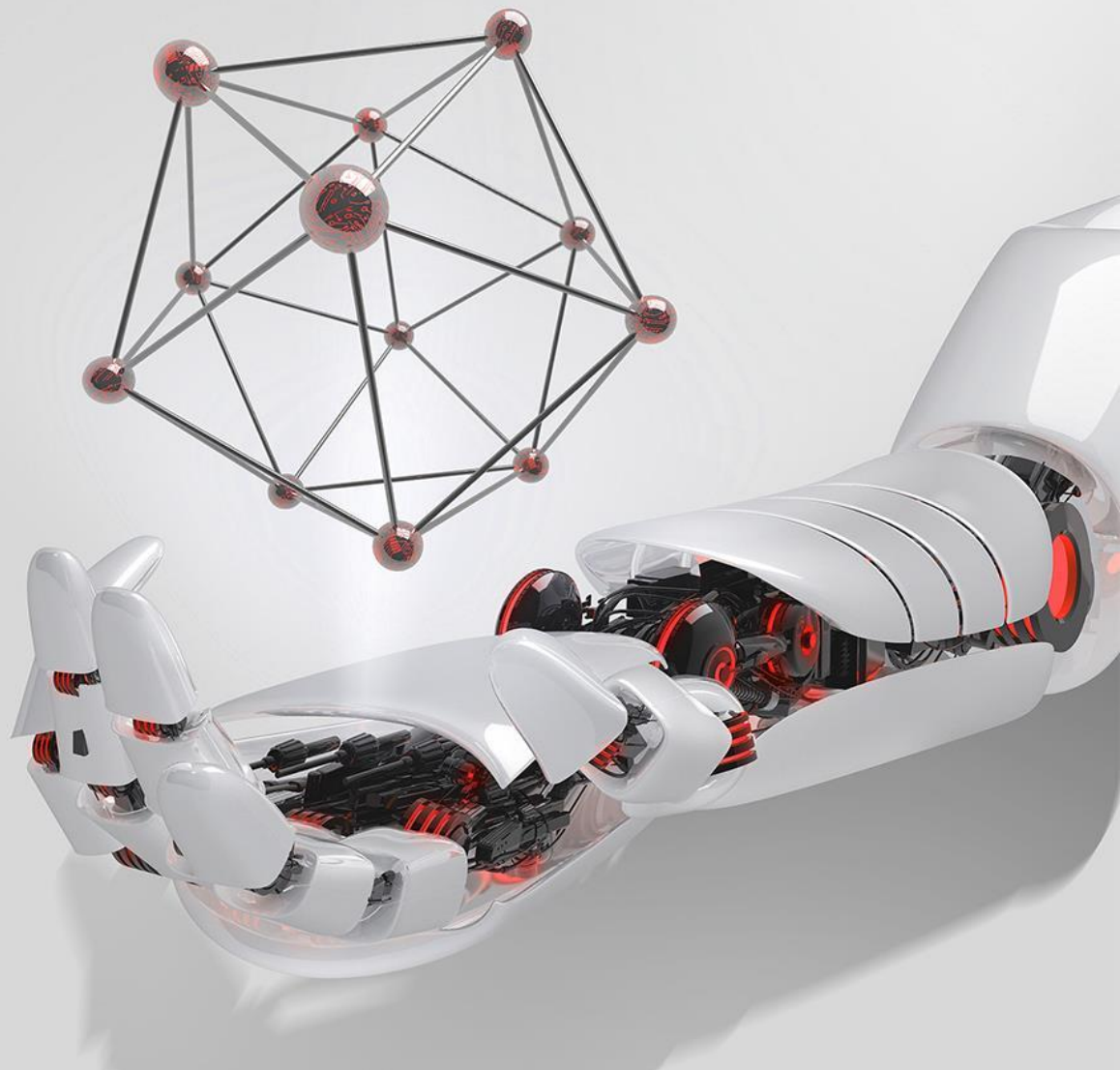


# 《盘古预训练大模型》

## 和我们对大模型的思考



谢凌曦

2021年7月



# 提纲

---

- 问题：作坊式AI开发所面临的挑战和工业化AI开发新模式
  - 面向云服务的AI开发面临的痛点
  - 工业化AI开发的整体设计理念
- 方案：**盘古预训练大模型**的特点和图像预训练技术简介
  - 盘古预训练大模型的特点
  - 盘古预训练大模型的视觉预训练算法
  - 盘古预训练大模型的算力支撑
- 思考：大模型之路走了多远，还有多远？它会通往何方？
  - 大模型在AI发展中的位置
  - 开放性的问题和对未来的期望

