

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。

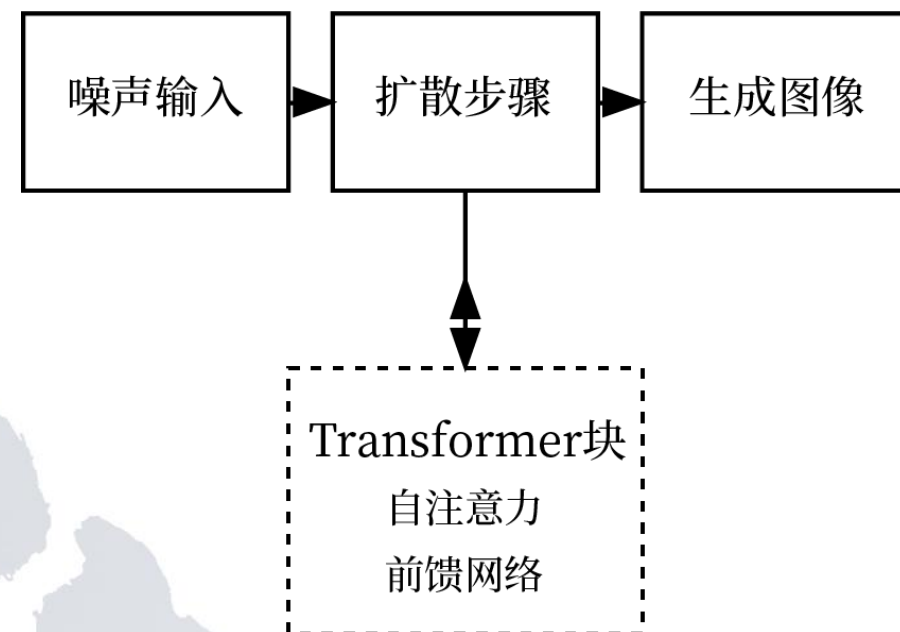
基于全年研究积累的三十余万字AI进展数据库，对当前AI发展进行阶段性总结具有重要意义。为了系统呈现AI发展的关键技术要点和趋势，该报告精选了50个年度关键词，覆盖大模型技术的八大领域：图像处理、视频生成、3D生成、编程助手、Agent、端侧智能、具身智能和基础模型；借鉴大模型的思维特征，创新性

的通过"快思考

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

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

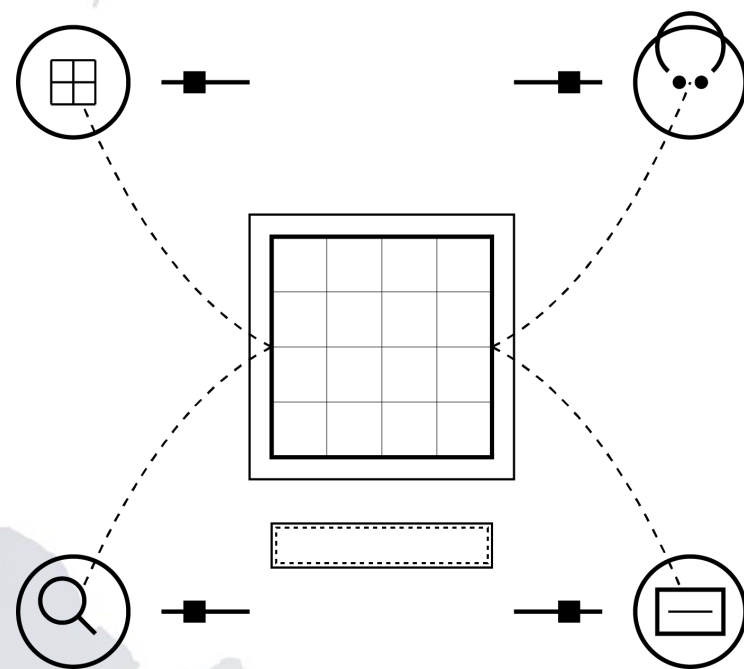
1. A → 技术演进：
 - 下一个字符 → 下一个像素
 - 突破：序列建模能力迁移
2. B → 架构优势：
 - 替代：U-Net → MMDiT (SD3、Flux、混元文生图)
 - 提升：- 空间关系理解 - 复杂提示处理 - 细节还原能力
3. A + B → C: 规模效应
 - 参数规模：800M → 12B
 - 涌现能力：- 真实度提升 - 控制力增强 - 细节完善

