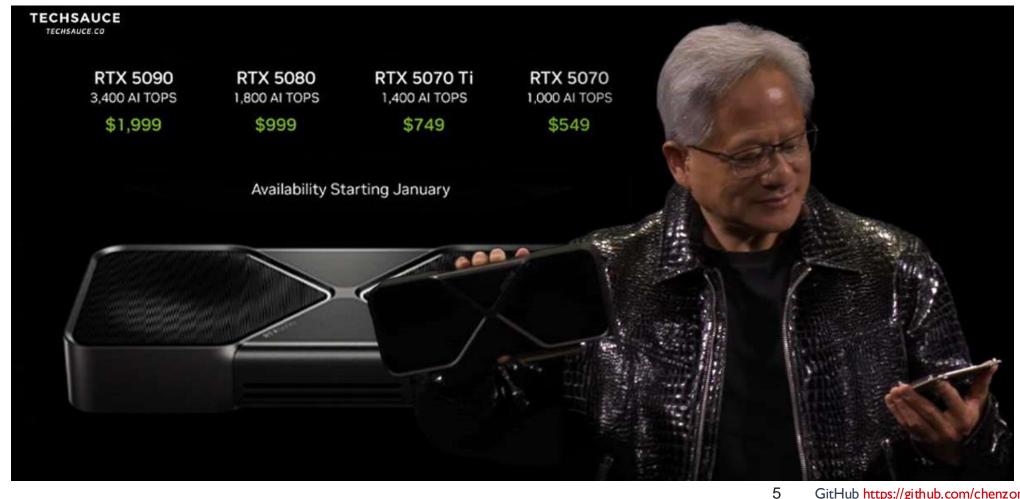
# NVIDIA CES 2025 頃代特篇



- •新一代 GeForce RTX 50 系列芯片重磅发布
- Grace Blackwell NVLink72 巨型芯片亮相
- AI 超级计算机 Project DIGITS 发布
- 一颗 Thor 芯片同时支持智舱、智驾、车联网等功能