# 从作坊式开发到工业化开发，是AI落地的迫切需求

AI走进**企业核心生产系统**，开始创造更大价值



2025年，企业对AI的采用率将达到**86%**（当前仅**4%**）  
华为云EI的AI实践：**600+**项目，其中**30%**进入生产系统

然而，AI落地需求高度**碎片化**  
传统的**作坊式开发模式**，成为AI落地的巨大阻碍

**预训练大模型**，引领AI开发模式向**工业化**转变



开发周期  
**月级 >> 天级**



模型精度  
**提升 10%**

通用能力

适用80%的AI任务

高效能比

绿色AI-高效训练推理

自学习

不需人类标签训练

**千亿规模参数 预训练大模型**

**大模型 ≠ 大算力+大数据**

**大规模数据挖掘**

多样性数据+多尺度多模态归一化预处理

**大规模模型设计**

表达能力更强+训练收敛更快

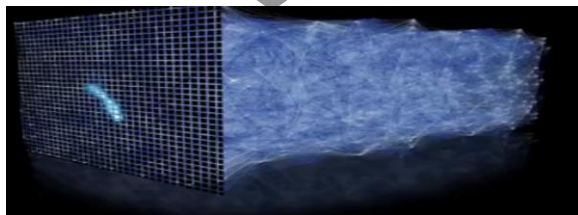
**大规模分布式优化**

大规模集群+数据模型分布

# 工业化AI开发的设计理念

## 超大的神经网络

首个千亿参数中文大模型  
CLUE打榜业界领先



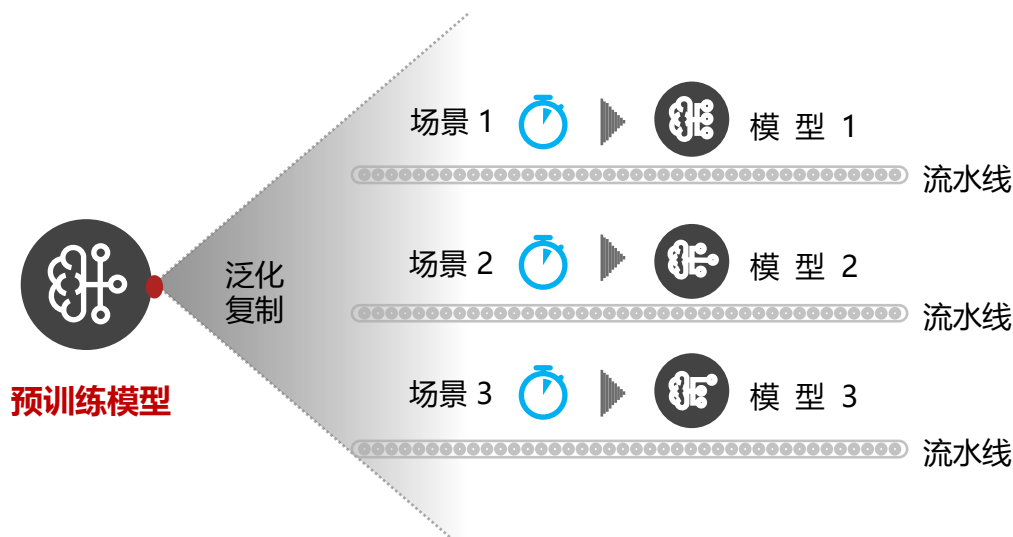
华为云盘古大模型  
「AI工业化开发」

## 强壮的网络架构

大模型综合性能比之前  
提升了10%以上

