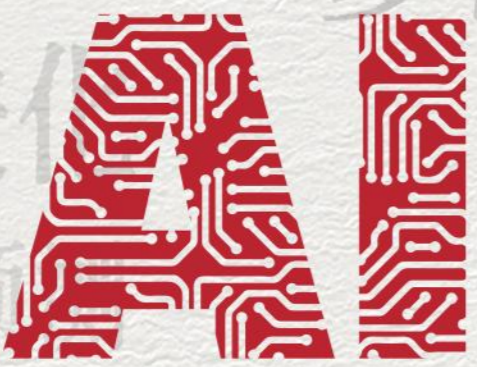


人形机器人 视频生成 3D生成 高分辨率处理  
Agent 多模态融合 图像生成控制 Scaling Law

自我进化  
下一帧预览



# 图景解码

材质还原  
极限压缩

社会模拟  
AI元宇宙

# 50 关键词

AGI  
AI 50  
KEYWORDS

MoE架构  
编程助手

AI芯片  
DiT架构

—— 快思考与慢思考 ——

世界模拟器  
长期记忆

隐私计算 全栈生成 自主执行 端云协同

基础模型

推理debug  
高斯泼溅

智能体协作 慢思考

端侧智



# 序言



AI 50  
KEYWORDS

在过去一年，人工智能技术的迅猛发展正深刻重塑着社会的运行方式。从前沿动态追踪到实际应用场景，从教育辅导到科研创新，AI已然成为推动社会进步的重要力量。

在AI技术快速迭代的背景下，系统性的信息整合与分析变得尤为重要。为降低信息获取成本，提升学习效率，腾讯研究院开发了一系列专业的AI资讯产品：

- AI每日速递，一份高度凝练的日报产品，帮助读者用3-5分钟快速掌握AI领域当日十大关键进展，在信息过载的时代，为学习与研究“标注”出最有价值的高质量数据；
- AI每周50关键词，作为周报产品，基于AI速递内容构建。通过梳理一周热点关键词并制作可交互索引，为研究者提供便捷的“检索增强”工具，助力快速定位所需信息；
- 科技九宫格，一档短视频栏目，以3-5分钟视频形式解读科技热点与关键技术原理。通过可视化呈现，促进读者对前沿技术的理解与讨论，为团队内容优化提供重要“反馈”；

这些产品的运营过程，恰如大语言模型的迭代优化——持续不断地吸收新数据，萃取新知识，产生新洞见。在此基础上，团队还同步开展了AGI专题分析、AGI线上圆桌、AI&Society高端研讨会与AI&Society百人百问等系列研究探讨。

基于全年研究积累的三十余万字AI进展数据库，对当前AI发展进行阶段性总结具有重要意义。为了系统呈现AI发展的关键技术要点和趋势，该报告精选了50个年度关键词，覆盖大模型技术的八大领域：图像处理、视频生成、3D生成、编程助手、Agent、端侧智能、具身智能和基础模型；借鉴大模型的思维特征，创新性

的通过"快思考"与"慢思考"两种维度进行分析，形成了50张AI技术图景卡片。

- "快思考"维度呈现印象卡片，采用人机协同方式完成。项目团队研究人员主导提示词工程与价值判断，把握内容方向；AI系统负责执行，最终绘制输出技术定义、图示与总结语；

- "慢思考"维度则深入分析技术发展的底层逻辑。重点整合研究团队在圆桌讨论和专题研究中的深度思考，借助AI辅助梳理出逻辑链条、本质洞见与趋势判断，为读者勾勒AI发展的脉络与方向。

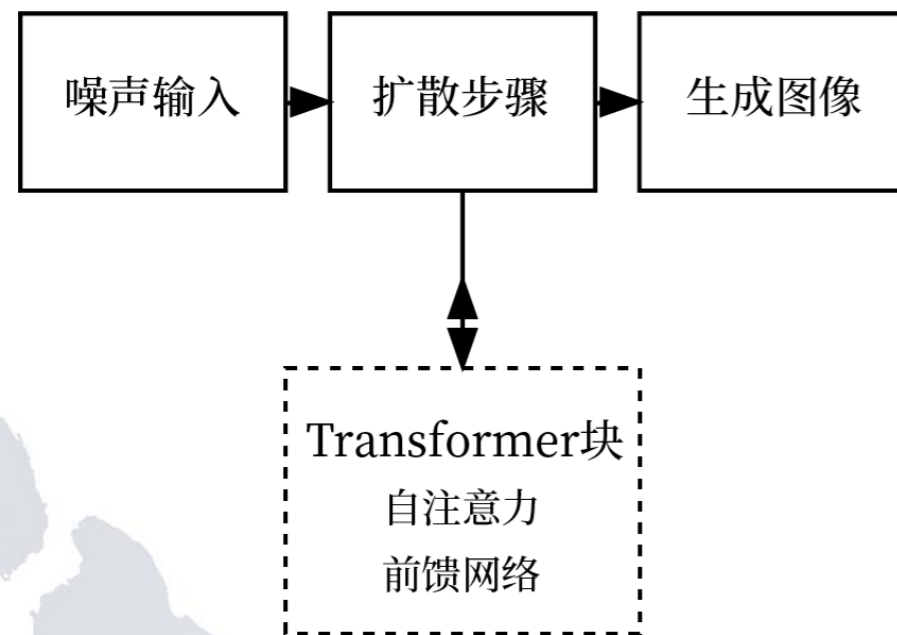
AI技术呈现持续演进、动态发展的特征。该报告通过50个关键词构建的技术图景，旨在展现AI发展的重点领域，把握未来关键趋势，为各界提供研究与决策参考。腾讯研究院将持续深化AI&Society领域的探索，并诚挚的邀请各界好友一共关注与参与，一起迈向一个智能共生的时代。

——腾讯研究院院长 司晓

商业应用 Transformer 早期融合  
DiT架构 开源模型 交叉注意力 商业生态 训练效率  
扩散模型 图像一致性 ControlNet 医学影像 艺术创作  
视觉编码 图像处理 光影效果  
多模态融合 Transfusion 视觉特效  
提示词优化 图像生成控制 细节还原  
像素优化 放射影像 推理速度 场景构图  
高分辨率处理 潜空间技术 光照控制 医疗AI

# DiT架构

结合扩散模型和Transformer的架构，  
用于高质量图像生成的深度学习模型。



**扩散变幻，意象成型**

## 核心观察

- A: Transformer从文本扩展至其它
- B: DiT架构带来图像生成质的飞跃
- C: Scaling Law在图像领域开始生效

## 逻辑链条

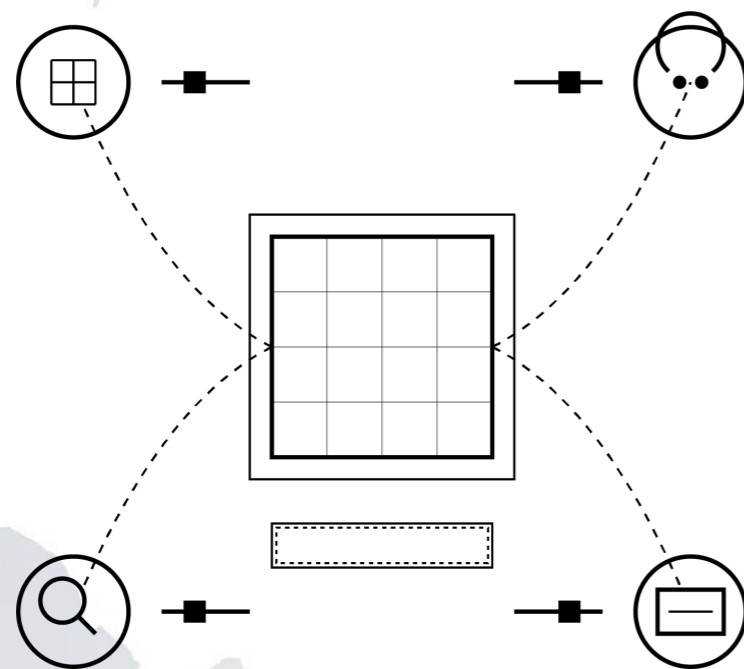
1. A → 技术演进：
  - 下一个字符 → 下一个像素
  - 突破：序列建模能力迁移
2. B → 架构优势：
  - 替代：U-Net → MMDiT (SD3、Flux、混元文生图)
  - 提升：- 空间关系理解 - 复杂提示处理 - 细节还原能力
3. A + B → C: 规模效应
  - 参数规模：800M → 12B
  - 涌现能力：- 真实度提升 - 控制力增强 - 细节完善

## 本质洞见

1. 图像生成正从传统扩散模型走向序列化建模
2. Transformer不同模态的底层范式可以实现迁移
3. 图像领域正在复制语言模型的缩放法则与能力涌现

# 图像生成控制

通过精确的提示词、参数和约束条件，引导AI模型生成符合预期的特定图像内容和风格。



参数为笔，意念成像

## 核心观察

- A: 图像控制从文本描述走向精确控制
- B: 控制方式呈现多层次演进
- C: ControlNet实现精确干预能力

## 逻辑链条

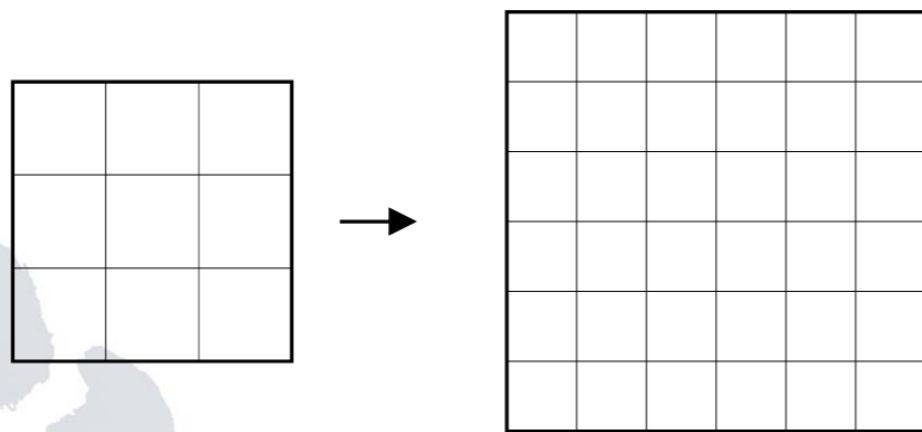
- A → 控制维度演进：
  - Prompt: 文本描述引导
  - LoRA: 低成本模型微调
  - ControlNet: 精确条件控制
- B → 技术路径分化：
  - 描述控制: 语义理解
  - 参数控制: 模型微调
  - 条件控制: 额外输入引导
  - workflow 控制: 外部编排
- C → 精确控制突破：
  - 光影: IC-Light照明控制
  - 轮廓: Paints-Undo创作追溯
  - 构图: Omost自动扩展

## 本质洞见

1. 控制正从"描述性"向"操作性"演进
2. 多层次控制机制形成互补优势，图像生成正走向"精工制造"时代
3. AI图像生成正从粗放生成走向精确控制，这将重塑创作生产流程。

# 高分辨率图像处理

对大尺寸、高精度图像进行分析、增强和变换，以提取信息、改善质量或适应特定应用需求。



细微入毫，尺显真容

## 核心观察

- A: AI图像处理已突破1K分辨率门槛
- B: 高分辨率对图像生成具有重要商业价值
- C: 高分辨率对图像理解同样关键
- D: 高分辨率处理仍存在多重技术限制

## 逻辑链条

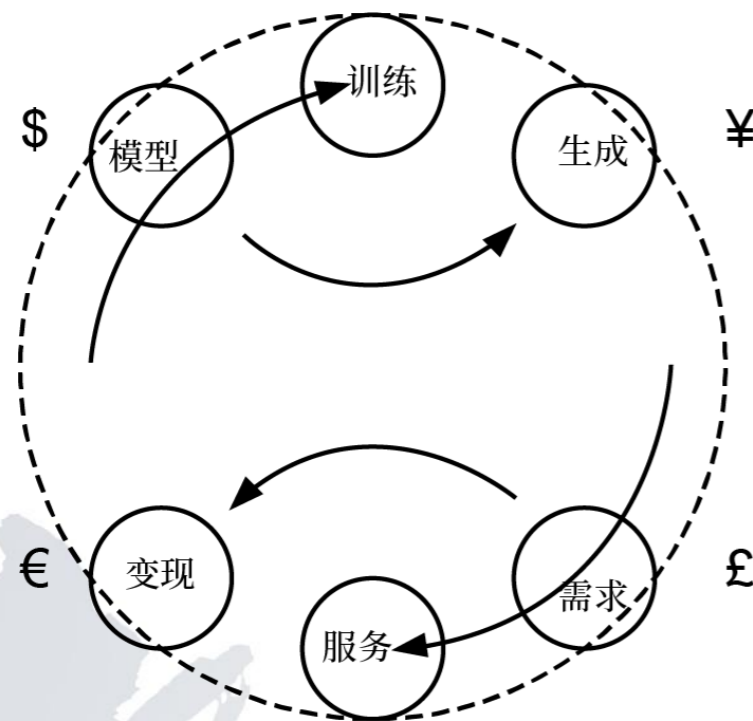
1.  $A \wedge B \rightarrow$  产业应用扩展  
生成模型支持1024×1024原生分辨率  
 $\rightarrow$  艺术创作/广告/游戏开发等应用提升
2.  $A \wedge C \rightarrow$  专业领域突破  
医学影像分辨率需求[256-1024]  $\wedge$  模型达到1K处理能力  
 $\rightarrow$  专业应用可行
3.  $D \rightarrow$  技术演进方向  
 $\forall$ (高分辨率处理)  $\rightarrow$   $\exists$ (架构创新  $\vee$  性能优化)  
例: Pixtral 12B, Eagle系列针对分辨率优化

## 本质洞见

1. 1K分辨率是AI图像处理由通用向专业化过渡的分水岭
2. AI图像价值实现需要技术与产业的双向驱动，专业需求  $\wedge$  技术突破  $\rightarrow$  应用深化

# AI图像商业化

将人工智能图像生成技术转化为可持续的商业服务，实现技术价值与市场需求的良性循环。



智造赋能，价值衍生

## 核心观察

- A: 生成式AI具有强大的技术能力和流量吸引力
- B: 技术能力需要转化为有效商业闭环
- C: AI企业被迫进行商业模式转型
- D: 产业整合成为主要出路

## 逻辑链条

1.  $A \wedge \neg B \rightarrow C$  为什么要转型?  
技术能力  $\neq$  商业价值  
流量优势  $\neq$  变现能力  
(技术优势  $\wedge$  商业化失败)  $\Rightarrow$  寻求转型
2.  $C \rightarrow D$  案例佐证:
  - Stability AI: 技术困境  $\rightarrow$  视效领域融合
  - Leonardo.ai: 独立运营  $\rightarrow$  平台整合
3.  $\forall$ (成功转型案例)  $\rightarrow$   $\exists$ (产业链整合  $\vee$  场景深耕)
  - 原生技术  $\rightarrow$  工具产品  $\rightarrow$  产业解决方案

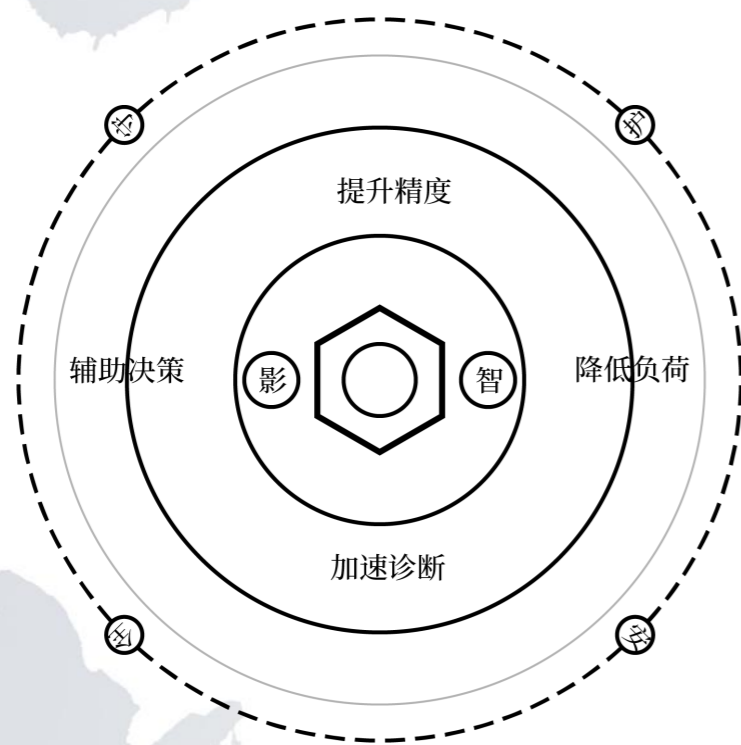
## 本质洞见

1. 想要跨越鸿沟，要么融入既有产业链，要么能够成功对接具体应用需求
2. AI图像生成企业将"技术驱动"向"场景驱动"转变，通过产业整合获得商业生态位



# 医疗AI

运用人工智能分析医学影像、临床数据，协助医生诊断决策，实现精准治疗增强。



慧眼穿透，微显著知

## 核心观察

- A: 图像理解在医疗领域率先实现商业化
- B: 科技巨头深耕医疗AI研发
- C: 学术界取得突破性进展
- D: 行业权威对医疗AI持积极态度

## 逻辑链条

1. A → 产业成熟度与应用价值  
图像理解 > 图像生成 专业应用 > 通用应用
2. B ∧ C → 技术进步  
企业投入: - Med-Gemini系列(2D/3D/基因组)  
学术突破: - Mirai(预测诊断) - SAT(3D分割)  
技术突破 → 临床验证 → 商业应用  
 $\forall(\text{成功医疗AI}) \rightarrow \exists(\text{专业性} \wedge \text{实用性} \wedge \text{可靠性})$
3. D → 发展趋势  
领域专家认可 (Hinton、吴恩达等) ⇒ 技术路线可靠性