• 新一代 GeForce RTX 50 系列芯片重磅发布



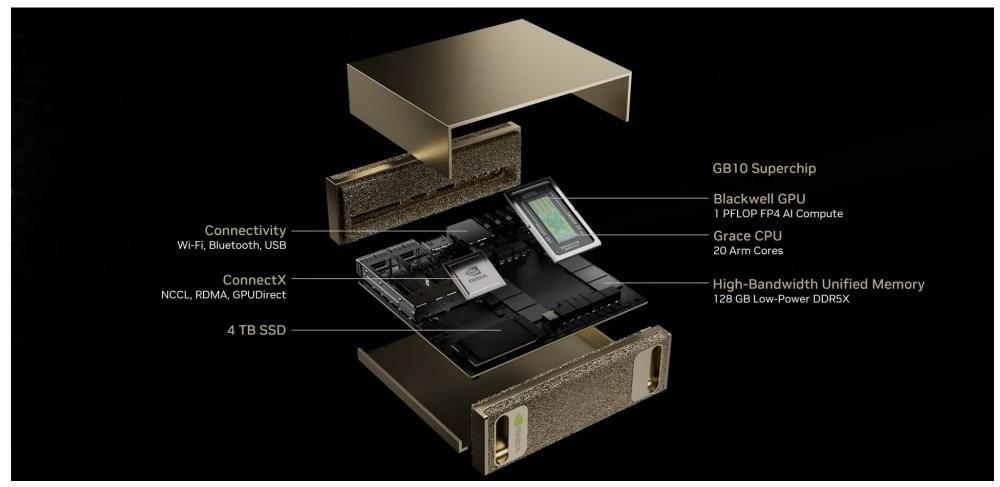


• Grace Blackwell NVLink72 巨型芯片亮相





• AI 超级计算机 Project DIGITS 发布





• 一颗 Thor 芯片同时支持智舱、智驾、车联网等功能



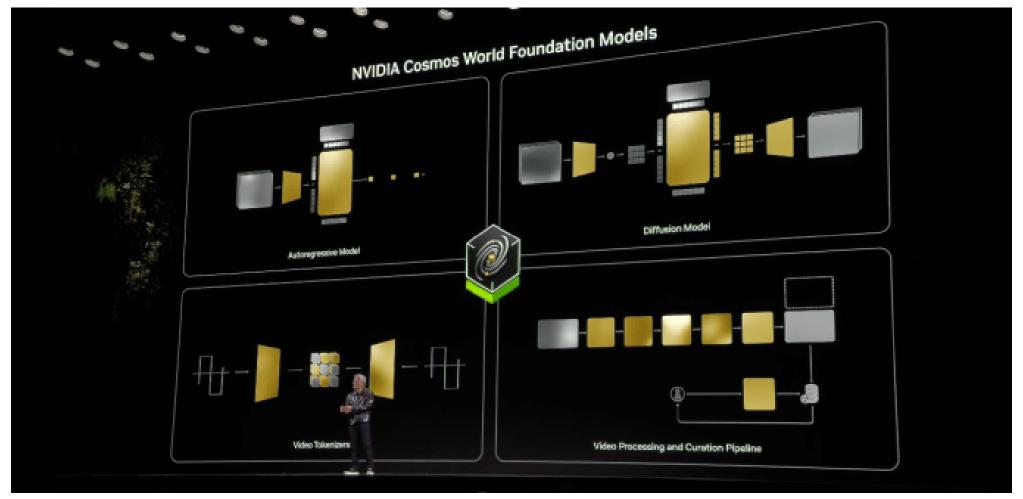


# NIDIACES 2025 探件篇



- 世界基础模型(WFM)开发平台 Cosmos 诞生
- AI 蓝图 (AI Blueprints) 基于 NIM 微服务构建,推动 AI 在 PC 端的普及和应用

· 世界基础模型(WFM)开发平台 Cosmos 诞生





• AI 蓝图 (Blueprints) 基于 NIM 微服务构建,推动 AI 在 PC 端的普及和应用





# NVIDIA CES 2025 总结



### 1. 显卡与AI融合:

全新发布的Blackwell RTX显卡搭载Neuro Shaders AI技术,展示了计算机图形学与人工智能的深度融合, 使图像生成效率和画质大幅提升。

### 2. 世界基础模型Cosmos:

这款模型具备洞察物理世界的能力,可以为自动驾驶和机器人行业提供高精度的仿真数据和动态场景生成。

### 3. Omniverse数字孪生平台:

。 用于训练和测试自动驾驶技术及机器人的虚拟环境,成为机器人技术发展的"试验场"。

# NVIDIA CES 2025 总结

预训练Pre Training: 从千卡到万卡集群,模型往两级分化 后训练 Post Training: 随模型规模增长而需要百卡资源,RL 强化、蒸馏 **Scaling Law** 应用测试时调优 Test Time Tuning: O1 类模型,强调逻辑推理Reasoning Scale out: A100 -> H100-> B100 -> GIX Scale up: NVLink -> NVL36 -> NVL72 大型机(训练 AI) 大型机发展: Grace Blackwell NVLink72 **NVIDA CES 2025** 垂直场景: Thor 车载性能提升,功耗降低,多流接入处理多模态 消费场景: GTX 4090 -> GTX5090, 越来越 AI, Tensor Core作用增加 消费级(部署 AI) AIPC: 推出 Project DIGITS 桌面小型工作站 链接世界: 大型机(DGX 训练 AI), 消费机(AGX 部署 AI), Omniverse+Cosmos 数字孪生 虚实结合: 通过 WFM 让虚拟世界的设计搬到现实世界训练 世界模型 生成世界: 而不是简单生成视频,用于娱乐,更多的是作为与物理世界的交互媒介