## 优秀的泛化能力

全场景覆盖率  
提升10倍以上



## 盘古大模型

模型泛化 极大节省训练投入  
流水线 工具集成 训练更快

效率提升

**10 - 100倍**

模型性能更优

预训练

微调&部署

迭代

# 大模型发展的时间线和《盘古大模型》的特点

NLP

首次使用Encoder-Decoder架构  
多任务优化和prompt-based微调



通用知识 x 行业经验

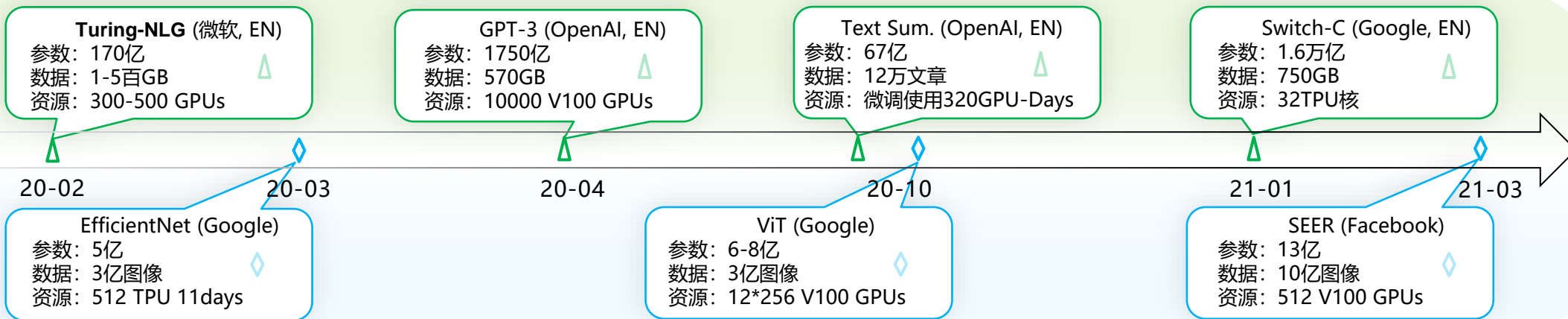


小样本学习超越GPT系列



生成与理解性能领先

大模型  
发展时间线



CV

首次实现模型的按需抽取  
生成-判别损失、基于样本相似度的对比学习



业界最大CV单模型



判别与生成联合预训练



100+ 场景验证

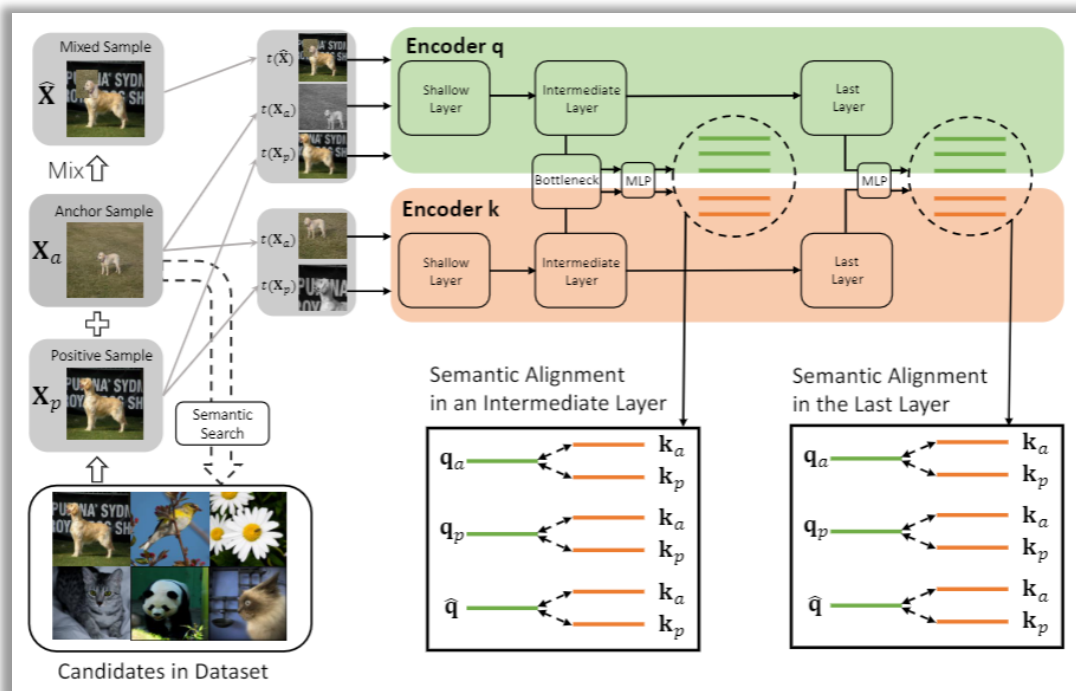


小样本学习性能领先



# 《盘古大模型》 图像预训练技术： 层次化语义混合

- 无监督预训练中， **并非**所有不同样本都应在语义空间中被推远！
  - 算法： 充分利用语义空间近邻样本， 构造混合样本并进行层次化语义对齐
  - 效果： 以ResNet50作为backbone， 线性测试达到76.6%的top-1精度（与有监督持平）， 1%和10%小样本训练分别达到66.7%和75.1%的top-1精度