## 本质洞见

1. 多模态识别能力提升，让AI在专业领域理解、分析应用成为可能
2. 医疗AI的成功得益于其深度对接专业场景，以解决实际临床需求为导向的发展路径

AI原生创作工具 影视制作  
创作工具 画面保真

下一帧预测 内容创作  
开发框架 世界模拟器

场景理解 动作流畅 细节还原 时序控制

光影效果 镜头语言 **视频生成** 实时生成  
运镜控制

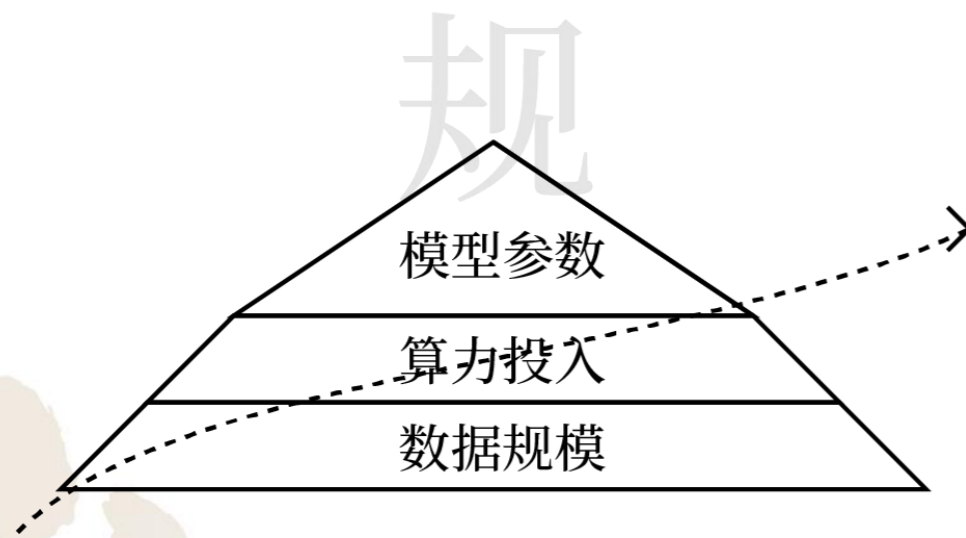
分辨率 生成式游戏 生成速度  
视觉编辑 多模态融合

主体一致性 高清渲染 规模化 物理模拟

艺术家共创 游戏视频  
帧间插值 摄像机控制

# 规模化训练

通过扩大模型参数、数据规模和算力投入，  
在量变中实现质变的训练范式。



以量取胜，跃迁超萃

## 核心观察

- A: 视频生成相比图像生成难度提升百倍
- B: 视频生成技术发展出自回归与扩散两大路线
- C: Sora引领DiT架构成为主流方向
- D: 规模化训练是实现高质量视频生成的关键

## 逻辑链条

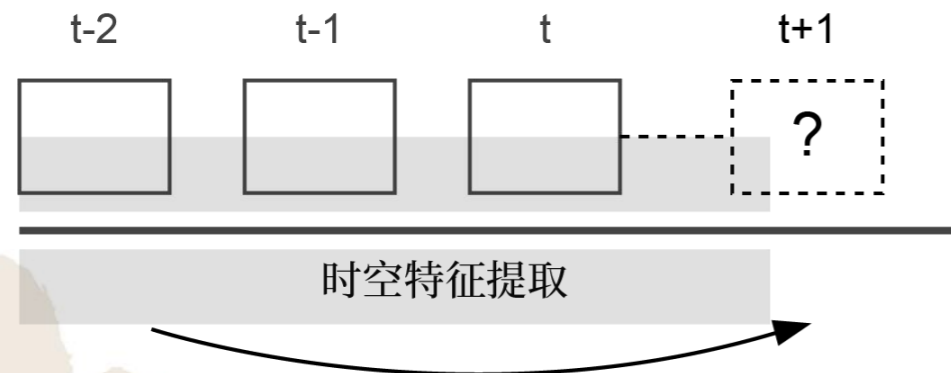
1. 问题难度跃升 (A)  
视频生成 / 图像生成  $\approx$  百倍复杂度: 大量连续帧、时序连贯性、主体一致性
2. 技术路线探索 (B)  
自回归Transformer方案 or 扩散模型  $\rightarrow$  规模化训练
3. 技术突破与统一 (B  $\rightarrow$  C  $\rightarrow$  D)  
DiT架构整合:
  - 融合Transformer与扩散模型优势
  - 引入时空块编码创新
  - 通过规模化训练实现性能突破

## 本质洞见

1. 视频生成的技术演进呈现"分散探索 $\rightarrow$ 路径统一"的特征
2. 规模化训练是解决复杂生成任务的通用范式
3. 架构创新 (DiT) + 训练范式 (规模化) 的组合是突破性进展的关键

# 下一帧预测

基于已知视频帧序列的时空特征，推演预测未来瞬间的画面内容。



窥今以知来，推果溯因

## 核心观察

- A: 视频生成模型的核心在于时序特征处理
- B: DiT架构通过扩散过程处理时序关系
- C: 自回归方案将视频离散为可预测的token序列
- D: 下一帧预测是视频连续性的关键保证

## 逻辑链条

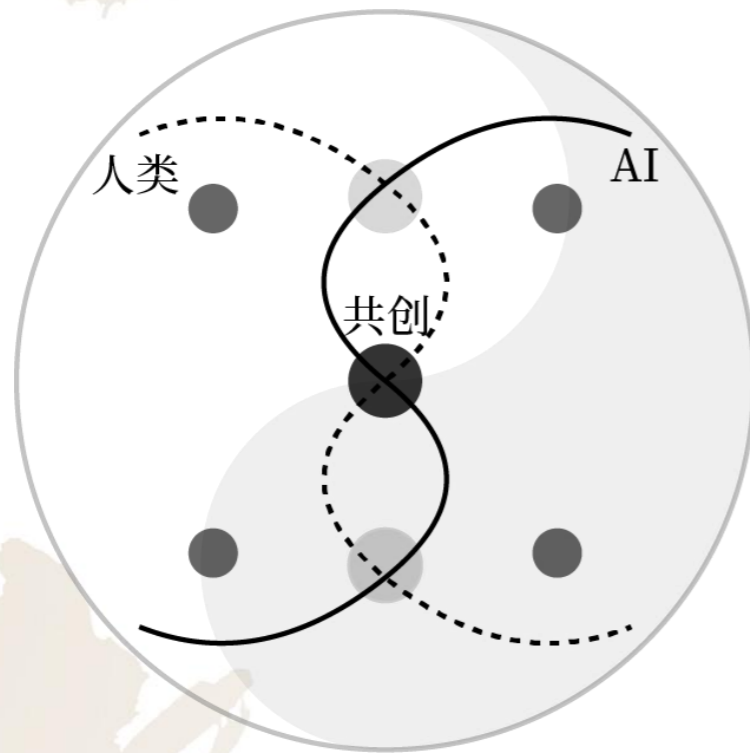
1. 技术路线分化 (A)
  - 扩散模型：噪声迭代 → 帧序列生成
  - 自回归Transformer模型：token预测 → 帧序列构建
2. 实现机制对比 (B  $\wedge$  C)
  - DiT方案：整体扩散过程、时空特征同步建模
  - 自回归Transformer方案：视频token化、序列化预测
3. 预测能力 (D)
  - 连续性保证：时序特征学习 + 运动规律理解 + 状态迁移预测

## 本质洞见

1. 下一帧预测是视频生成的核心任务，但不同技术路线有不同实现方式
2. 自回归预测通过将视频离散化，把复杂的时序预测转化为token预测问题
3. 预测范式的选择直接影响模型的生成能力与效率权衡

# 艺术家共创

人类艺术家与AI模型通过交互式创作，在视频生成过程中实现创意的双向激发与融合。



天人合一，机艺双馨

## 核心观察

- A: 模型厂商通过工具+社区培育创作生态
- B: 厂商频繁举办各类创作比赛扩大影响
- C: 与艺术家合作已成为行业标配
- D: 艺术家参与可反哺模型训练形成数据飞轮

## 逻辑链条

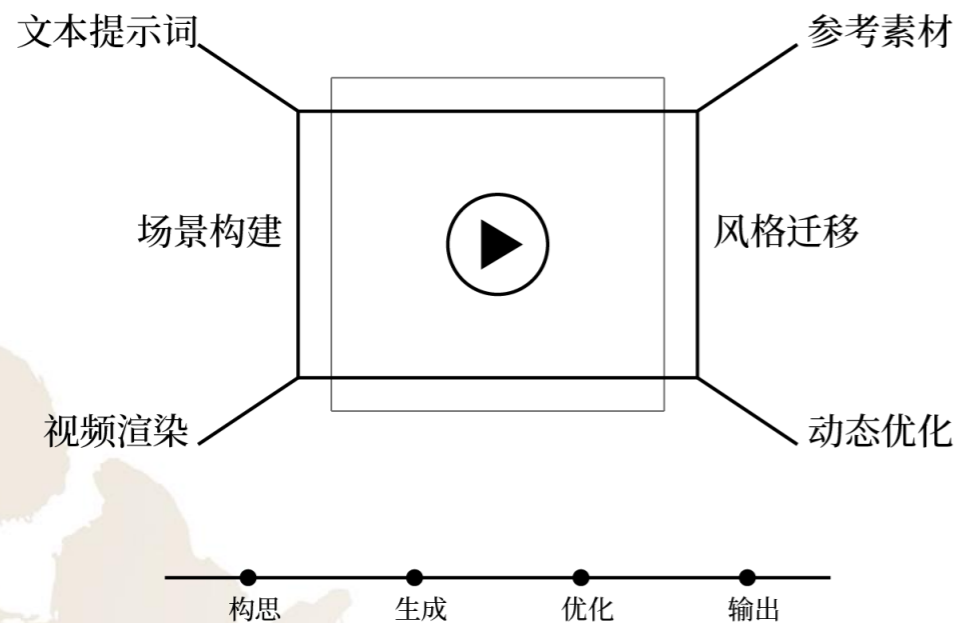
1. 生态构建路径 ( $A \wedge B$ )
  - 打造工具产品 → 运营社区
  - 举办比赛活动 → 扩大影响
2. 艺术家价值链 ( $C \rightarrow D$ )
  - 前端：优质作品展示
  - 中端：专业反馈收集
  - 后端：训练数据优化
3. 闭环形成 ( $A \wedge B \wedge C \rightarrow D$ )
  - 工具应用 → 社区运营 → 艺术家合作 → 数据反馈 → 模型优化

## 本质洞见

1. AI视频生态正在从"工具提供"向"价值共创"演进
2. 艺术家在生态中扮演双重角色：既是内容生产者，也是模型优化的关键贡献者
3. 数据飞轮成为商业闭环的核心驱动力，将持续提升AI创作的质量边界

# AI原生创作

以AI视频生成模型为核心  
重构视频创作的思维范式与 workflow



智成影像，创意无界

## 核心观察

- A: 视频生成模型的交互逻辑比文本模型更复杂
- B: 模型能力支持多样化输入（文本、图片、视频）
- C: 配套工具提供细粒度控制选项
- D: AI原生创作工具正在向全流程方向演进

## 逻辑链条

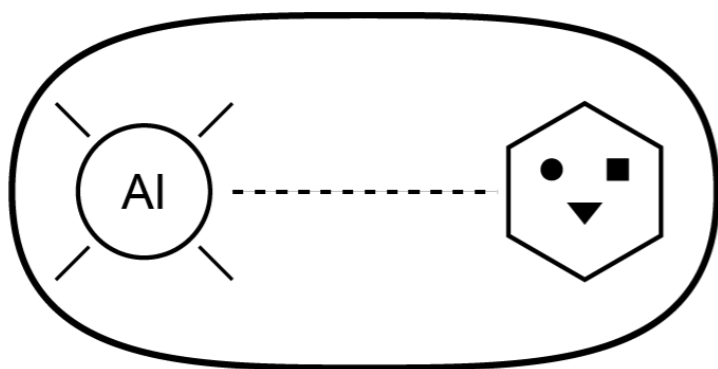
- 交互使用难度差异 (A): 文本模型(简单) → 视频模型(复杂)
- 能力扩展 (A → B)  
文生视频 → 图生视频 → 首尾帧控制 → 视频生视频
- 控制增强 (B → C)  
精细化控制 = 运镜控制 + 运动笔刷 + 主体选定 + ...
- 范式升级 (C → D)  
传统创作工具 → AI原生创作平台 (多工具集成+ workflow 适配+专业功能对标)

## 本质洞见

- 视频生成模型的应用正在从"单一生成"向"创作生态"演进
- 成功的AI创作工具需要在保持AI能力优势的同时，兼顾传统创作习惯
- 降低使用门槛与提供专业控制是视频生成模型应用的双重任务

# 生成式游戏

通过AI生成技术动态创造游戏内容，  
实现无限可能的交互叙事与世界构建。



程序生成

动态交互

玩家创造

无限想象，生生不息

## 核心观察

- A: 视频生成模型正向游戏世界模拟方向发展
- B: 多个研究团队在游戏生成领域取得突破
- C: 游戏引擎本质是一种受限的世界模型
- D: 从游戏模拟到现实世界模拟存在复杂度跨越

## 逻辑链条

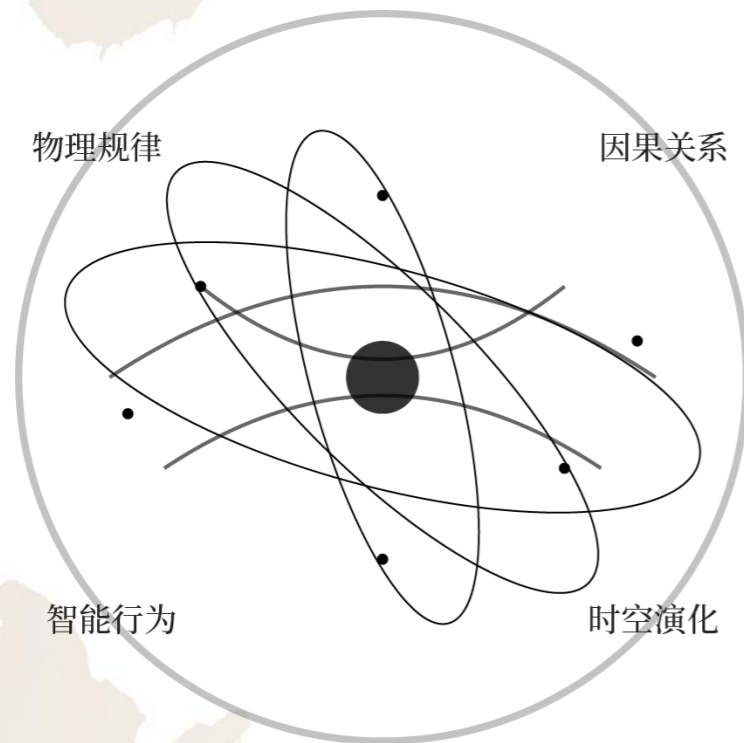
1. 技术演进路径 (A → B)
  - Oasis: Minecraft式开放世界生成
  - Genie-2: 通用可交互游戏生成基础模型
2. 概念拓展 (B → C)  
游戏引擎 ⇔ 受限世界模型  
特征: 有限世界尺寸、封闭规则系统、可预测状态转移
3. 仍待探索 (C → D)  
现实世界 = 复杂度指数级增长 + 数据收集成本激增 + 状态空间爆炸

## 本质洞见

1. 游戏生成是通向世界模拟的"缩微实验场", 提供了可控的技术验证环境
2. 从游戏到现实的跨越不仅是量的积累, 更需要在模型架构和学习范式上的质变

# 世界模拟器

一个能够模拟现实世界物理规律、因果关系与智能行为的计算机系统



虚实之间，万象归一

## 核心观察

- A: 多实验室主张视频生成模型是实现世界模拟器的可行路线
- B: 当前视频生成模型存在成本高、模态不全、长视频不稳定等问题
- C: 游戏生成模型在实时模拟方面取得突破性进展
- D: 游戏引擎可视为特定范围的世界模型

## 逻辑链条

1. 现状认知 ( $A \wedge B$ )
  - 技术路线已明确
  - 现实困难：成本高、缺模态、不稳定
2. 游戏生成带来希望 ( $C \rightarrow D$ )
  - DIAMOND：可交互游戏画面预测
  - Oasis 开放世界实时模拟
3. 演进路径 ( $B \wedge C \Rightarrow$  未来发展)  
游戏模拟  $\rightarrow$  受限世界模拟  $\rightarrow$  通用世界模拟

## 本质洞见

1. 世界模拟器的实现路径正在从"完整模拟"转向"分级模拟"
2. 游戏生成模型作为受限环境的世界模拟，为解决视频生成核心问题提供了新思路
3. 实现真正的世界模拟器需要解决的根本挑战是复杂度控制与数据获取



U-Net UV展开 多视角生成 生成速度 扩散模型

开发工具 几何形态还原 DiT架构  
元宇宙

3D训练数据 光照效果 3D UGC  
数字孪生

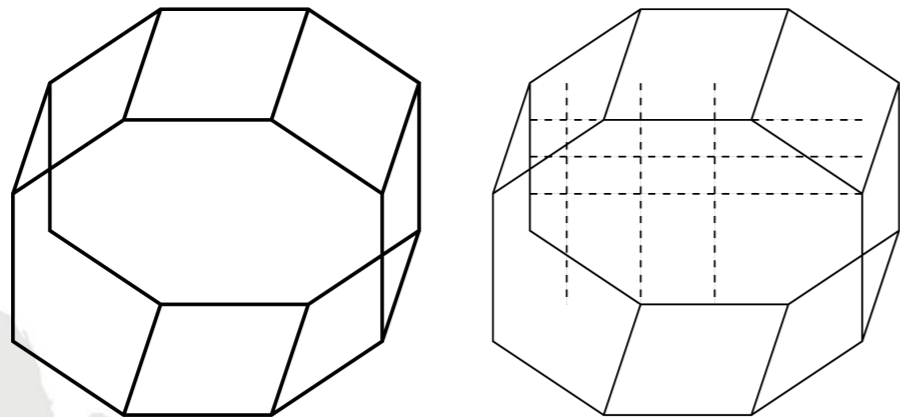
深度估计 3D生成 纹理质量  
游戏场景 几何精度  
实时渲染  
虚拟世界