# Scaling Law 失效了

1. Scaling Law 到头了吗? 2025 年了还能延续多久?

# Ilya承认Scaling Law到头了,Meta和谷歌寻找替代o1的方法

2024-11-15 20:13:13 来源: 大数据文摘 ② 北京

举报

## OpenAI大改下代大模型方向,scaling law撞墙? AI 社区炸锅了

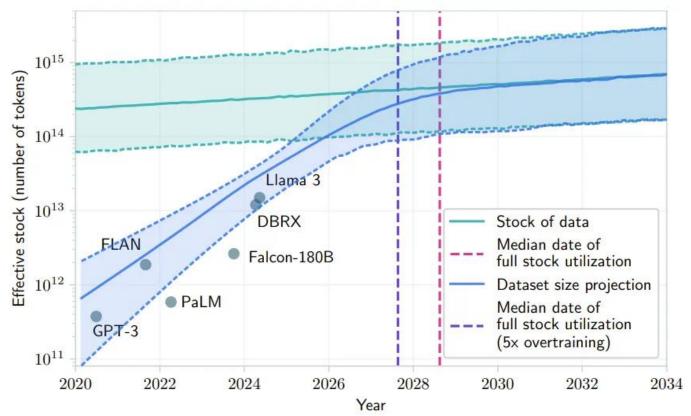
大模型的 scaling law 到头了? 行业龙头 OpenAI 在转换策略。





# Scaling Law 失效了

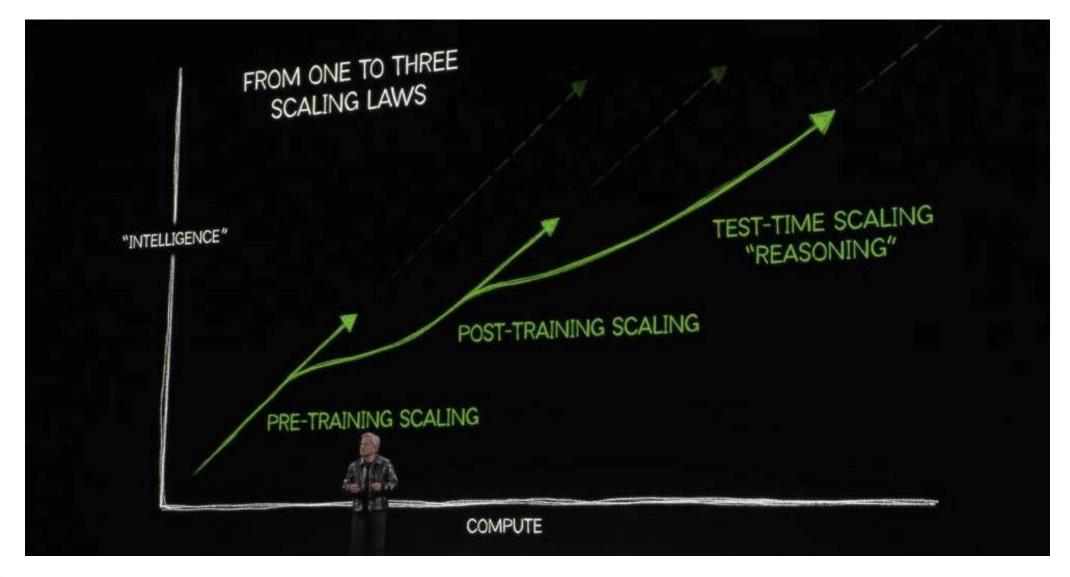
1. 如果 LLM 保持现在的发展势头,预计在 2028 年左右,已有的数据储量将被全部利用完。届时,基于大数据的大模型的发展将可能放缓甚至陷入停滞。







# From ONE TO THREE Scaling Laws





# NVIDIA CES 2025 总结

预训练Pre Training: 从千卡到万卡集群,模型往两级分化 后训练 Post Training: 随模型规模增长而需要百卡资源,RL 强化、蒸馏 **Scaling Law** 应用测试时调优 Test Time Tuning: O1 类模型,强调逻辑推理Reasoning **Scale out:** A100 -> H100-> B100 -> GTX Scale up: NVLink -> NVL36 -> NVL72 大型机(训练 AI) 大型机发展: Grace Blackwell NVLink72 **NVIDA CES 2025** 世且功京· Inor 主载性能提升,切耗降低,多流接入处理多层心 消费场景: GTX 4090 -> GTX5090, 越来越 AI, Tensor Core作用增加 消费级(部署 AI) AIPC: 推出 Project DIGITS 桌面小型工作站 链接世界: 大型机(DGX 训练 AI), 消费机(AGX 部署 AI), Omniverse+Cosmos 数字孪生 虚实结合: 通过 WFM 让虚拟世界的设计搬到现实世界训练 世界模型 生成世界: 而不是简单生成视频,用于娱乐,更多的是作为与物理世界的交互媒介