Method	Accuracy(%)
BigBiGAN[7]	56.6
Local aggregation [37]	58.8
SeLa [33]	61.5
PIRL [19]	63.6
CPCv2 [14]	63.8
PCL [17]	65.9
SimCLRv2 [3]	71.7
MoCo v2 [4]	71.1
BYOL [10]	74.3
SwAV [1]	75.3
CsMI 200 Epochs (w/o multi-crop)	71.6
CsMI 800 Epochs (w/o multi-crop)	74.4
CsMI 200 Epochs(w/ multi-crop)	74.6
CsMI 800 Epochs(w/ multi-crop)	<b>76.6</b>

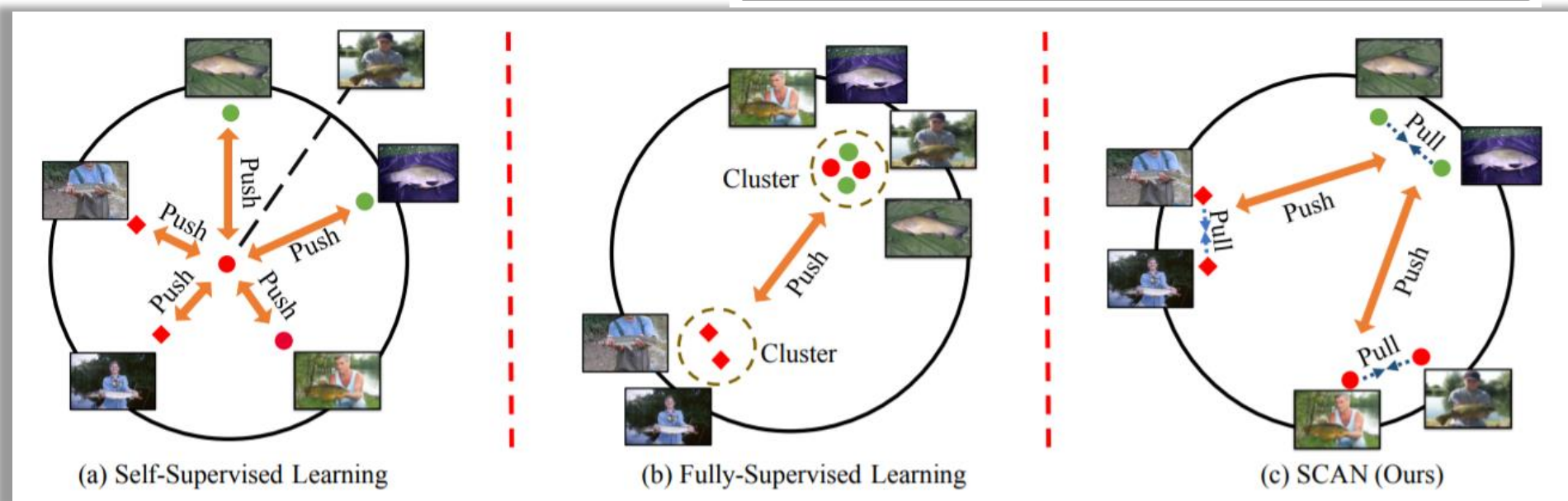
Method	1% labels		10% labels	
	Top-1	Top-5	Top-1	Top-5
Supervised	25.4	48.4	56.4	80.4
UDA[32]	-	-	68.8	88.5
FixMatch[26]	-	-	71.5	89.1
PIRL[19]	30.7	57.2	60.4	83.8
PCL[17]	-	75.6	-	86.2
SimCLR[2]	48.3	75.5	65.6	87.8
MoCo v2[4]	52.4	78.4	65.3	86.6
SwAV[1]	53.9	78.5	70.2	89.9
SimCLRv2[3]	57.9	82.5	68.4	89.2
CsMI	<b>62.2</b>	<b>83.0</b>	<b>72.9</b>	<b>90.7</b>
CsMI+data mining	<b>66.7</b>	<b>87.7</b>	<b>75.1</b>	<b>92.1</b>