角色生成 高斯泼溅 物理碰撞 AI元宇宙  
Roblox

视角一致性 PBR渲染  
场景重建 材质还原 神经场

# 几何形态还原

从复杂物体中提取基础几何特征，  
重建物体的本质形态结构与空间关系。



化繁为简，归元返真

## 核心观察

- A: 各类技术方案快速涌现
- B: 性能与效率大幅提升
- C: 应用场景不断拓展
- D: 技术挑战仍待突破

## 逻辑链条

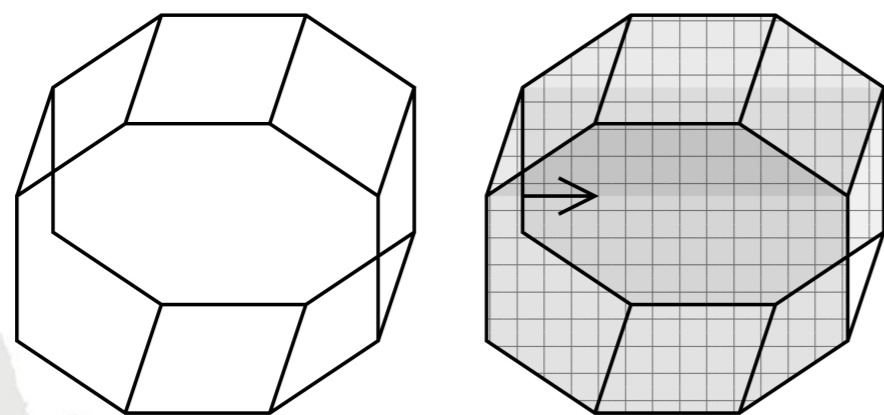
1. A  $\rightarrow$  主流技术方案
  - Tripo 2.0: DiT+U-Net架构
  - GRM: 四视图transformer重建
  - Unique3D: 多视图及法线扩散
2. B  $\wedge$  C  $\rightarrow$  应用价值
  - $\forall$ (形态还原)  $\rightarrow$   $\exists$ (速度  $\wedge$  质量  $\wedge$  效率) 广泛应用:
    - 游戏开发: 场景建模 - 影视制作: 特效场景
    - 工业设计: 产品迭代 - VR构建: 虚拟环境
3. D  $\rightarrow$  发展挑战
  - 数据积累  $\wedge$  模型优化  $\rightarrow$  产业赋能
  - 当前瓶颈: - 训练数据稀缺 - 泛化能力有限
  - 未来方向: - GANs数据生成 - 迁移学习优化

## 本质洞见

1. 图片、视频的生成均为帧的叠加，3D生成难度加大，须解决空间几何难题
2. 几何形态还原技术通过持续创新，推动3D生产效率与应用范围双提升

# 材质还原

基于几何模型的空间结构特征，  
生成真实感材质与纹理映射。



质感重构，真实再现

## 核心观察

- A: 纹理生成方法多元化
- B: 核心技术不断突破
- C: AI赋能加速发展
- D: 应用场景持续扩展

## 逻辑链条

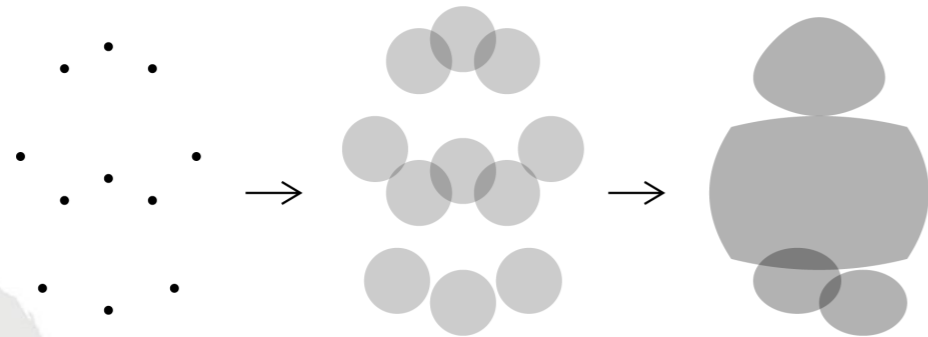
1. A → 基础方法体系 生成路径:
  - 基于图像: 纹理映射/无缝拼接
  - 基于模型: 细节增强/PBR渲染
  - 基于深度学习: GAN/CNN架构
2. B → 关键技术 技术要素:
  - UV展开与优化 - 程序化生成算法
  - PBR材质系统 - 纹理合成修复
3. C ∧ D → 发展趋势  
算法突破 → AI赋能 → 应用拓展  
AI应用: - 自动生成/转换 - 风格迁移/增强  
 $\forall$ (纹理生成) →  $\exists$ (自动化 ∧ 真实感)

## 本质洞见

1. 材质还原跟图片生成存在不同，比如存在光影问题，需要针对性解决
2. 材质生成通过多元技术融合，推动3D内容制作提质增效

# 高斯泼溅

采用空间点云数据作为基础表示，  
生成柔和连续的三维表面形态。



点云弥散，形态重现

## 核心观察

- A: 高斯泼溅是一种实时3D渲染技术
- B: 使用高斯分布描述三维空间点
- C: 具备多项技术优势
- D: 应用场景广泛

## 逻辑链条

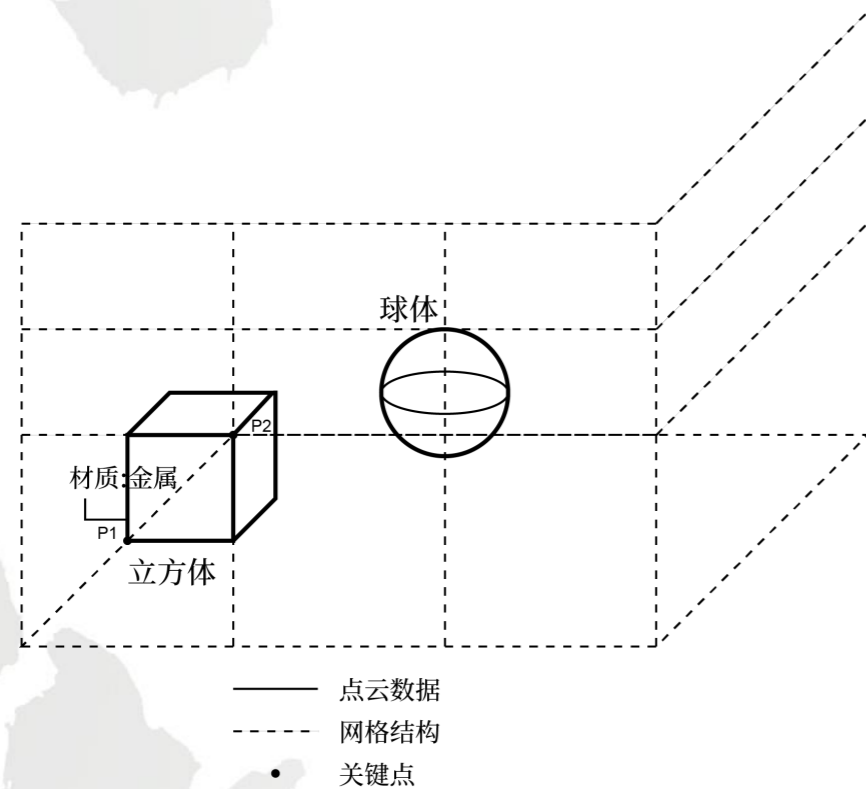
1.  $A \wedge B \rightarrow$  技术原理  
光栅化  $\wedge$  高斯函数描述  $\rightarrow$  精确场景表达  
属性: 位置  $\wedge$  协方差  $\wedge$  颜色  $\wedge$  不透明度
2.  $C \rightarrow$  性能三角:
  - 高品质渲染( $>$ NeRF)
  - 实时性能( $>100\text{fps}@1080\text{p}$ )
  - 训练效率( $<1\text{h}$ )创新价值: - 数据采集简化 - 优化机制灵活 - 实时渲染提升
3.  $D \rightarrow$  应用价值  
 $\forall$ (高斯泼溅)  $\rightarrow \exists$ (效率  $\wedge$  质量  $\wedge$  易用性)  
传统渲染  $\rightarrow$  高斯分布描述 离线处理  $\rightarrow$  实时交互  
VR/AR  $\rightarrow$  沉浸体验 游戏/动画  $\rightarrow$  视觉质量  
场景重建  $\rightarrow$  自动化

## 本质洞见

1. 高斯泼溅技术通过创新的数学模型，实现了3D渲染的效率与质量的统一
2. 在此基础上，高斯泼溅适合用于大规模的重建任务，积累3D数字资产

# 3D训练数据

用于训练空间感知与理解模型的三维数据集，  
包含几何、材质与场景语义信息。



**格物穷理，数据生境**

## 核心观察

- A: 3D数据稀缺形成系统性瓶颈
- B: 获取成本与应用规模互相制约
- C: 技术创新寻求突破性解法
- D: 行业需求驱动发展方向

## 逻辑链条

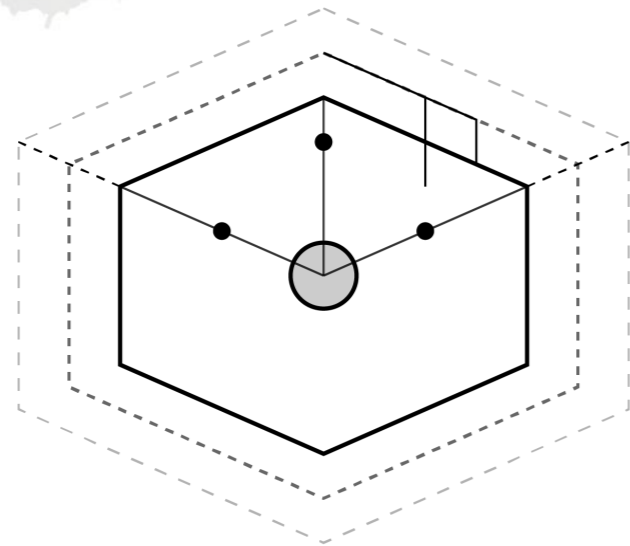
1.  $A \wedge B \rightarrow$  困境闭环，3D训练数据的稀缺性影响：
  - 高成本限制应用 - 小规模制约投入 - 低投入加剧稀缺
2.  $C \rightarrow$  破局路径，尝试中的技术演进：
  - 合成数据(Bootstrap3D) - 多模态融合(ULIP)
  - 领域适应(Swin3D++) - 单图生成(VFusion3D)构建正向循环：  
技术创新  $\rightarrow$  成本下降  $\rightarrow$  规模扩大  $\rightarrow$  持续优化
3.  $D \rightarrow$  应用牵引，明确的场景需求：
  - 自动驾驶(感知安全)
  - 机器人(精准控制)
  - 建筑工程(数字孪生)

## 本质洞见

1. 3D数据生态需要打破成本-规模困局，通过技术创新和场景落地形成良性循环

# AI元宇宙

由人工智能驱动虚拟世界生态系统，  
通过3D生成技术构建无限可能的数字空间。



—— 已生成空间  
----- 待生成区域  
● 智能实体

虚实相生，智境无垠

## 核心观察

- A: AI改变内容生产范式
- B: 交互模式发生质变
- C: 基础设施智能化升级
- D: 价值体系重构

## 逻辑链条

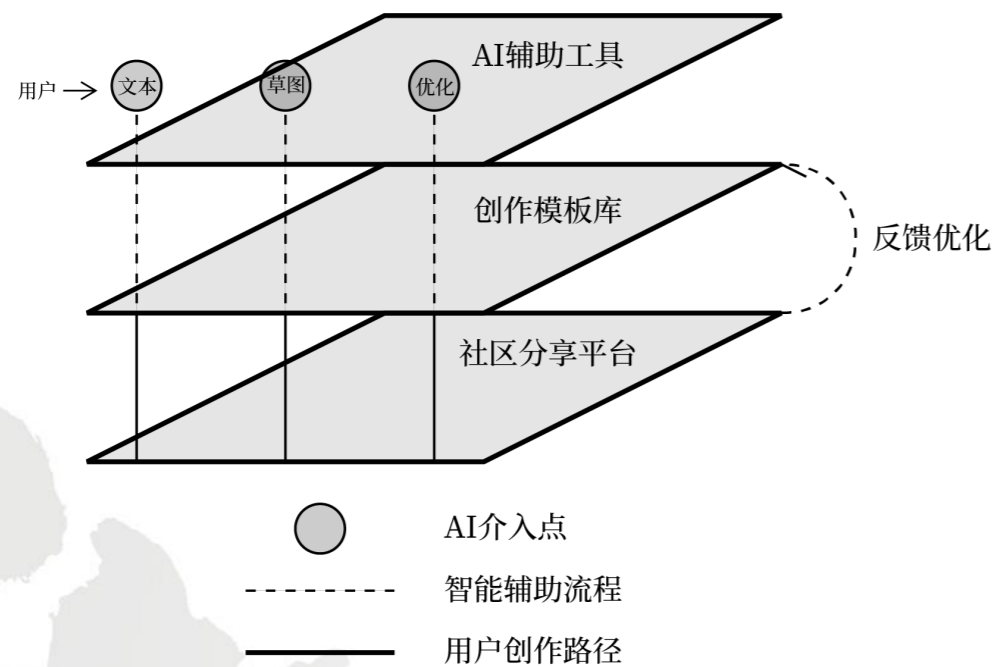
1. A → 内容生态重构  
生产模式转变: 人工创作 → AI生成  
供给特征: 内容稀缺 → 注意力稀缺  
创作门槛: 专业技能 → 提示工程
2. B → 交互深度演进  
功能 → 情感 → 价值, 每一维度都需要特定AI能力支撑  
一阶交互: 功能响应 ⇒ 二阶交互: 情感链接 ⇒ 三阶交互: 价值共创  
-场景生成-角色互动    -个性适配-情绪共鸣    -协同创作-价值交换
3. C ∧ D → 系统性升级  
基础设施: - 智能优化(资源分配) - 自适应调节(负载均衡)  
价值重构: 内容价值 → 交互价值 → 网络价值

## 本质洞见

1.  $\forall$ (AI元宇宙发展) →  $\exists$ (规模化 ∧ 个性化), 原有的规模化与个性化成本过高
2. 通过重构生产范式、深化交互维度和重塑价值体系, 形成全新的数字世界形态

# 3D UGC

借助AI技术赋能的3D内容创作生态，  
实现从专业产出到大众创作的范式转变。



## 降维创作，众智共建

## 核心观察

- A: 用户角色边界重构
- B: 创作门槛显著降低
- C: 创作生态持续演进
- D: 平台化趋势凸显

## 逻辑链条

- A → 角色边界模糊化  
用户分层演进  
- AI爱好者(技术探索) - 设计师(效率提升) - 普通用户(自我表达)  
结果: 创作者 ⇌ 消费者 边界模糊  
⇌ 消费者 → 创作者 被动接受 → 主动参与  
⇌ 案例: Minecraft/Roblox生态
- B → 生产效率提升  
传统工作流 → AI辅助流程  
应用领域: - 游戏设计 - 工业设计 - 3D打印
- C ∧ D → 未来发展方向  
工具简化 ∧ 社区驱动 → ∃(创作生态系统)

## 本质洞见

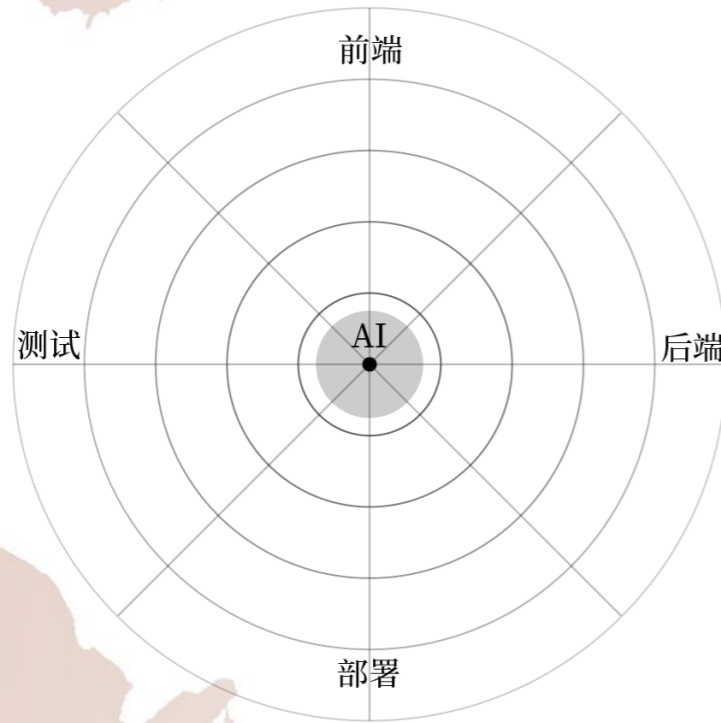
- AI提供正循环动力: 工具降维(技术门槛) → 场景扩张(应用范围) → 生态融合(平台协同)
- 3D UGC正从专业工具走向社交平台, 重构创作生态与价值体系

语义理解 画布工坊 代码质量 智能补全  
协作开发 零代码开发 超级补全 代码审查  
全栈生成 性能优化 快速迭代 多语言支持 自动部署 问题诊断  
项目管理 编程助手 代码优化  
自动重构 测试生成  
云端IDE 多模型集成 错误检测  
实时编辑 云端沙盒 工具链集成 动态UI  
提示工程 代码生成 推理debug  
版本控制



# 全栈生成

AI系统自主完成从需求理解到部署运维的完整软件开发生命周期



智构全链，万物生成

## 核心观察

- A: AI编程工具从代码补全起步
- B: 工具功能扩展至全栈代码生成
- C: 引入跨文件上下文理解能力
- D: 开发者仍需把控代码质量

## 逻辑链条

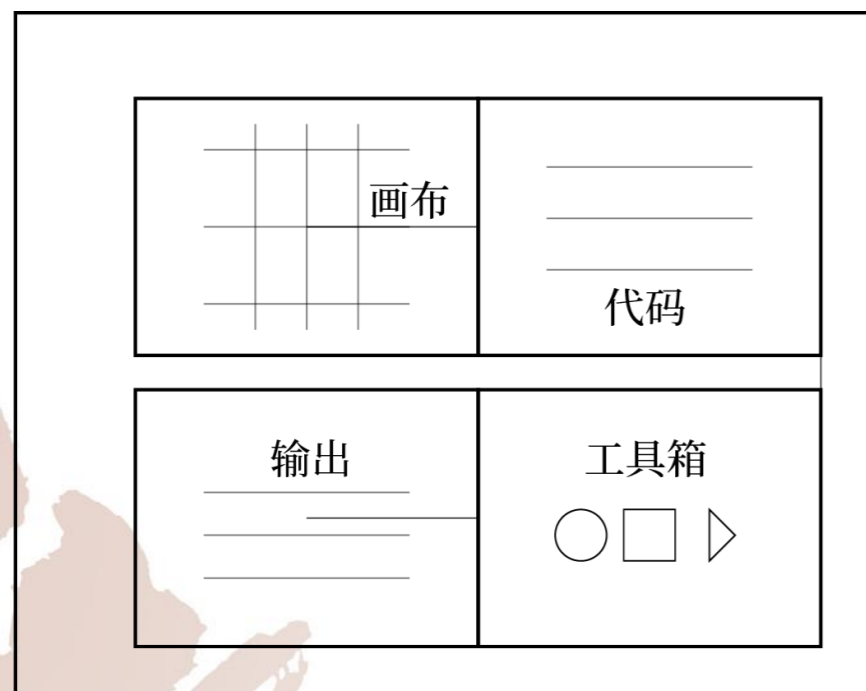
1. 功能演进链 (A → B)  
代码补全 → 代码片段生成 → 完整函数生成 → 全栈应用生成
2. 能力扩展链 (B ∧ C)
  - 前后端代码理解
  - 跨文件上下文关联
  - 工程架构把握
  - 自然语言转换
3. 应用深化链 (B → D)  
全栈生成促进：开发效率提升、编程门槛降低  
但仍需要：人工质量监督、规范性把控