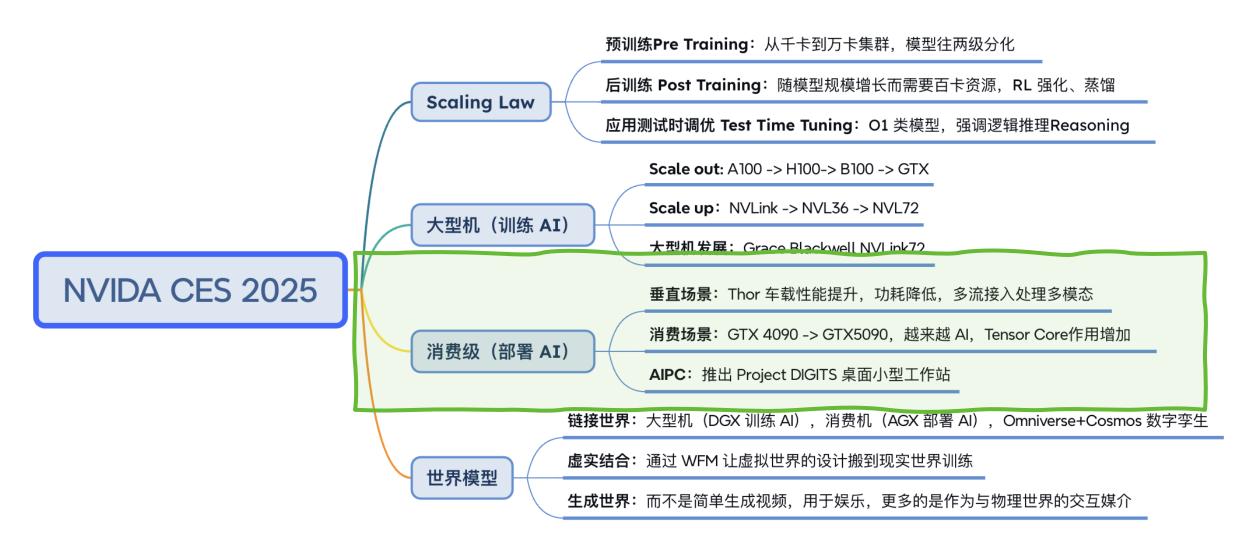


• Grace Blackwell NVLink72 巨型芯片亮相



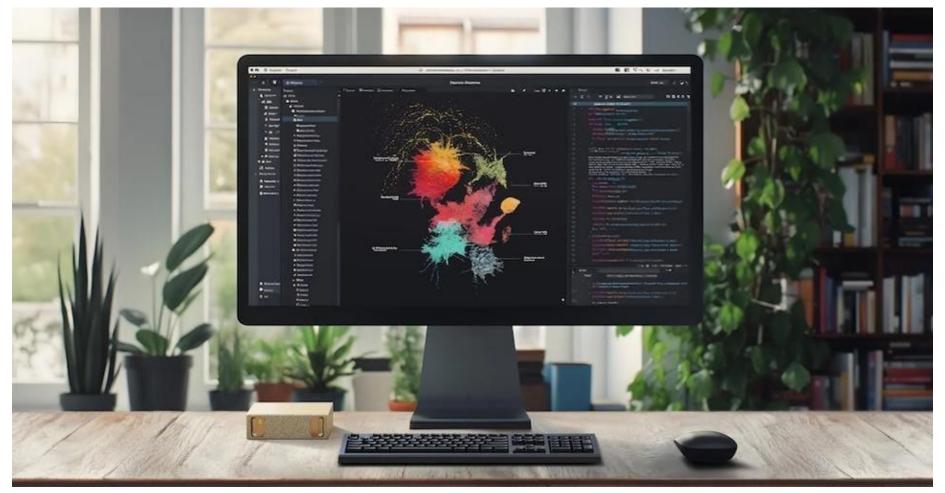


# NVIDIA CES 2025 总结





• AI 超级计算机 Project DIGITS 发布





# NVIDIA CES 2025 总结

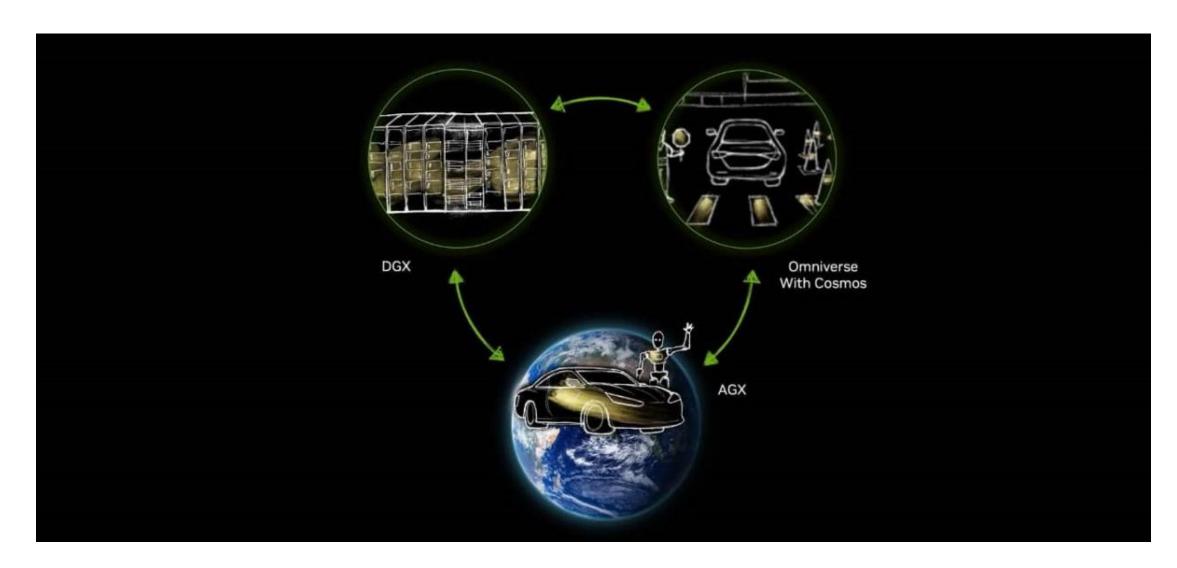
**NVIDA CES 2025** 

预训练Pre Training:从千卡到万卡集群,模型往两级分化 后训练 Post Training: 随模型规模增长而需要百卡资源,RL 强化、蒸馏 **Scaling Law** 应用测试时调优 Test Time Tuning: O1 类模型,强调逻辑推理Reasoning **Scale out:** A100 -> H100-> B100 -> GTX Scale up: NVLink -> NVL36 -> NVL72 大型机(训练 AI) 大型机发展: Grace Blackwell NVLink72 垂直场景: Thor 车载性能提升,功耗降低,多流接入处理多模态 消费场景: GTX 4090 -> GTX5090, 越来越 AI, Tensor Core作用增加 消费级(部署 AI) AIPC: 推出 Project DIGITS 桌面小型工作站 链接世界: 大型机(DGX 训练 AI), 消费机(AGX 部署 AI), Omniverse+Cosmos 数字孪生 虚实结合:通过 WFM 让虚拟世界的设计搬到现实世界训练 世界模型



生成世界: 而不是简单生成视频,用于娱乐,更多的是作为与物理世界的交互媒介

# NVIDIA CES 2025 总结







把AI系统带入每个开发者、每个家庭、 每个组织,构建万物互联的智能世界

Bring Al System to every person, home and organization for a fully connected, intelligent world.

Copyright © 2024 XXX Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. XXX may change the information at any time without notice.



 $Git Hub\ https://github.com/chenzomi I\ 2/AIF oundation$ 

# Github 链接

- https://github.com/NVIDIA/Cosmos/tree/main
- https://github.com/orgs/NVIDIA/repositories?q=cosmos