# 《盘古大模型》 图像预训练技术：有监督语义调整

- 有监督预训练中，**并非**所有同类样本都应在语义空间中被拉近！
  - 算法：在相同标签样本中寻找表征最近邻样本，仅拉近这些样本，推远其余样本
  - 效果：以ResNet50作为backbone，预训练效果同时超越有监督和无监督预训练，在各项下游任务上保持领先

Dataset	Task	Metric	Pre-Training Method		
			FSup	MoCo	SCAN
VOC	Det	AP <sub>50</sub>	81.4	82.5	<b>83.3</b>
	Seg	mIoU	74.4	73.4	<b>76.6</b>
COCO	Det	AP <sup>bb</sup>	38.9	39.2	<b>40.9</b>
	InsSeg	AP <sup>mk</sup>	35.4	35.7	<b>37.2</b>
LVIS	InsSeg	AP	24.4	25.3	<b>25.4</b>
Cityscapes	InsSeg	AP	32.9	33.1	<b>33.8</b>
	Seg	mIoU	74.6	75.2	<b>76.5</b>



# 《盘古大模型》背后的算力支持



提供澎湃算力

调度鹏城云脑II E级算力

- AIPerf500 AI 算力榜世界排名第一
- IO500吞吐性能全机测试冠军
- 4096颗昇腾910+2048颗鲲鹏
- 每秒10的18次方的浮点算力



智能资源规划调度

训练性能提升35%

- 自适应硬件亲和性调度技术
- 分布式软路由规划技术
- E级算力调度
- 支持万亿级模型训练

CANN

算子性能提升

单算子性能提升30%

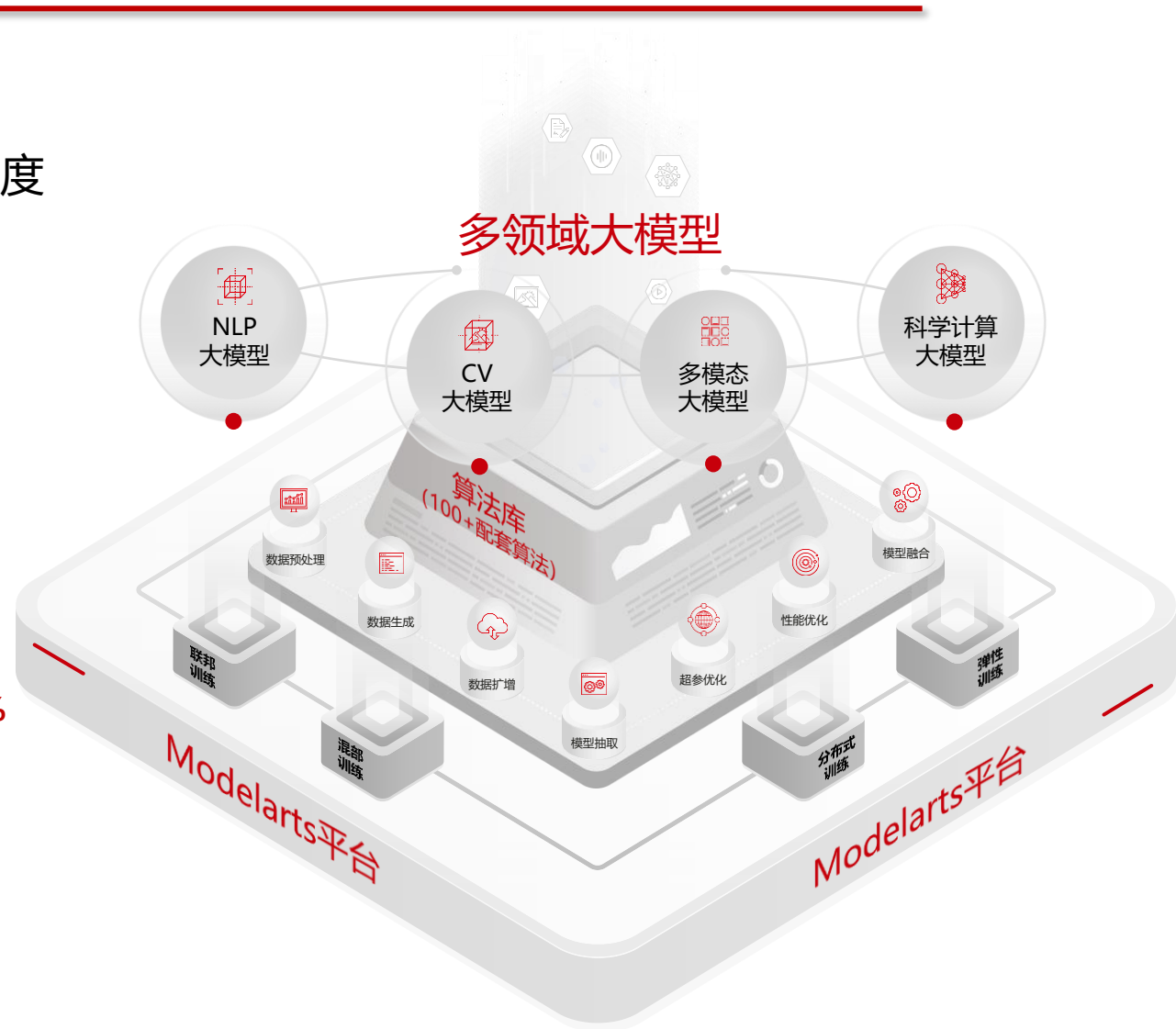
- 算子量化技术
- 混合精度训练技术
- 算子融合优化技术
- 并行通讯优化技术