## 本质洞见

1. AI编程工具正在从"局部辅助"向"整体构建"转变
2. 技术演进呈现"深度+广度"双向发展趋势
3. 人机协作模式正在从"替代性工具"转向"赋能性伙伴"

# 画布工坊

交互式AI编程环境的创作空间  
集代码执行、可视呈现与协同创作于一体



方圆之间，演绎造化

## 核心观察

- A: 主流AI平台纷纷推出画布/工坊功能
- B: 画布工坊实现代码执行和协作功能整合
- C: 开发工具正向智能化协作方向演进
- D: 新一代开发环境强调实时交互与协同

## 逻辑链条

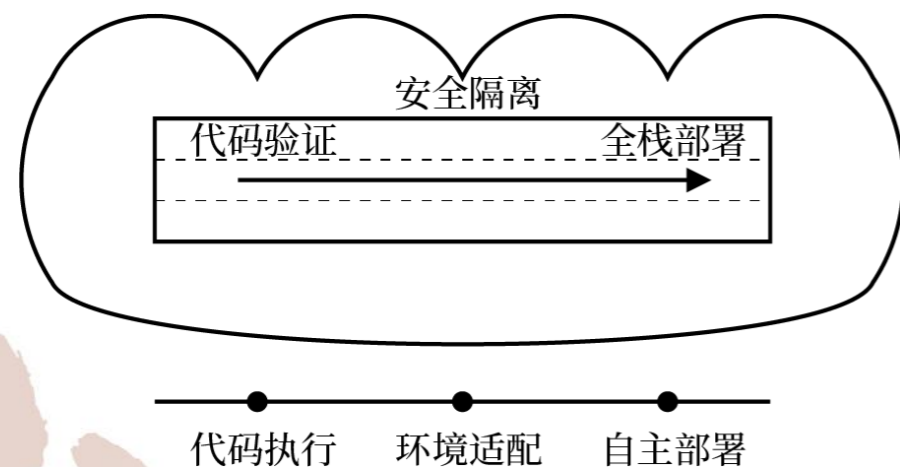
1. 功能整合趋势 (A  $\wedge$  B)
  - ChatGPT Canvas集成写作和编程
  - Claude Artifacts提供专用工作空间
2. 技术演进路径 (B  $\rightarrow$  C)  
传统IDE  $\rightarrow$  AI辅助编程  $\rightarrow$  智能协作环境  
特征：从单一功能向多维协作转变
3. 应用模式升级 (C  $\rightarrow$  D)
  - 实时代码建议与反馈
  - 多模态内容处理能力
  - 高效协作、快速迭代

## 本质洞见

1. "画布工坊"代表了AI辅助开发从工具向平台的范式转变
2. 未来开发环境将以"人机协同+团队协作"为核心特征
3. AI编程工具正在重塑传统开发流程，形成新的生产力模式

# 云端沙盒

AI代码的安全隔离执行环境  
从代码验证到全栈部署的进化容器



安全孕育，赋能进化

## 核心观察

- A: 云端沙盒为AI代码生成提供隔离执行环境
- B: AI Agent 需要安全可控的实验场地
- C: 沙盒正在向自主智能体运行时平台演进

## 逻辑链条

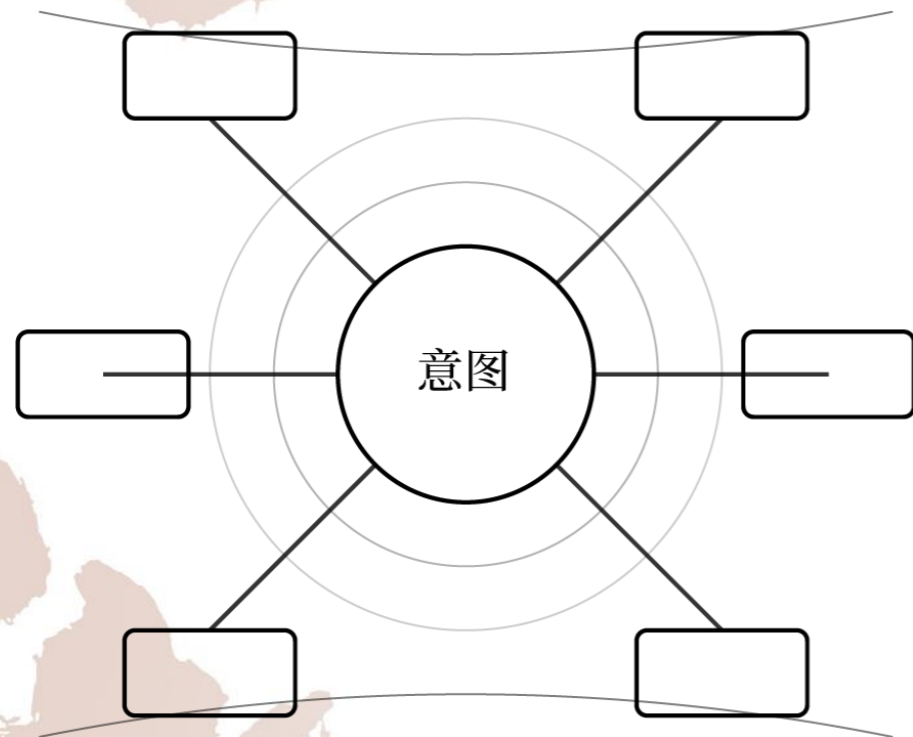
1. 安全基础设施 (A)  
代码隔离执行 + 资源使用限制 + 实时行为监控
2. Agent能力构建 (A → B)  
代码生成验证 + 环境状态模拟 + 交互行为测试 + 错误优雅回滚
3. 运行时平台进化 (B → C)
  - 单次执行 → 持续运行
  - 被动验证 → 主动学习
  - 独立沙盒 → 多智能体协作
  - 静态环境 → 动态适应

## 本质洞见

1. 云端沙盒正从"代码验证工具"演进为"AI智能体孵化器"
2. "安全边界 + 能力构建 + 自主进化"构成了AI系统迭代优化的闭环
3. 沙盒平台化趋势体现了从"受控执行"到"自主协作"的范式转变

# 动态UI

基于用户意图实时生成和调整的  
智能化界面呈现系统



界面如水，随心而动

## 核心观察

- A: 传统UI是静态预设的界面结构
- B: AI代码生成能力显著提升
- C: UI正在向动态适配方向发展
- D: LLM在UI生成中扮演核心角色

## 逻辑链条

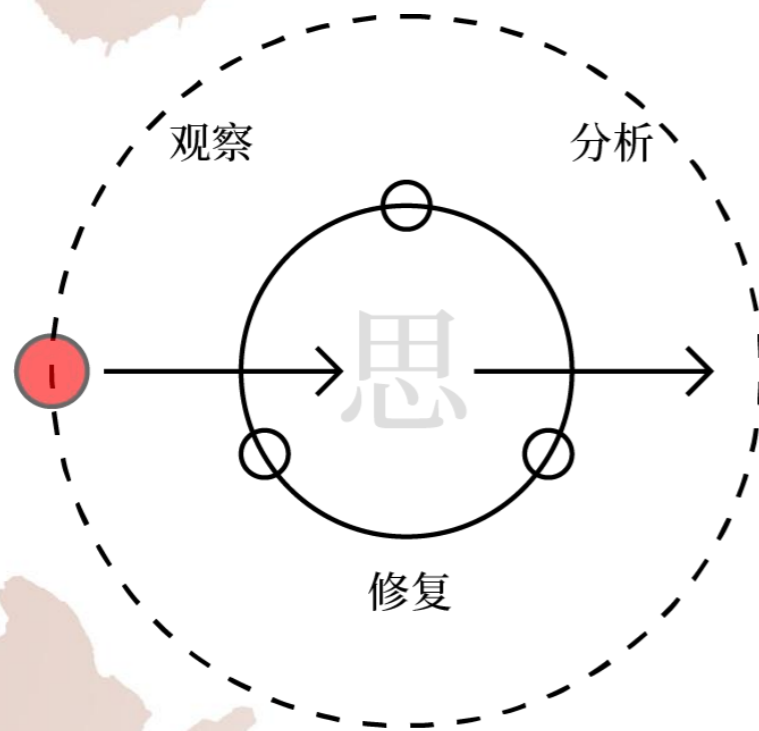
1. 技术基础演进 ( $A \wedge B \rightarrow C$ )
  - 静态UI限制  $\rightarrow$  需求痛点浮现
  - AI代码生成提升  $\rightarrow$  技术可能性出现
  - 两者交互  $\rightarrow$  促进动态UI发展
2. 实现路径展开 ( $C \wedge D$ )
  - 生成式设计 + 自适应界面  $\rightarrow$  个性化呈现
3. 发展趋势形成 ( $B \rightarrow D \rightarrow C$ )
  - 设计开发自动化
  - 交互方式自然化

## 本质洞见

1. UI正从"预设式"向"生成式"范式转变
2. 界面交互正从"人适应机器"向"机器适应人"演进
3. UI设计正从"静态产品"向"动态服务"转型

# 推理Debug

通过深度推理能力对程序进行诊断和修复，  
实现智能化的程序调试与优化。



循证推理，源流自明

## 核心观察

- A: 强化学习增强了LLM的推理能力
- B: 通过长时间推理可解决困难问题
- C: 推理能力可应用于编程Debug场景

## 逻辑链条

1. 基础能力构建 (A)  
思维链 + 强化学习 → 推理能力提升
2. 能力延展 (A → B)  
增强推理 → 复杂问题分解、长时深度思考、逐步验证优化
3. Debug场景应用 (B → C)
  - 代码逻辑分析
  - 错误原因追溯
  - 解决方案生成

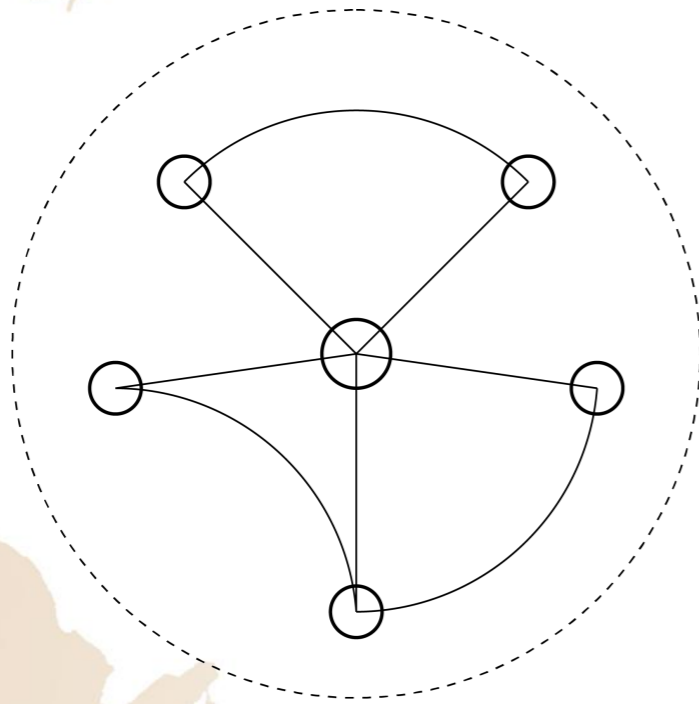
## 本质洞见

1. AI推理正从"快速响应"向"深度思考"模式转变
2. Debug过程正在实现从"症状处理"到"根因分析"的进化
3. 推理型AI正在重构传统编程的问题解决模式

知识积累 社会模拟 工作流程 自主执行  
层次结构 模型协同 评估框架  
任务规划 智能体网格 自我进化  
自主决策 科研助手 泛化能力  
终身学习 Agent 金融支付  
虚拟医疗  
基准评估 强化学习 智能体应用  
教育辅导  
环境适应 执行效率 智能体协作架构  
安全控制  
交互系统 上下文理解  
内外循环 多代理协作 长期记忆

# 社会模拟

在虚拟空间中构建群体行为与社会关系的映射，  
观察个体互动中涌现的集体智慧与复杂模式。



万象生灵，群智涌现

## 核心观察

- A: AI社会模拟系统由生成式代理、记忆流和反思模块构成
- B: Agent模型正从通用型向个性化方向发展
- C: 系统开发过程中需要考虑准确性、偏见和伦理问题
- D: "Agent Banks"正在成为社会科学研究的新工具

## 逻辑链条

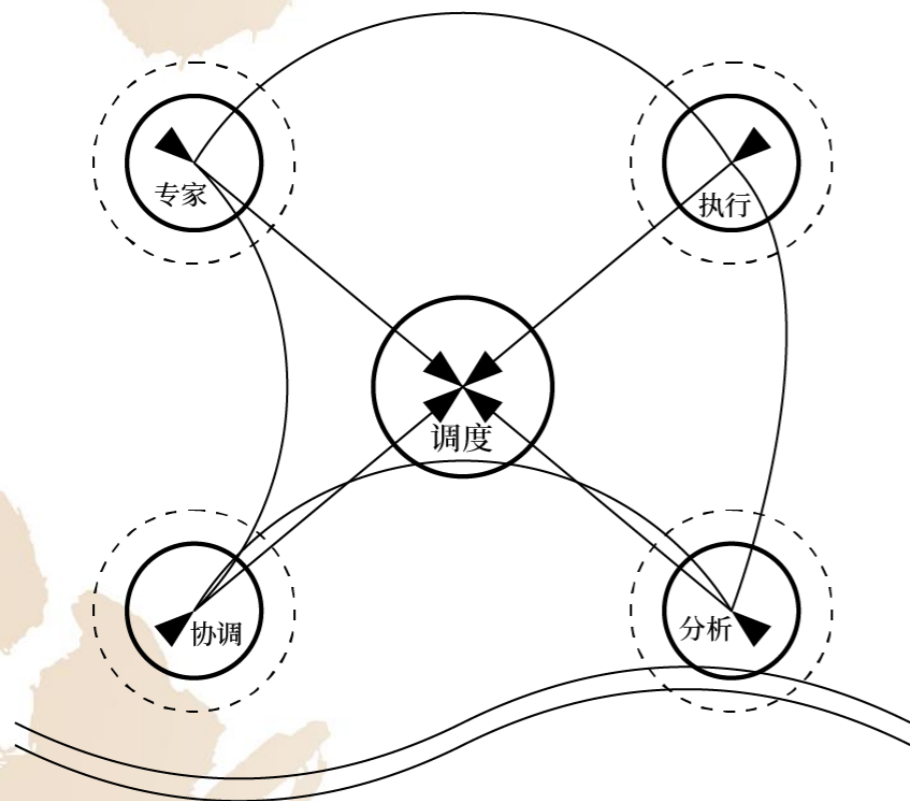
1. 技术基础构建 (A)  
生成式代理  $\wedge$  记忆流  $\wedge$  反思模块  $\rightarrow$  基础架构形成
2. 发展方向演进 (A  $\rightarrow$  B)
  - 通用模型  $\rightarrow$  个性化代理
  - 真实数据  $\rightarrow$  代理行为塑造
  - 简单交互  $\rightarrow$  复杂社会模拟
3. 约束与平衡 (B  $\rightarrow$  C)
  - 准确性要求  $\Rightarrow$  行为仿真优化
  - 偏见问题  $\Rightarrow$  公平性机制
  - 隐私保护  $\Rightarrow$  伦理框架建立
4. 应用价值实现 (C  $\rightarrow$  D)  
Agent Banks构建  $\rightarrow$  社会科学研究工具

## 本质洞见

1. AI社会模拟正在从"技术可行性"向"社会适用性"转变
2. 系统发展面临技术进步与伦理约束的双重挑战

# 智能体协作架构

多智能体通过角色分工、信息共享和任务协同，构建具有涌现能力的协作系统。



分工有序，协同涌智

## 核心观察

- A: 多智能体系统正从对称协作转向非对称协作模式
- B: 专门的编排工具正成为管理复杂协作的关键
- C: 通信和数据交换的健壮性日益重要
- D: 系统的可扩展性和稳定性成为核心关注点

## 逻辑链条

1. 协作模式演进 (A)
  - 对称协作 → 非对称协作 (允许对立)
  - 辩论者-评判者系统
  - 加权投票机制
2. 技术支撑体系 (B ∧ C)
  - 编排工具提供:
  - workflow 定义与管理
  - 实时监控与调试
  - 可视化调试环境
  - API集成能力
3. 系统优化方向 (C → D)
  - 关注重点:
  - 标准化通信协议
  - 容错与错误处理