[M]<sup>s</sup>

提升并行效率

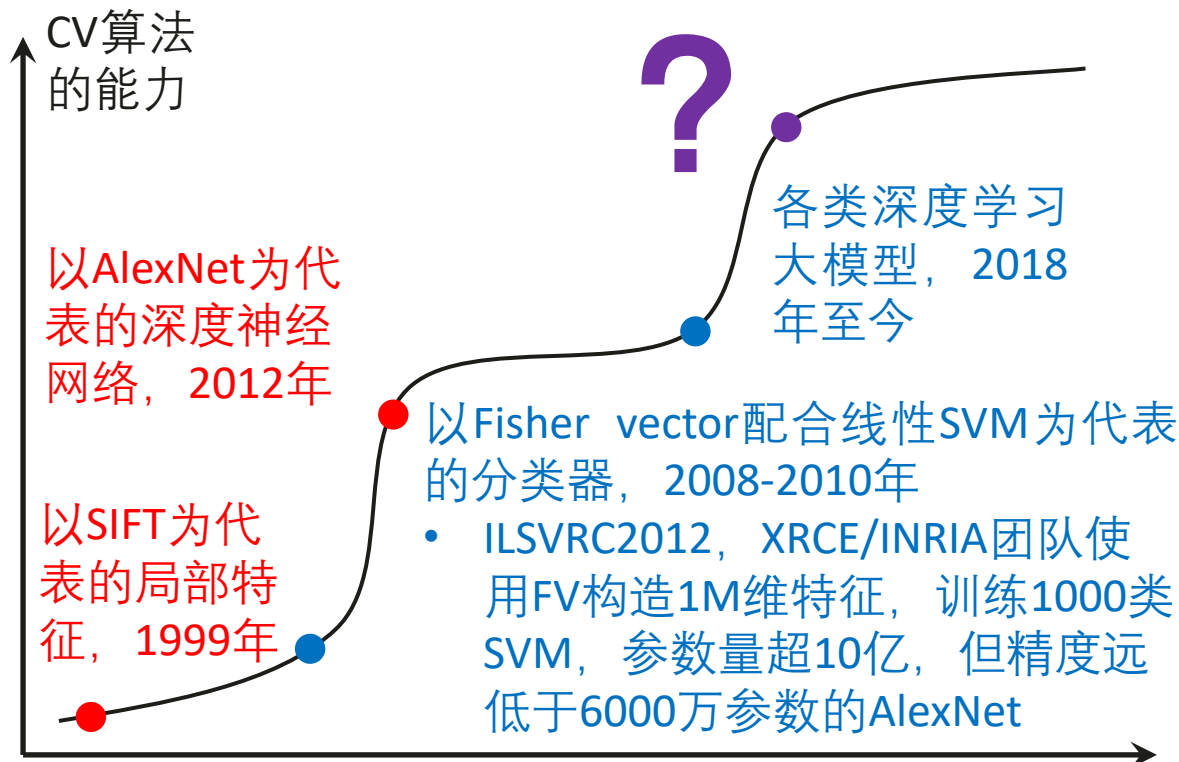
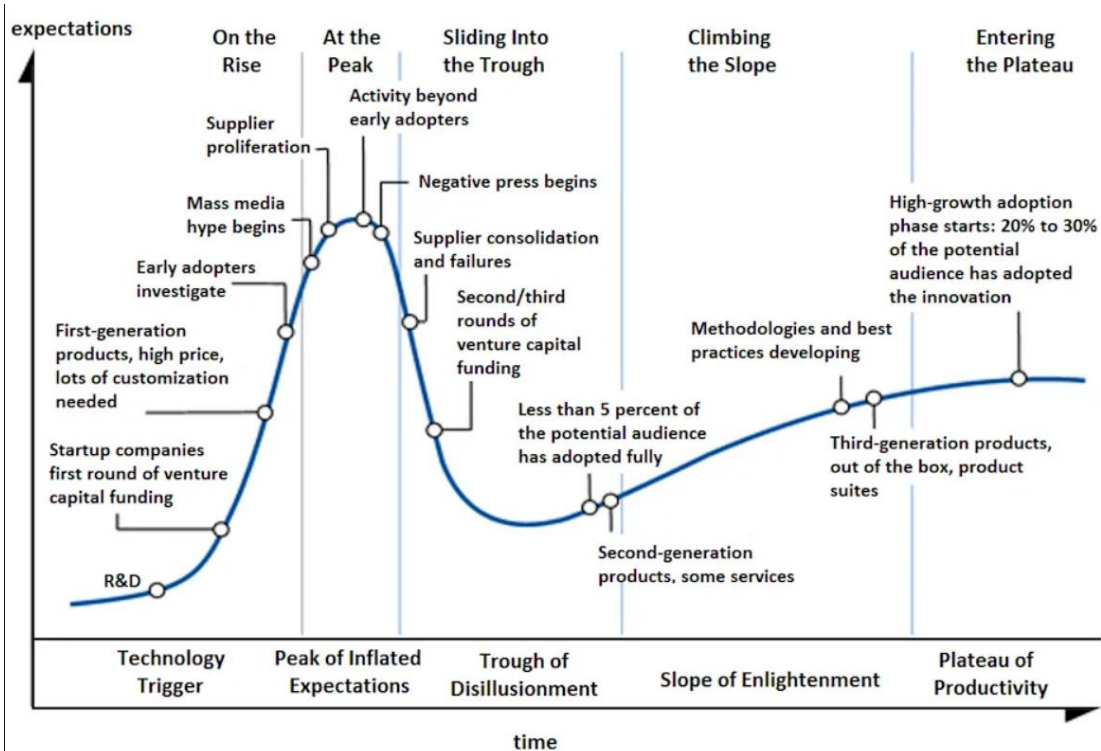
集群线性度提升20%

- 多维自动混合同并行技术
- 模型并行+数据并行
- Pipeline并行
- 前向反向+优化器混合同并行





# 思考：大模型处于AI发展的什么位置？



Gartner 技术成熟度曲线：任何新技术的发展，都要经历科技诞生的驱动期、过高期望的峰值、泡沫化的低谷期、稳步爬升的光明期、实质产出的平原期  
预训练大模型，有可能处于“过高期望的峰值”上！

任何领域的进步，都不是平稳增长的，而是伴随着一段时间出现一次的技术跳跃  
CV领域的上一次技术跳跃，是2012年以AlexNet为代表的深度学习技术，它所取代的正是上一代大模型！

# 一些开放性的问题

---

- 如何评价大模型的强弱？
  - 参数量显然不是最好的标准，下游评测也很tricky——有更好的指标吗？
- 大模型和下游小样本学习会冲突吗？
  - 大模型的基石是大数据，大数据则意味着更广、更散的数据分布
  - 下游迁移，特别是小样本学习，就需要迅速适应小范围的数据分布
- 大模型能否拥有更高级的内部结构？
  - 层次化或者组合式的内部结构，将更有利于大模型的训练和推理
- 大模型如何学习常识和逻辑？
  - 大模型难以掌握基本常识和数理逻辑
  - 学习常识和逻辑，将使大模型完成从记忆到推理的升华

# 总结

---

- 预训练大模型，是深度学习时代，科学和工程思维结合的产物
  - AI的新基建工程，大厂重器
  - 帮助人工智能走向下一个时代
- 喧闹过后，落地更加重要
  - 没有良性商业模式的持续输血，大模型走不了太远
  - 大场景可遇不可求&小场景仍存在挑战
- 我们对大模型应该有怎样的期望？
  - 孕育新技术，就像上一代大模型所做的那样
  - 改变人机交互模式，提升人机交互效率

# 谢谢!

---

- 如果有任何问题和建议，欢迎随时讨论!