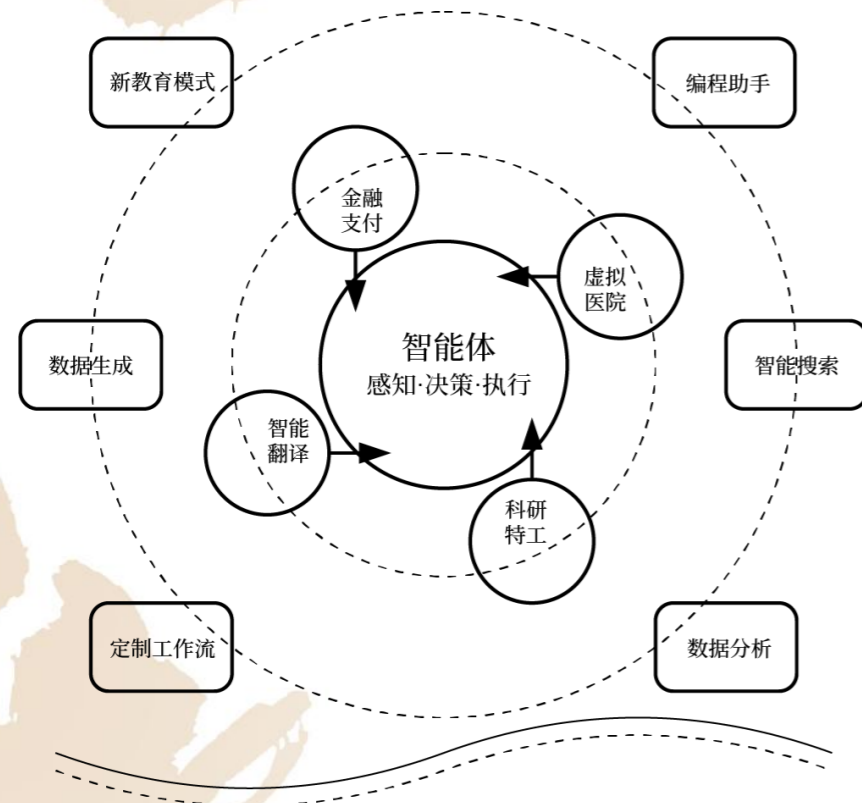
## 本质洞见

1. 多智能体系统正在从"平等协作"向"专业分工"转变
2. 系统复杂度的提升推动了配套工具的专业化发展
3. 未来的多智能体系统将更注重适应性和可靠性，而非单纯的功能扩展



# 智能体应用

基于智能体技术构建的实际应用场景，  
通过感知、决策、执行的闭环能力服务现实任务。



智融万象，悉听人命

## 核心观察

- A: Agent技术正在全面进入商业化应用阶段
- B: 多样化场景需求推动Agent能力持续进化
- C: Agent应用呈现专业化、自动化、协作化特征

## 逻辑链条

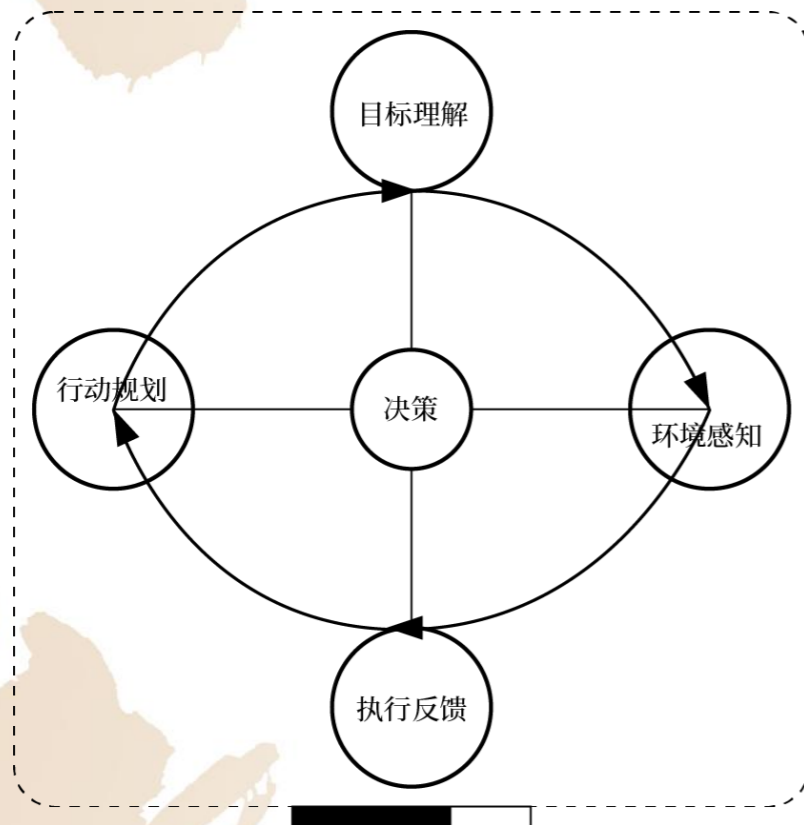
1. 技术能力进化 (A → B)
  - 基础能力: 任务执行 → 决策制定 → 自主学习
  - 协作能力: 单体运行 → 多智能体协同 → 生态协作
  - 专业能力: 通用服务 → 垂直领域 → 复杂场景
2. 应用场景扩展 (B → C)
  - 金融交易: 支付结算、风险控制
  - 医疗服务: 诊断咨询、治疗方案
  - 科研教育: 实验设计、个性化教学
  - 软件开发: 代码生成、程序测试
  - 数据分析: 建模预测、见解生成
3. 发展特征演进 (C)
  - 单一功能 → 多维协作 → 生态集成
  - 人工辅助 → 半自动化 → 全自动化
  - 通用服务 → 专业分工 → 场景定制

## 本质洞见

1. Agent技术正从"能力积累期"进入"应用爆发期"
2. 多智能体协作成为解决复杂问题的主流范式
3. 专业化分工与场景深耕是Agent发展的必经之路

# 自主执行

智能体自动感知环境并制定决策，持续执行任务并实现目标闭环。



择径自主，循迹前行

## 核心观察

- A: AI Agent正从规则驱动转向学习驱动模式
- B: 自主能力由多个核心组件协同支撑
- C: 多智能体系统(MAS)成为复杂任务解决方案

## 逻辑链条

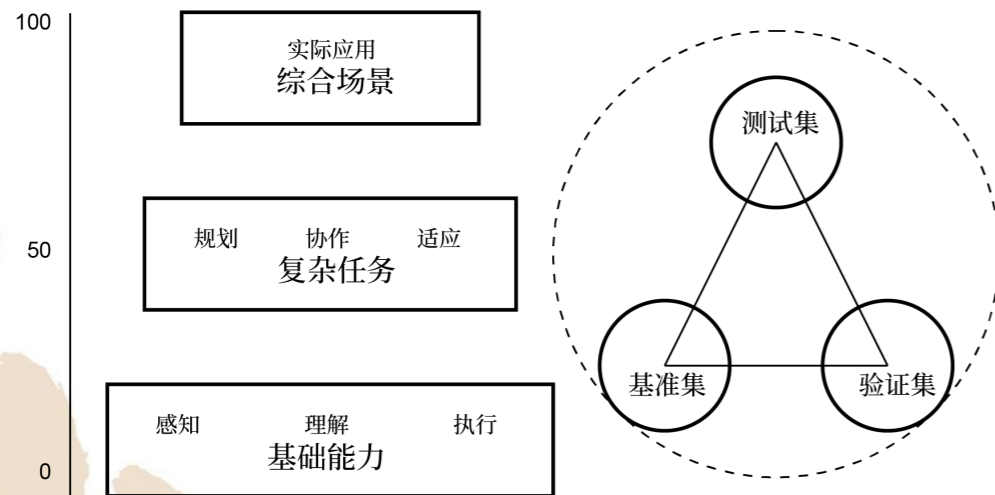
1. 技术范式转变 (A)  
规则驱动 → 学习驱动
  - 数据学习能力
  - 动态适应能力
  - 不确定性处理
2. 核心能力构建 (A ∧ B)
  - 决策中枢
  - 大规模模型
  - 推理规划
  - 记忆管理
  - 工具调用
  - 执行机制
  - 学习系统
3. 演进路径: A ∧ B → C (有机融合)

## 本质洞见

1. Agent技术正经历从"确定性"到"学习型"的根本性转变
2. 多组件协同架构是实现真正自主能力的关键

# 智能体基准评估

构建智能体能力评估的统一标准与方法，建立可度量、可对比的评价体系。



度衡有据，优劣可鉴

## 核心观察

- A: 评估指标包含多个维度(准确性、响应时间、可靠性等)
- B: 高级基准测试强调交互性和动态性( $\tau$ -bench、AgentBench等)
- C: 评估趋势关注用户满意度和商业价值

## 逻辑链条

### 1. 基础评估框架 (A)

核心指标体系:

- 准确性  $\rightarrow$  输出正确性
- 响应时间  $\rightarrow$  执行效率
- 可靠性  $\rightarrow$  一致性表现
- 完成率  $\rightarrow$  运营效能
- 错误率  $\rightarrow$  质量控制
- 成本效益  $\rightarrow$  经济可行性

### 2. 高级评估方法 (A $\rightarrow$ B)

新型基准测试:

- $\tau$ -bench: 动态对话模拟
- AgentBench: 交互环境评估
- Meta-Benchmarking: 自优化能力

### 3. 评估维度扩展 (B $\rightarrow$ C)

新增关注点:

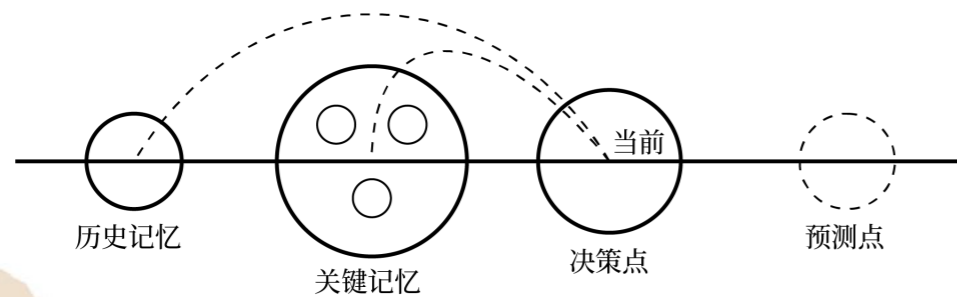
- 用户体验度量
- 商业价值评估
- 持续优化能力

## 本质洞见

1. AI智能体评估正从"结果验证"走向"过程理解"
2. 评估维度正在从单一技术指标扩展到人机协作
3. 效能自动化评估与人工评估的混合模式将成为主流

# 长期记忆

构建深层次的信息存储与提取机制，  
实现跨时空的知识关联与统筹。



承前启后，思维贯通

## 核心观察

- A: AI系统正从固定上下文窗口向多层次记忆架构演进
- B: 检索增强和外部存储成为扩展记忆的主要方案
- C: 动态总结和分层管理是处理长期记忆的关键技术
- D: 系统正在发展出类人的记忆形成与遗忘机制

## 逻辑链条

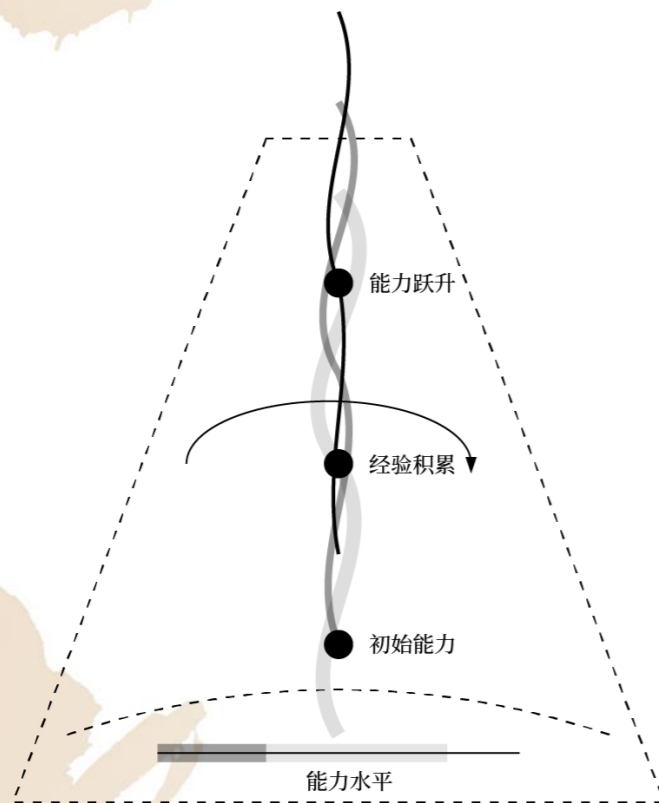
1. 架构创新 (A → B)
  - 向量数据库集成
  - 稀疏注意力机制
  - 情节式记忆表示
2. 记忆分层 (B → C)  
层级划分:
  - 短期工作记忆
  - 中期扩展记忆
  - 长期知识库储存
3. 智能管理 (C → D)  
关键技术:
  - 递归摘要生成
  - 上下文感知压缩
  - 混合推理机制

## 本质洞见

1. AI记忆系统正从"单一存储"向"分布式认知"架构演进
2. 记忆管理的重点从"信息存取"转向"智能筛选与整合"
3. 真正的AI智能体需要在"记忆"与"遗忘"之间找到平衡

# 自我进化

通过持续学习与经验积累，智能体不断优化自身能力模型，实现认知边界的动态扩展。



积厚流光，破茧成蝶

## 核心观察

- A: AgentGym提出智能体自我进化框架
- B: AgentEvol实现探索-学习双循环机制
- C: LeanAgent引入数学定理证明的课程学习
- D: 评估体系从结果导向转向过程导向

## 逻辑链条

1. 框架构建 (A)  
三要素支撑：
  - 多样化环境任务
  - 轨迹集知识储备
  - 有效进化方法论
2. 实践应用 (B → C)  
创新特征：
  - 基于难度的课程学习
  - 动态知识库管理
  - 渐进式训练平衡
3. 评估演进 (C → D)  
新范式建立：
  - 过程导向评估
  - 中间反馈机制
  - 可扩展评价体系

## 本质洞见

1. AI智能体进化正从"静态训练"转向"动态适应"
2. 持续学习能力依赖于"探索-反馈-优化"的闭环系统
3. 智能体评估标准从"能力验证"转向"进化潜力"

读屏操作 内存占用 端侧Agents  
计算效率

极限压缩 推理速度 屏幕交互 上下文优化

MiniCPM 数据安全 参数优化 端云协同

跨应用操作 混合专家 端侧智能 实时翻译  
低延迟 注意力机制

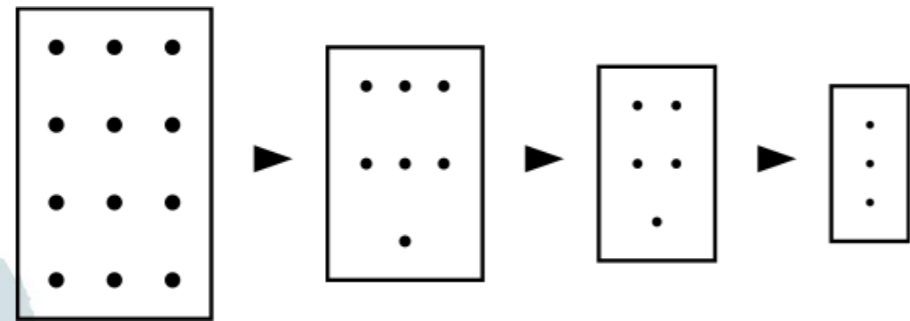
小模型 UI识别 算力优化 AI芯片 智能搜索

视觉理解 NPU加速 多语言支持 模型蒸馏

移动处理器 端侧多模态隐私计算

# 极限压缩

在保持核心能力的前提下，将深度学习模型压缩到最小规模的技术与方法。



精简得要，去芜存菁

## 核心观察

- A: 7B曾被认为是端侧模型的入门门槛
- B: Apple Intelligence实现了3B模型的端侧部署
- C: Gemma 2 2B将通用端侧模型极限推至2.6B
- D: 小参数量模型可实现与大模型相当的性能

## 逻辑链条

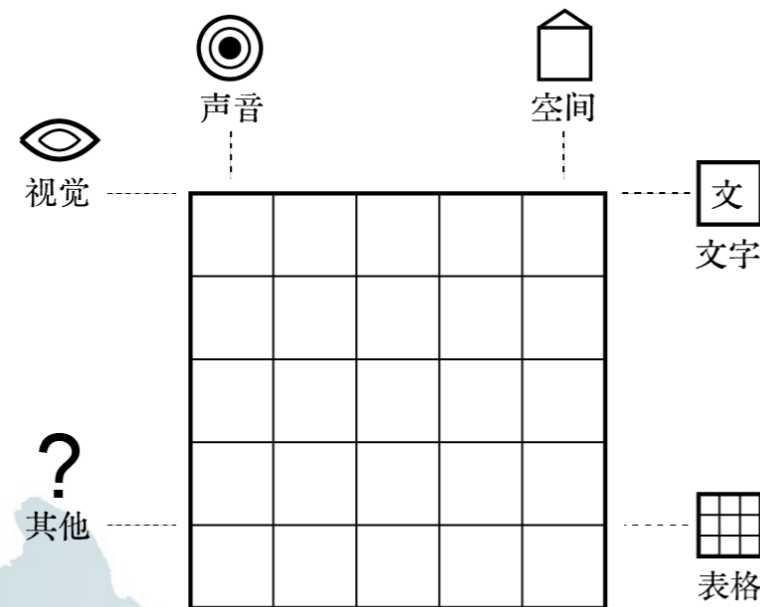
1. 初始认知 (A): 7B → 曾被视为端侧模型最小可用参数量  
↓
2. 一次突破 (A → B) Apple Intelligence通过:
  - 专项任务小模型生成
  - 适配器动态加载交换
  - 端侧模型整体量化压缩↓
3. 二次突破 (B → C) Gemma 2 2B通过:
  - 上级模型蒸馏
  - MLX Swift优化
  - ShieldGemma分类器稳定输出

## 本质洞见

1. 端侧部署的技术路线正在从"压缩大模型"转向"优化小模型"
2. 关键技术突破点:
  - 任务特化
  - 动态适配
  - 量化优化
  - 结构化提示
3. 性能与参数量的解耦趋势明显，更高效的架构设计正在取代简单的规模追求

# 端侧多模态

在终端设备上实现图像、语音、文本等多模态数据的轻量级感知与融合理解。



多源汇聚，融贯通达

## 核心观察

- A: 把7B设定为端侧模型的基准参考点
- B: 模型发展出现双轨并行趋势：极限压缩与多模态
- C: 不同终端对多模态需求差异显著

## 逻辑链条

1. A → B: 以7B为分水岭
  - 向下：追求极致压缩
  - 向平：扩展多模态（如MiniCPM-V系列）
  - 最终：两个方向叠加，实现2-3B级别多模态
- ↓
2. B + C → 差异化发展必然性：
  - 手机/PC：文本为主、小任务为主 → 倾向小型化
  - 车载/专业设备：多模态刚需 → 保持规模+图片、视频、语音交互
3. 技术价值判断：
  - 若 性能 = 效率 × 规模
  - 则 提升效率 > 单纯压缩规模
  - 则 提升效率 → 模态叠加

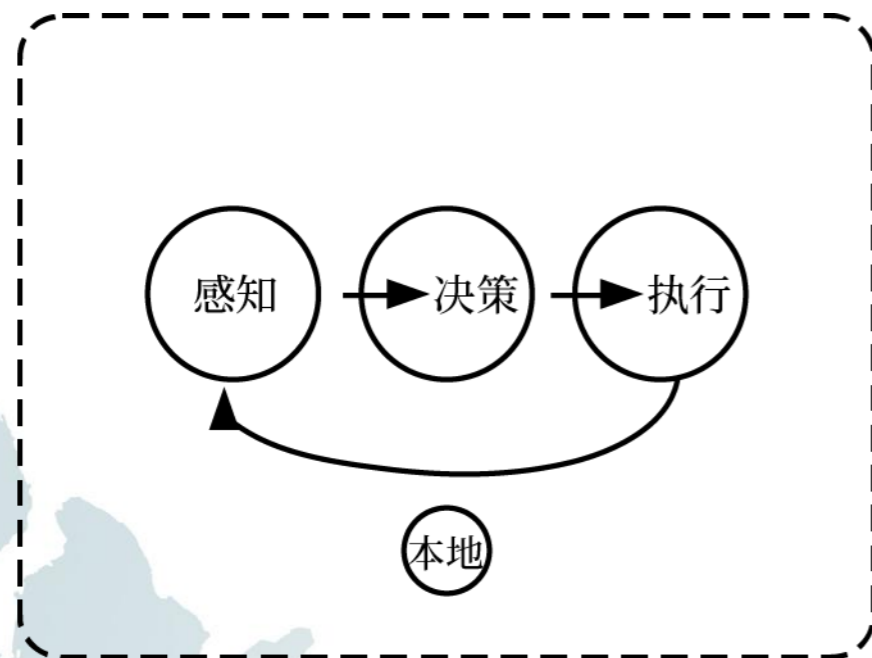
## 本质洞见

1. 端侧模型不是"越小越好"，而是"够用更好"、"好用最好"
2. 设备算力决定了最优模型规模，未来终端需求将呈多样化
3. 多模态能力是突破通用计算的关键，仍有强需求



# 端侧Agents

在终端设备上自主运行的智能代理程序，  
具备感知、决策、执行的闭环能力。



自主循环，智在边端

## 核心观察

- A: 大模型发展有两种路线：单一系统（快思考） vs 双系统结合（快+慢思考）
- B: 端侧资源约束要求最优化使用
- C: Agent 是连接模型与应用的关键

## 逻辑链条

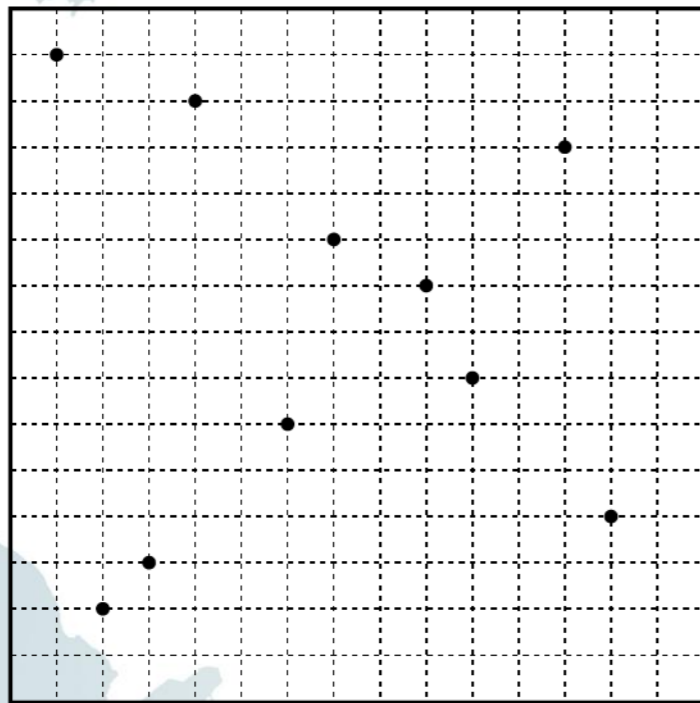
1. 为什么需要Agent？  
A + B  $\Rightarrow$  双系统路线更适合端侧
  - 资源有限性
  - 任务复杂性
  - 生态多样性
2. Agent如何创造价值？  
B + C  $\Rightarrow$  双重价值实现
  - 资源优化：任务分解、按需调用
  - 生态连接：跨应用协作、UI理解
3. 发展趋势：
  - 技术：从单一模型到多智能体协作
  - 生态：从封闭应用到开放服务
  - 交互：从指令执行到场景理解

## 本质洞见

1. Agent不是锦上添花，而是端侧AI的必需品
2. Agent的作用远不止于目前，未来，平台竞争将围绕Agent能力展开

# AI芯片

高效执行人工智能算法的专用集成电路，  
通过并行计算单元阵列实现模型加速。



算力凝芯，效能焕发

## 核心观察

- A: 传统GPU架构面临存储墙瓶颈
- B: 存算一体架构实现突破性提升
- C: 超快推理或将重构AI应用形态

## 逻辑链条

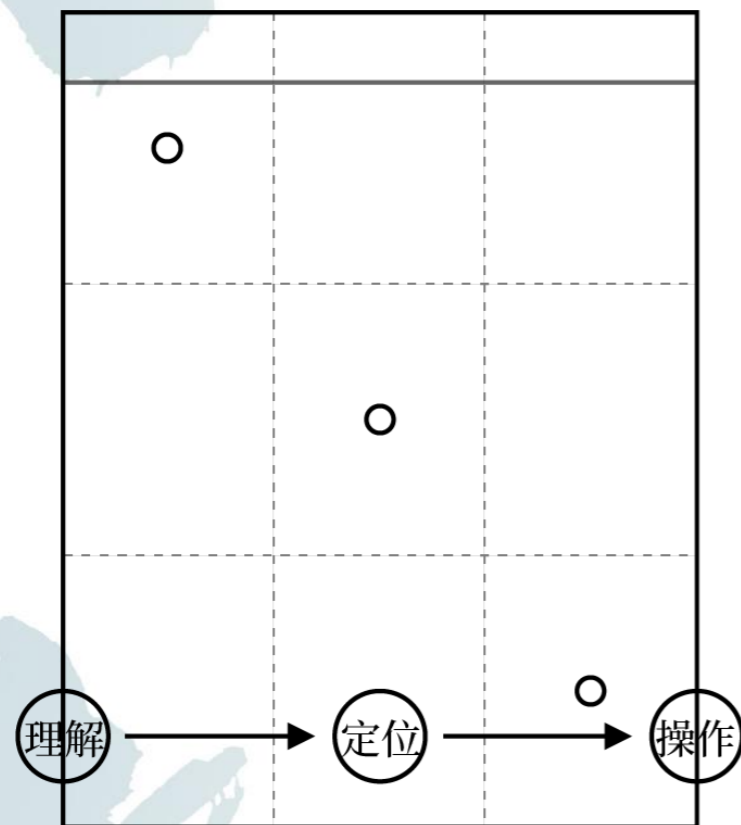
1. A → 传统架构局限：
  - 存算分离制约性能
  - 数据搬运消耗巨大
  - 并行计算效率受限
2. B → 架构创新：
  - Groq: LPU一维处理器阵列
  - Cerebras: 晶圆级存算集成
  - 英伟达: HBM近存优化
3. A + B → C: 应用变革
  - 思考更敏捷: - 深度推理成本降低 - 多方案实时对比
  - 交互更自然: - 实时语音对话 - 动态应用生成

## 本质洞见

1. 存算分离到存算一体是计算架构的范式转换
2. 推理速度的量级提升，将重构人机交互模式
3. 市场成熟度与技术突破之间存在时间差，需要通过应用场景培育推动发展

# 读屏操作

通过智能视觉分析解构界面结构与交互元素，实现对屏幕内容的理解与自动操作。



解构识别，如臂使指

## 核心观察

- A: 传统读屏停留在机械识别层面
- B: Ferret-UI等实现了智能理解突破
- C: 读屏或将成为端侧智能的基础设施
- D: 读屏或带来数据安全与隐私等新问题

## 逻辑链条

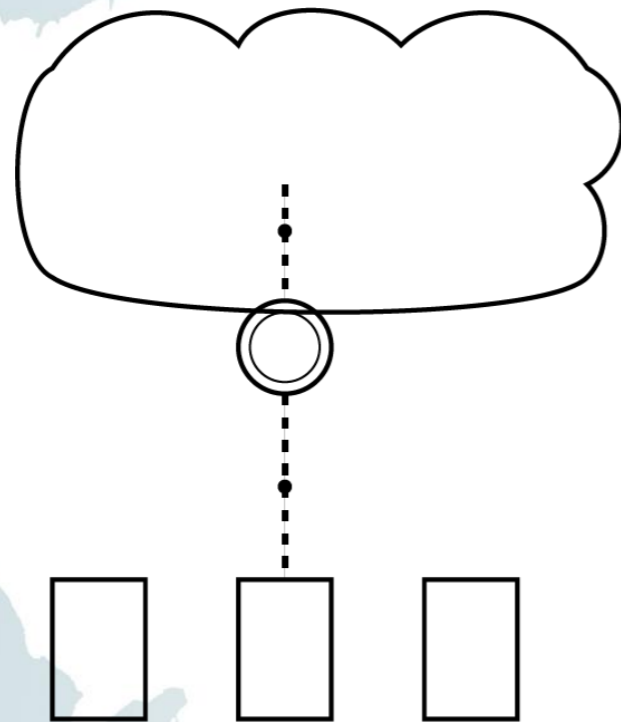
1. A → B: 能力跃迁
  - 从简单识别到智能理解
  - 从固定流程到动态适配
  - 从单一操作到任务推理
2. B → C: 基础设施化
  - 视觉理解: 屏幕内容完整解析
  - 意图理解: 自然语言指令转换
  - 行为执行: 跨应用任务编排
3. C+D: 技术实现路径:
  - 底层突破:     ⇒     能力构建:     ⇒     生态价值:
  - 多分辨率处理     • 基础识别定位     • 应用协作
  - 细节增强采样     • 高级推理交互     • 智能交互

## 本质洞见

1. UI理解是智能交互的核心入口
2. 读屏技术正从工具走向平台，但须解决数据安全问题
3. 进一步或将重塑整个端侧交互范式和应用生态

# 端云协同

端与云的计算资源动态调配机制，  
实现智能任务的最优分发与协作处理。



智能调度，云端共生

## 核心观察

- A: 云端模型始终领先端侧一个量级
- B: 用户数据实际大量存储于云端
- C: 端云协同或不是选择，而是必然

## 逻辑链条

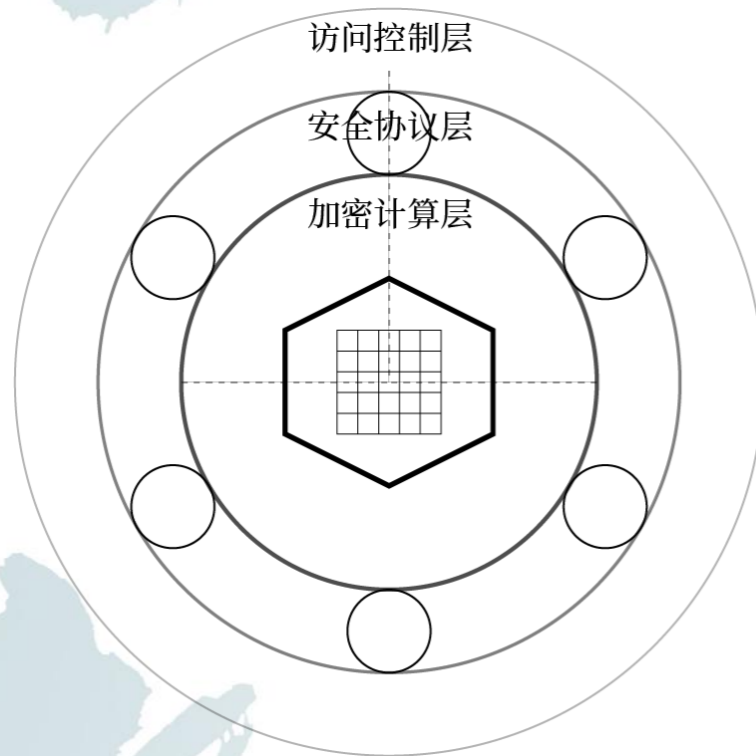
1. 能力差异的演化：
  - A → Q1: 两种条路线：
    - 云端：追求极限能力
    - 端侧：追求效率平衡
  - Q1 → R1: 差异将持续存在且加大，不存在端侧"赶超"的可能性
2. 数据分布的现实
  - B → Q2: "纯端侧"假设已被打破；
  - Q2 → R2: 数据天然呈现混合分布：
    - 敏感数据：端侧存储
    - 通用数据：云端存储
    - 交互数据：实时流动
3. 协同的深层必然性
  - (R1 + R2) → Q3: 单一架构难以满足需求
  - Q3 → R3: 协同不是技术选择，而是架构必然
    - 能力协同：优势互补
    - 数据协同：动态流转
    - 成本协同：资源优化

## 本质洞见

1. 端云不是简单分工，而是能力的有机融合，协同效应超越单一能力的叠加
2. 端云的边界正在消失，未来将形成动态流转的统一计算范式与AI基础设施

# 隐私计算

在数据加密状态下进行分布式协同计算，  
确保数据应用与隐私保护的动态平衡。



密而不绝，算而不泄

## 核心观察

- A: 传统隐私保护依赖物理隔离
- B: 新型隐私方案突破物理限制
- C: 多元化解决方案正共存发展

## 逻辑链条

逻辑链条：

1. 认知转变：A → Q1: 物理隔离的局限性
  - 成本高昂
  - 效率受限
  - 体验割裂Q1 → R1: 需要新型隐私保护范式
2. 技术突破：B → Q2: 突破传统边界约束：  
Q2 → R2: 形成新的保护模式
  - 算力：云端托管
  - 数据：加密流转
  - 控制：用户主权
3. 生态演化  
(R1 + R2) → Q3: 不同场景要求不同；  
Q3 → R3: 多元化是必然选择
  - 封闭生态：极致控制，如苹果模式：垂直整合
  - 开放生态：灵活平衡，Cohere模式：水平协作
  - 混合模式：场景适配

## 本质洞见

1. 范式转换：空间隔离 → 逻辑隔离、静态保护 → 动态保护、被动防御 → 主动赋权
2. 隐私计算将从简单的“隔离保护”走向“智能协同”，重塑技术架构和服务模式

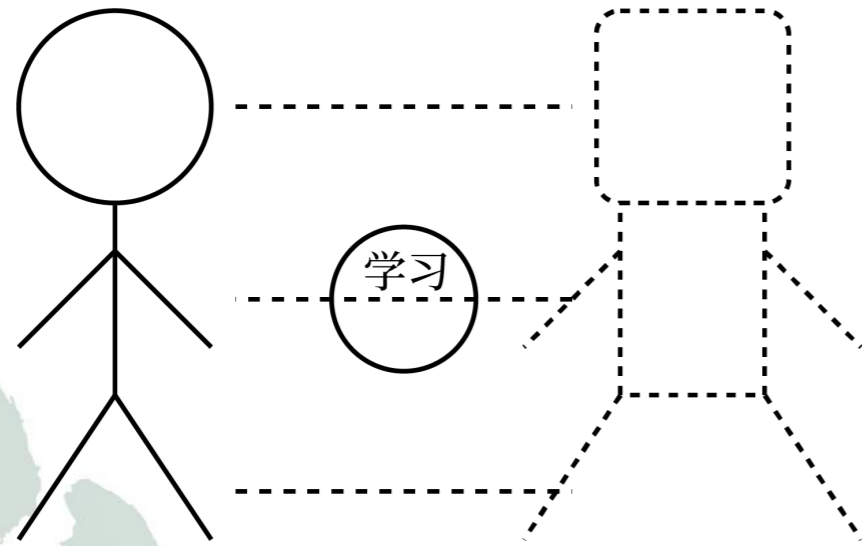
服务机器人 人形机器人 医疗机器人  
运动规划 医疗手术 触觉皮肤 深度感知  
双足机器人 Sim2Real 轮足机器人  
模仿学习 传感系统  
力控制 自主学习  
动作精度 环境适应  
强化学习 空间智能 共创平台 续航能力  
执行器 任务规划 老年陪护 仿真训练 工业制造  
灵巧手 运动控制 机器人产业链  
安全交互

# 具身智能

空间智能 共创平台  
运动控制 机器人产业链

# 人形机器人

融合形态结构与认知交互的智能体，  
实现类人化的感知、决策与动作能力。



形神兼备，智行合一

## 核心观察

- A: 大模型不等于具身智能
- B: 传统液压技术存在工程局限
- C: 电动化成为新技术路线
- D: 硬件基础设施仍需突破

## 逻辑链条

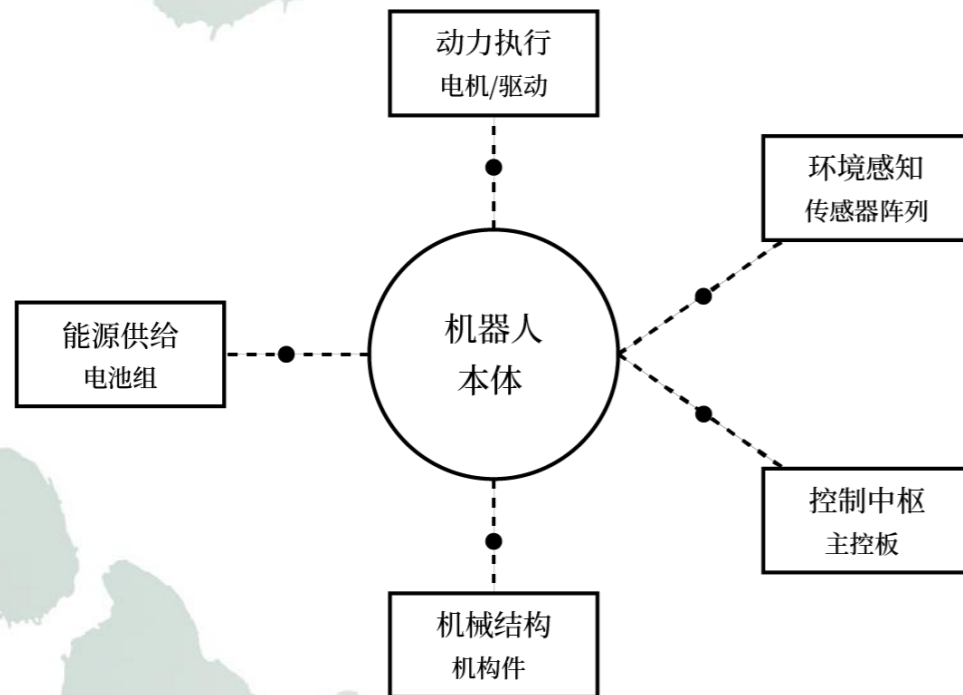
1.  $\neg(A) \rightarrow$  具身智能具有复杂性  
硬件限制 > 算法限制    工程实现 > 理论突破  
AI能力  $\neq$  物理实现能力  $\Rightarrow$  演示性能  $\rightarrow$  实用性能
2. B  $\rightarrow$  C (技术路线转变) 案例分析:
  - Atlas: 液压  $\rightarrow$  电动
  - Optimus: 全电动架构
  - Figure 02: 电动+AI融合
3.  $\forall$ (技术突破)  $\rightarrow$  多维度进展:
  - 本体控制    - 灵巧手    - 触觉传感    - 表情模仿

## 本质洞见

1. 仍须提升: 基础设施  $\wedge$  控制算法  $\wedge$  应用场景  $\rightarrow$  综合能力提升
2. 人形机器人需要在"身体"能力上实现突破, 才能真正释放"大脑"的潜力

# 机器人供应链

机器人核心部件（动力、传感等）及标准化模块的供应体系与产业链条。



硬核汇聚，零整相成

## 核心观察

- A: 供应链成熟度制约产业发展
- B: 成本远超预期目标
- C: 中国供应链展现竞争优势
- D: 国内企业推动多样化创新

## 逻辑链条

1.  $A \wedge B \rightarrow$  产业困境
  - Atlas(液压): \$2M - Optimus(电动): \$60K vs 目标\$20K
  - 原因: 零部件定制化需求
2.  $C \rightarrow$  突破路径 案例佐证:
  - 宇树G1: ¥99K - 众擎SE01: \$20-30K
  - $\forall$ (中国供应链)  $\rightarrow$   $\exists$ (成本优势  $\wedge$  快速迭代)
3.  $D \rightarrow$  发展趋势 供应链创新:
  - 技术创新: - 傅利叶: FSA执行器 - 众擎: 谐波力控关节
  - 实践验证: 技术验证  $\rightarrow$  小规模量产  $\rightarrow$  供应链成熟  $\rightarrow$  规模化突破

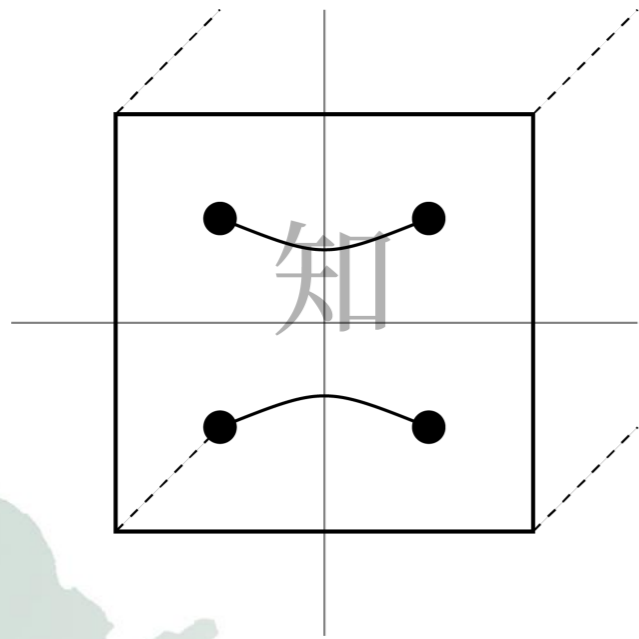
## 本质洞见

1. 机器人零部件的成本问题，将决定具身智能的落地进展
2. 中国供应链有望重复新能源汽车的成功，通过成本优化推动产业规模化发展



# 空间智能

通过多维感知和理解来构建三维世界模型，  
实现空间定位、场景理解与环境交互的认  
知系统。



洞悉维度，空间致知

## 核心观察

- A: 空间智能是具身智能的关键基础
- B: 3D数据获取是主要瓶颈
- C: 数据与算法双轨并进
- D: 空间智能复制ImageNet成功路径

## 逻辑链条

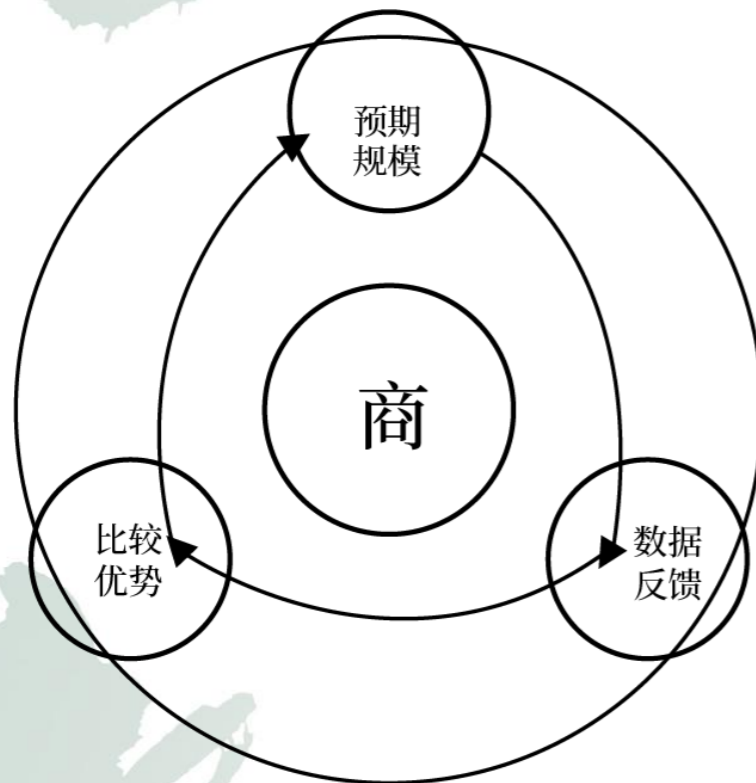
1. A → 发展必要性  
具身智能需求: - 多维感知 - 空间理解 - 物理交互
2. B ∧ C → 突破路径  
数据创新: - 数字表亲(ACDC) - WonderWorld(FLAGS)  
技术突破: - ReKep框架 - 关系关键点约束  
 $\forall(\text{空间智能}) \rightarrow \exists(\text{数据基础} \wedge \text{算法创新})$
3. D → 发展模式  
ImageNet路径复制: 经验复制 ∧ 维度升级  
2D图像 → 3D场景    标注分类 → 空间关系

## 本质洞见

1. 空间智能要做的是ImageNet路径复制与维度升级
2. 价值巨大，通过数据积累与算法创新，为具身智能构建认知世界的基础能力

# 机器人商业闭环

技术创新与市场需求相互促进、循环迭代，  
形成可持续发展的商业生态系统。



势成于市，道在循环

## 核心观察

- A: 市场需求驱动技术发展
- B: 三类关键价值方向浮现
- C: 实际落地需考虑比较优势
- D: 数据积累是重要目标

## 逻辑链条

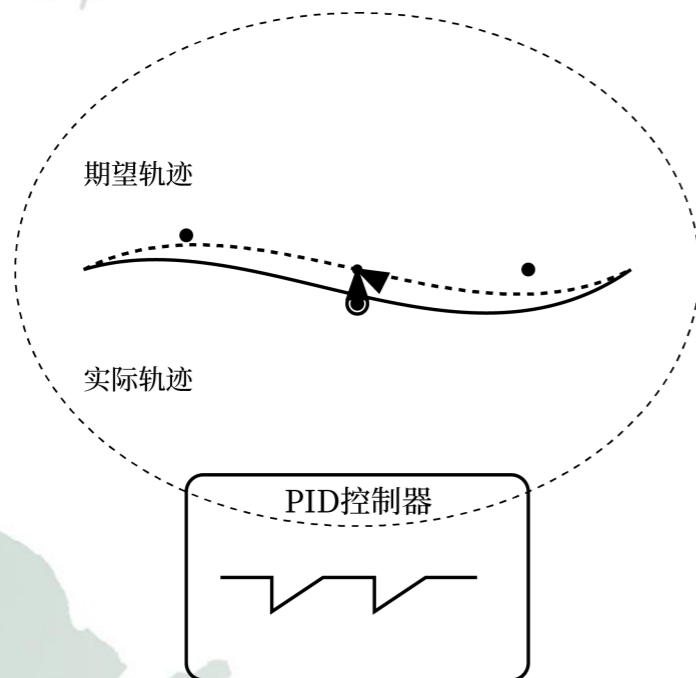
1. A → 市场优先原则：技术 ≠ 市场成功  
案例佐证：- 增程式电动车 - 苹果终端  
技术创新 < 市场匹配    工程实现 < 场景价值
2. B → 价值方向  
规模预期：- 老年护理(人口老龄化)  
数据反馈：- 工厂应用(验证优化)  
比较优势：- 手术机器人(专业场景)  
 $\forall$ (成功落地) →  $\exists$ (规模  $\vee$  数据  $\vee$  优势)
3. C  $\wedge$  D → 落地策略  
验证场景：数据积累(短) → 能力提升(中) → 规模应用(长)

## 本质洞见

1. 机器人产业发展需要找准市场切入点，通过场景应用积累数据与经验
2. 不一定非要先进人形机器人，可能是手术机器人、养老半人形机器人等

# 运动控制

基于动力学理论和反馈机制的执行系统，实现关节驱动的精准、稳定与柔顺控制。



动静统御，刚柔相济

## 核心观察

- A: 虚实差距是核心挑战
- B: 环境适应需要进化
- C: 动作规划追求通用
- D: 控制策略要求泛化

## 逻辑链条

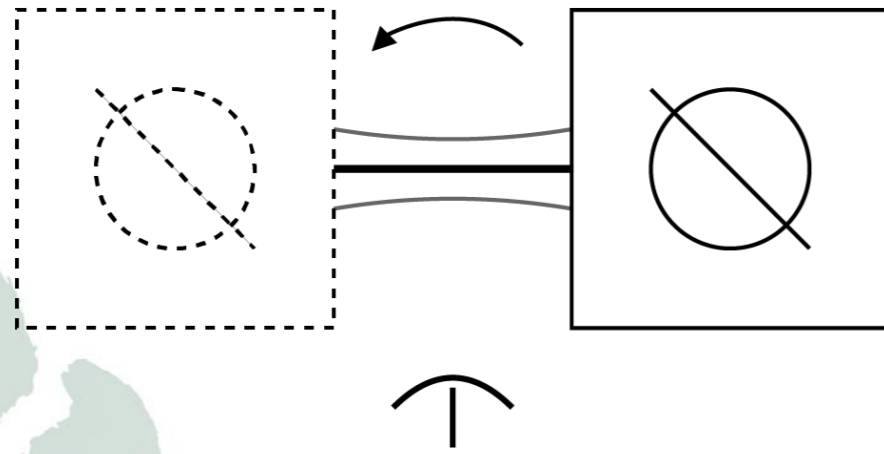
1. A → 突破方向  
虚拟环境 vs 现实环境 ⇒ 简单变量 vs 复杂变量  
单一任务 → 通用智能 ⇒ 固定场景 → 多样环境
2. B ∧ C → 技术路径  
环境适应: - DrEureka: AI生成奖励函数  
动作规划: - DeepMind: 自我进化学习  
参数穷举 → 自主学习 → 通用控制
3. D → 通用控制的统一方案  
HOVER模型: - 150万参数 - 多任务整合 - 双向适配  
 $\forall$ (运动控制) →  $\exists$ (环境适应 ∧ 策略泛化)

## 本质洞见

1. 运动控制是机器人研究的核心领域，生成式AI、强化学习带来新思路
2. 机器人控制需要从特定任务向通用能力演进，并通过统一模型实现多场景适应

# Sim2Real

通过虚拟仿真环境训练智能模型，实现向真实世界的高效迁移与泛化，弥合模拟与现实的差异鸿沟。



虚实共生，跨域涅槃

## 核心观察

- A: Sim2Real是连接虚拟与现实的桥梁
- B: 游戏环境提供理想训练平台
- C: 训练方法需多层次协同
- D: 现实迁移面临系统性挑战

## 逻辑链条

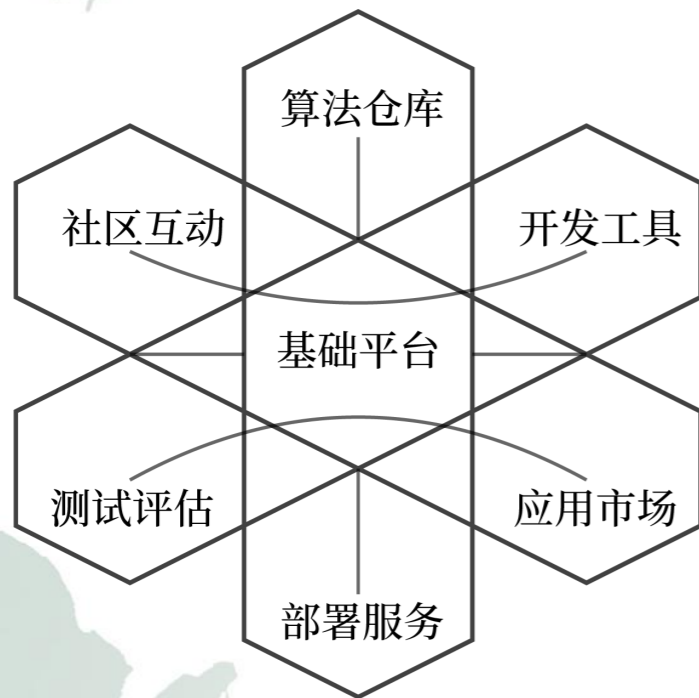
1. A → 基础框架  
仿真训练流程: 虚拟环境 → 策略学习 → 现实迁移  
核心机制: - 强化学习 - 自监督学习 - 领域随机化  
↓
2. B → 游戏价值  
平台优势: - 真实物理引擎 - 丰富交互机制 - 自动数据标注  
技术融合: 游戏引擎 + AR/VR + 强化学习 ⇒ 实现虚实无缝过渡  
训练特性:  $\forall(\text{游戏环境}) \rightarrow \exists(\text{可控性} \wedge \text{可重复性} \wedge \text{安全性})$   
价值实现: - 降低训练成本 - 提供安全环境 - 加速迭代验证  
↓
3. C  $\wedge$  D → 实现路径  
技术协同: - 虚拟训练(基础) - 迁移学习(过渡) - 现实微调(适应)

## 本质洞见

1. 机器人的大规模训练在物理世界行不通，尤其在训练初期危险度高
2. 游戏环境为Sim2Real提供了理想的中间训练场，是实现虚实迁移的关键。

# 共创平台

连接开发者、算法与应用场景的开放生态，  
加速机器人技术创新与产业化落地。



开放共生，创智荟萃

## 核心观察

- A: 具身智能需要多学科融合
- B: 独立研发效率低下
- C: 当前，两种协作路径并存
- D: 混合模式或成趋势

## 逻辑链条

1.  $A \wedge B \rightarrow$  协作必要性  
跨领域融合: - 机械工程 - 自动化控制 - 机器学习 - 认知科学  
协作动力: 技术复杂性  $\wedge$  资源互补性  $\rightarrow$  多方参与
2.  $C \rightarrow$  平台模式  
闭源平台: - Project GR00T(产业联盟)  $\Leftrightarrow$  深度整合、安全可靠  
开源社区: - LeRobot(共创工具包)  $\Leftrightarrow$  快速迭代、创新活跃
3.  $D \rightarrow$  发展趋势  
 $\forall$ (协作平台)  $\rightarrow \exists$ (技术共享  $\vee$  商业保护)  
演进方向: 单一模式  $\rightarrow$  混合模式 竞争关系  $\rightarrow$  互补共生

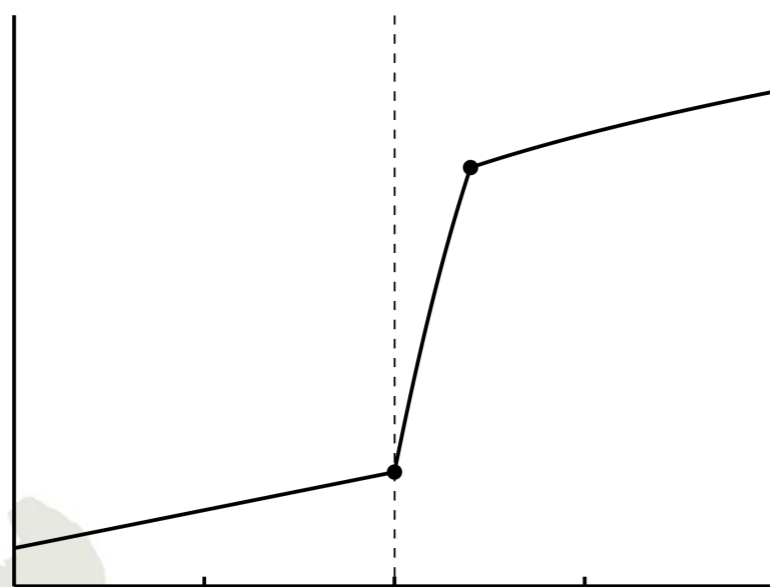
## 本质洞见

1. 具身智能的发展是一项长链的艰巨任务，需要群体的智慧，建立共创
2. 商业模式存在差异差异，需要兼容开源与闭源，最大限度多方协作加速技术突破

基准测试 Transformer 加速推理  
语音交互 Scaling Law 训练效率  
数学推理 推理能力 AI芯片 MoE架构  
代码生成 推理速度 API集成  
多模态融合 基础模型 强化学习  
推理速度 性能评估  
规模优化 开源生态 模型微调 长文本处理  
分布式训练 系统部署  
慢思考 计算成本 高级语音 函数调用  
RNN架构 思维链 合成数据

# Scaling Law

模型规模与能力的基础增长规律，  
揭示智能涌现的量变质变辩证关系。



量变质变，智慧涌现

## 核心观察

- A: 预训练模式将遇瓶颈
- B: 数据增长速度受限
- C: 传统Scaling模式难以持续
- D: 新技术路径亟需探索

## 逻辑链条

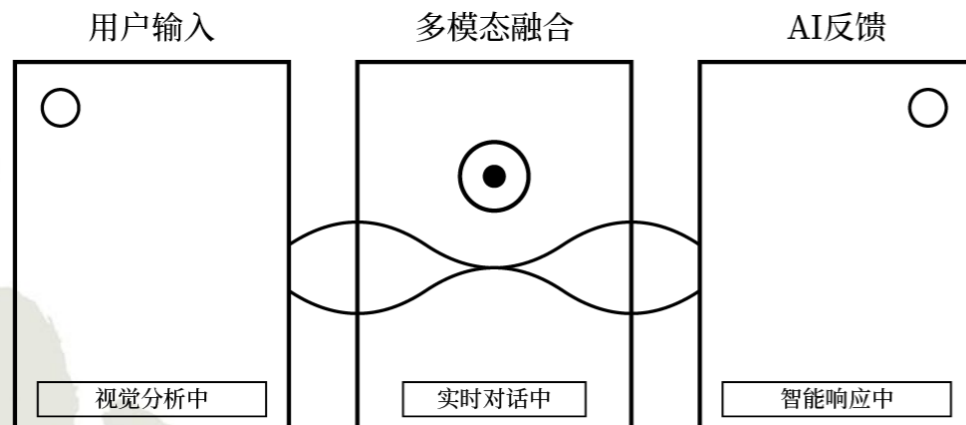
1.  $A \wedge B \rightarrow$  发展困境  
Ilya观点: - 数据如化石燃料 - 互联网数据有限 - 算力  $>$  数据增长  
 $\forall$ (规模扩张)  $\rightarrow \exists$ (数据限制  $\wedge$  效益递减)
2. C  $\rightarrow$  技术转折点  
传统路径局限: 预训练依赖数据量, 需要寻找新的增长点  
- 模式匹配  $\rightarrow$  推理能力  $\Leftrightarrow$  - 直觉模仿  $\rightarrow$  自主思考
3. D  $\rightarrow$  突破方向  
新范式探索: 更多类似人类进化的新Scaling模式, 慢思考、稀疏数据等  
- Agent自主性 - 合成数据生成 - 推理时计算  
- 生物学方法

## 本质洞见

1. AI正在突破传统Scaling范式, 探索更高效的智能涌现机制
2. 新的增长点, 或来自于生物学而非仅是物理学启发
3. 类似人类进化的新Scaling模式, 而非简单的线性扩展

# 高级视频语音模式

将AI助手能力扩展到实时视频通话场景，  
通过多模态交互实现沉浸式对话体验。



视听通感，如临其境

## 核心观察

- A: 多模态视觉交互能力上线
- B: 实时互动体验显著提升
- C: 人格化特征逐步完善
- D: 竞争压力日益加剧

## 逻辑链条

- A  $\wedge$  B  $\rightarrow$  功能突破  
能力拓展: - 视频通话理解 - 屏幕内容共享 - 实时教学指导  
创新方向: 单一模态  $\rightarrow$  多维交互 被动响应  $\rightarrow$  主动引导  
 $\forall$ (交互升级)  $\rightarrow$   $\exists$ (多模态  $\wedge$  实时性)
- C  $\rightarrow$  交互升级  
人性化特征: - 情感化反馈 - 多样化声音 - 记忆与学习
- D  $\rightarrow$  市场竞争  
OpenAI与谷歌对比:  
- 功能相似度高 - 时间差劣势 - 技术实力待证  
技术跟进 < 创新引领 功能对标 < 场景创新

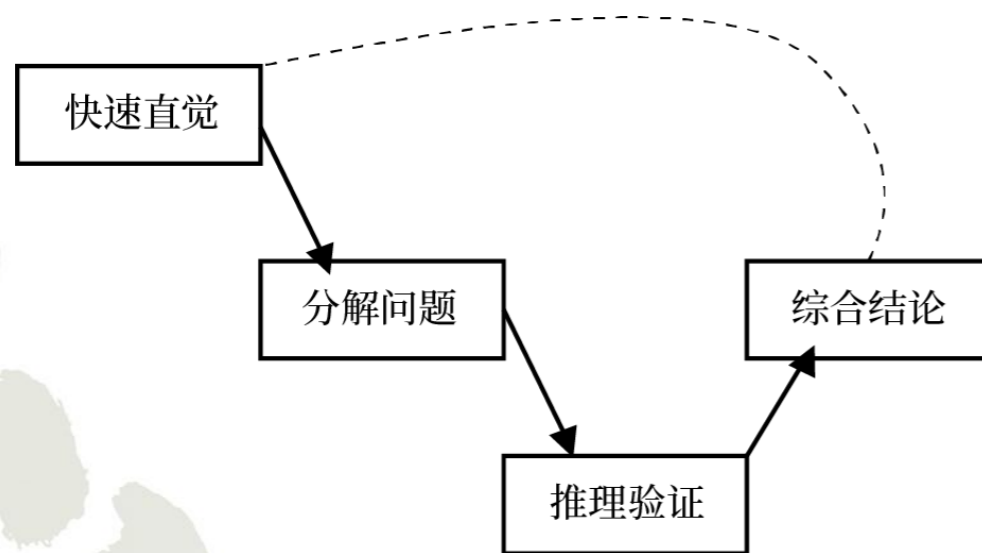
## 本质洞见

- 高级视频语音效果惊艳，但同样在技术上不存在明显的竞争壁垒
- 竞争者需在多模态交互基础上，探索独特价值与应用场景



# 慢思考

通过递进式的深度推理和验证机制，  
构建系统化的思维链路以获得可靠结论。



循序渐进，深思熟虑

## 核心观察

- A: AI进入通用智能新阶段
- B: 规模定律遇到数据瓶颈
- C: 强化学习成为新范式
- D: 产品能力由模型决定，正在发生改变

## 逻辑链条

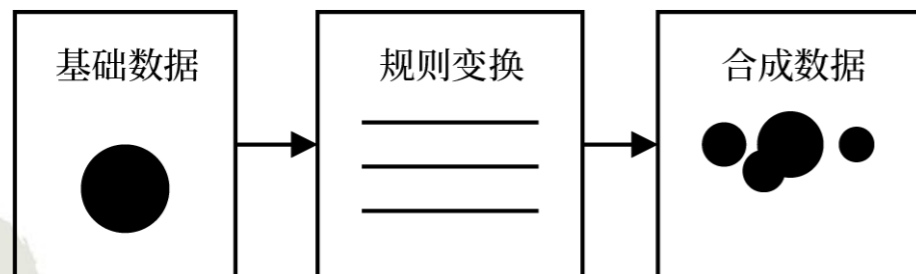
1.  $A \rightarrow$  通用智能形成条件  
必要因素: - 互联网数据积累 - 算力突破  
- Transformer架构  
 $\forall$ (因素缺失)  $\rightarrow$   $\neg$ (通用智能)  
数据规模  $\wedge$  算力限制  $\rightarrow$  范式创新需求
2.  $B \wedge C \rightarrow$  范式转变  
数据瓶颈: - 优质数据耗尽 - 专业数据不足  
强化学习突破: - 自我对弈 - 思维链生成 - 推理时间延长  
突破路径: 被动学习  $\rightarrow$  主动思考 即时响应  $\rightarrow$  延时推理
3.  $D \rightarrow$  产品发展规律  
算力转移: 训练端  $\rightarrow$  推理端  
技术能力  $\rightarrow$  产品能力 通用助理  $\rightarrow$  超级应用

## 本质洞见

1. AI正从规模定律向强化学习转变，通过主动思考突破数据瓶颈限制

# 合成数据

利用规则和模型构造的人工数据集，  
补充现实数据的不足并增强训练效果。



衍生多样，增益求真

## 核心观察

- A: 合成数据易，高质量合成数据难
- B: 合成数据的研究方法不断发展（如英伟达、World Labs等）
- C: 合成数据的评估是自循环的关键

## 逻辑链条

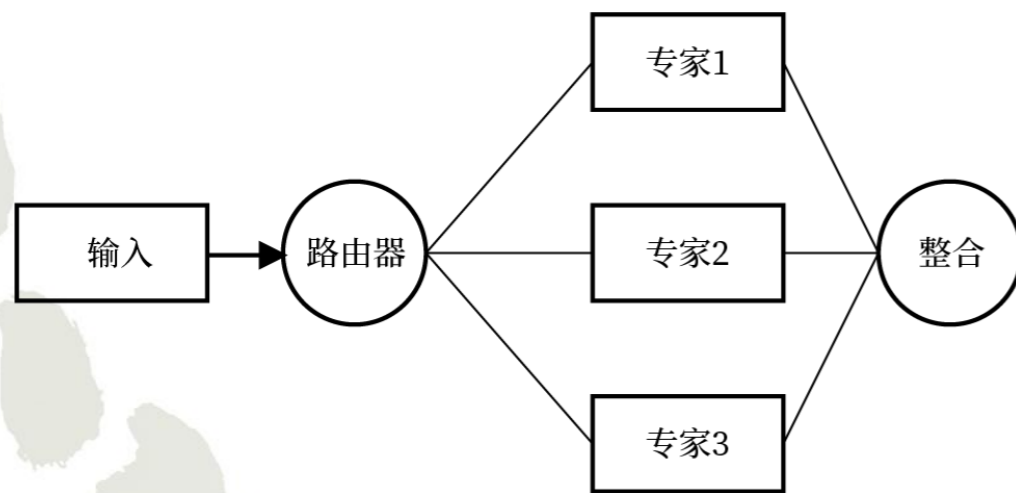
1. A → 价值递进：  
数据量 → 数据质量 → 数据结构  
质量提升的两难：结构化程度 ↑ 真实性 ↓
2. B → 系统进化：  
工具属性：- 数据制造 - 规模扩充  
生态属性：- 知识重组 - 自我优化  
结构化 ∧ 语义化 ∧ 模块化 → 数据的可理解性  
范式转变：合成数据是知识创造过程，而非简单的数据制造
3. C → 未来方向：  
合成范式：单向生成 → 对抗验证 → 系统共生 → 自我进化  
最终目标：数据工具 → 知识系统 规模扩张 → 质量提升

## 本质洞见

1. 合成数据的量变并不能直接产生质变，数据的质量、结构至关重要
2. 合成数据需要从扩充数据转向创造知识，通过系统化方法提升训练数据质量

# MoE架构

动态路由到不同专家网络的混合系统，  
实现大规模模型的条件计算和稀疏激活。



分工协作，智慧共融

## 核心观察

- A: MoE由专家模型与门控网络组成
- B: 架构具有稀疏性与扩展性特征
- C: 支持多任务学习与知识共享
- D: 在多领域展现应用价值

## 逻辑链条

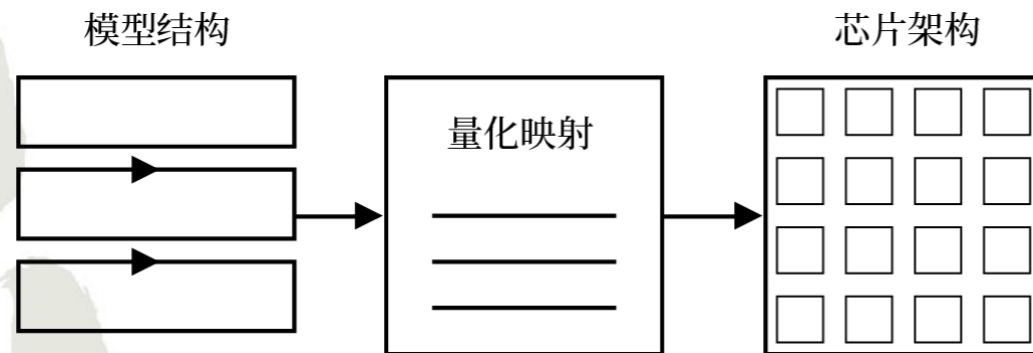
1. A → 基础架构  
组件构成：
  - 专家模型(任务处理) - 门控网络(动态路由)
$$\forall(\text{MoE系统}) \rightarrow \exists(\text{专家分工} \wedge \text{动态调度})$$
2. B  $\wedge$  C → 核心特征  
技术优势：
  - 稀疏激活降低开销 - 灵活扩展应对复杂性 - 多任务共享知识优化方向：计算效率  $\wedge$  模型性能  $\wedge$  架构扩展性
3. D → 应用价值  
落地场景：
  - NLP: 翻译/问答 - CV: 分类/检测 - 推荐: 个性化服务挑战权衡：性能提升 vs 复杂性增加 灵活性 vs 训练稳定性

## 本质洞见

1. MoE架构为模型预训练提供了一条有效的路径，突破规模上限
2. MoE通过专家分工与动态调度，实现计算效率与模型性能的优化

# 加速推理

深度定制模型结构与芯片架构，  
实现高效能的推理计算。



**芯模相融，提速增效**

## 核心观察

- A: 多种硬件架构各有特色
- B: 算法与芯片深度融合
- C: 不同场景有差异化需求
- D: 应用领域持续扩展

## 逻辑链条

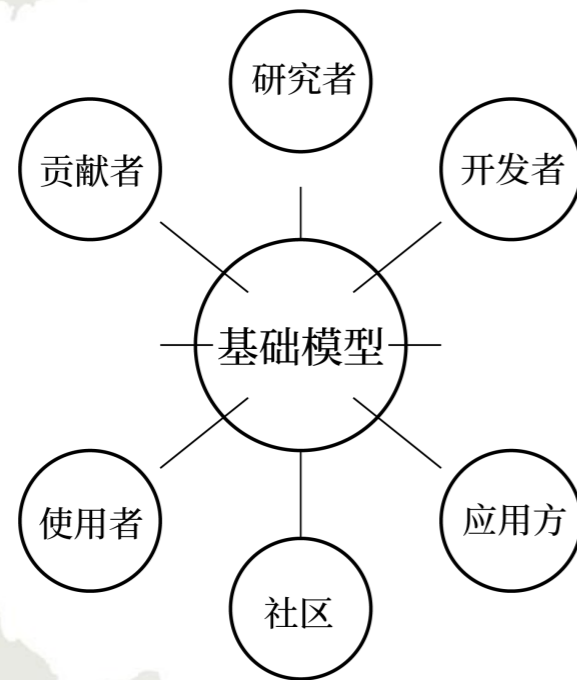
1. A → 硬件特征分析  
处理器对比:
  - CPU: 高主频(3.51GHz)/大缓存优势
  - ASIC: 专用高效/能耗优化
  - GPU: 并行计算强/功耗高限制
  - FPGA: 灵活可编程/场景定制
2. B → 技术融合案例
  - Groq LPU: 内嵌SRAM/500T ops
  - AlphaChip: 强化学习优化布局
3. C ∧ D → 场景应用  
 $\forall(\text{应用场景}) \rightarrow \exists(\text{性能需求} \wedge \text{资源约束}) \rightarrow \text{最优架构选择}$ 
  - 自动驾驶: 实时性要求
  - 金融分析: 风险控制
  - 医疗诊断: 精确性需求
  - 智能家居: 交互体验

## 本质洞见

1. 除GPU外，将从通用处理迈向专用优化加速，并推进AI基础设施协同优化
2. AI加速推理需要基于场景、性能、能效等，选择合适架构实现性能优化

# 开源生态

基础模型通过开放协议释放能量，  
形成自组织协同进化的创新共同体。



汇智共创，生生不息

## 核心观察

- A: 开闭源争论反映商业策略差异
- B: 开源模型与开源软件有区别
- C: 大模型开源受到高度关注
- D: 商业可持续性面临挑战

## 逻辑链条

1. A → 商业策略分化  
路线选择: 开闭源并存 → 差异化竞争
  - 纯开源: Meta等重社区
  - 双轨并行: 微软等重生态
  - 纯闭源: OpenAI等重商业
2. B → 开源特点  
区别开源软件:
  - 仅开放部分参数
  - 缺乏完整技术细节
  - 社区贡献受限
3. C ∧ D → 市场现状  
需求维度: - 企业研发验证 - 低成本应用 - 安全自主可控  
生存挑战: - 负毛利运营 - 商业模式待解 - 行业整合加速  
技术路线 < 商业模式 开源闭源 < 市场需求

## 本质洞见

1. 区别于开源软件，大模型的开源并不是拿来即用
2. 开源模型受到广泛关注，但与闭源比较各有优劣，最终将由市场需求决定竞争格局

# 逻辑操作符说明

AGI  
AI 50  
KEYWORDS

## 逻辑符号

1.  $\neg$  非（否定）：表示对一个命题的否定。例如， $\neg p$  表示“不是  $p$ ”或“ $p$  不成立”。
  2.  $\forall$  全称量词：表示“对于所有”，例如  $\forall x P(x)$  表示“对于所有  $x$ ， $P(x)$  都成立”。
  3.  $\exists$  存在量词：表示“存在”，例如  $\exists x P(x)$  表示“存在某个  $x$ ，使得  $P(x)$  成立”。
  4.  $\rightarrow$  充分条件： $p \rightarrow q$  表示“如果  $p$  成立，那么  $q$  也成立”，即“ $p$  是  $q$  的充分条件”。
  5.  $\wedge$  且（与）：表示两个命题都为真时，整体结果才为真。例如， $p \wedge q$  表示“ $p$  且  $q$  都为真”。
- 

## 推理符号

1.  $\Rightarrow$  推导关系：表示“可以推导出”，例如  $p \Rightarrow q$  表示“从  $p$  可以推导出  $q$ ”。
  2.  $\Leftrightarrow$  等价关系：表示两个命题可以互相推导。例如， $p \Leftrightarrow q$  表示“ $p$  和  $q$  是等价的”，即  $p$  与  $q$  可以互相推导。
- 

## 推理法则

1. 双重否定律： $\neg\neg p \Leftrightarrow p$ ，表示否定一个命题的否定与原命题等价。
2. 对置律： $(p \rightarrow q) \Leftrightarrow (\neg q \rightarrow \neg p)$ ，表示“如果  $p$  推出  $q$ ，那么  $\neg q$  推出  $\neg p$ ”。
3. 传递律： $(p \rightarrow q) \wedge (q \rightarrow r) \Rightarrow (p \rightarrow r)$ ，表示如果  $p$  推出  $q$ ，并且  $q$  推出  $r$ ，则  $p$  推出  $r$ 。

# AI50关键词附录

AGI  
AI 50  
KEYWORDS



腾讯研究院AI速递



AI每周关键词Top50



科技九宫格短视频



AGI路线图圆桌



AGI路线图专题



AIS百人百问

# 研究团队

AGI  
AI 50  
KEYWORDS

**项目顾问：**司晓、冯宏声、刘琼

**项目策划：**李瑞龙、袁晓辉

**研究团队：**李瑞龙、曹士圯、袁晓辉、徐思彦

**联合研究：**李继刚、数字生命卡兹克