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

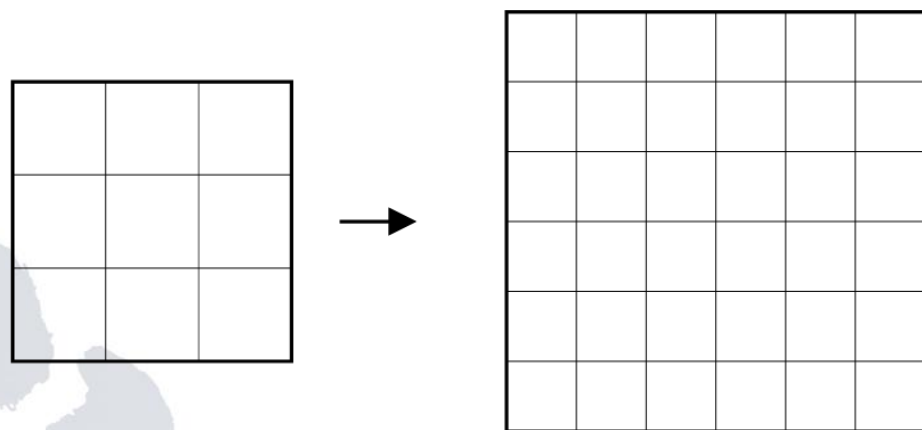
1. A → 控制维度演进：
 - Prompt: 文本描述引导
 - LoRA: 低成本模型微调
 - ControlNet: 精确条件控制
2. B → 技术路径分化：
 - 描述控制: 语义理解
 - 参数控制: 模型微调
 - 条件控制: 额外输入引导
 - 工作流控制: 外部编排
3. C → 精确控制突破：
 - 光影: IC-Light照明控制
 - 轮廓: Paints-Undo创作追溯
 - 构图: Omost自动扩展

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

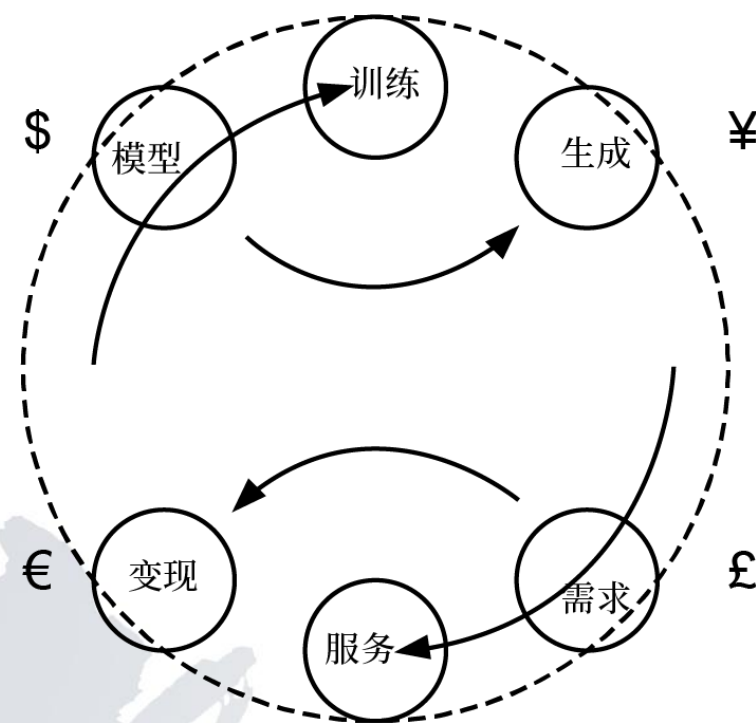
1. $A \wedge B \rightarrow$ 产业应用扩展
生成模型支持1024×1024原生分辨率
 $\rightarrow$ 艺术创作/广告/游戏开发等应用提升
2. $A \wedge C \rightarrow$ 专业领域突破
医学影像分辨率需求[256-1024] $\wedge$ 模型达到1K处理能力
 $\rightarrow$ 专业应用可行
3. $D \rightarrow$ 技术演进方向
 $\forall$ (高分辨率处理) $\rightarrow \exists$ (架构创新 $\vee$ 性能优化)
例: Pixtral 12B, Eagle系列针对分辨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

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

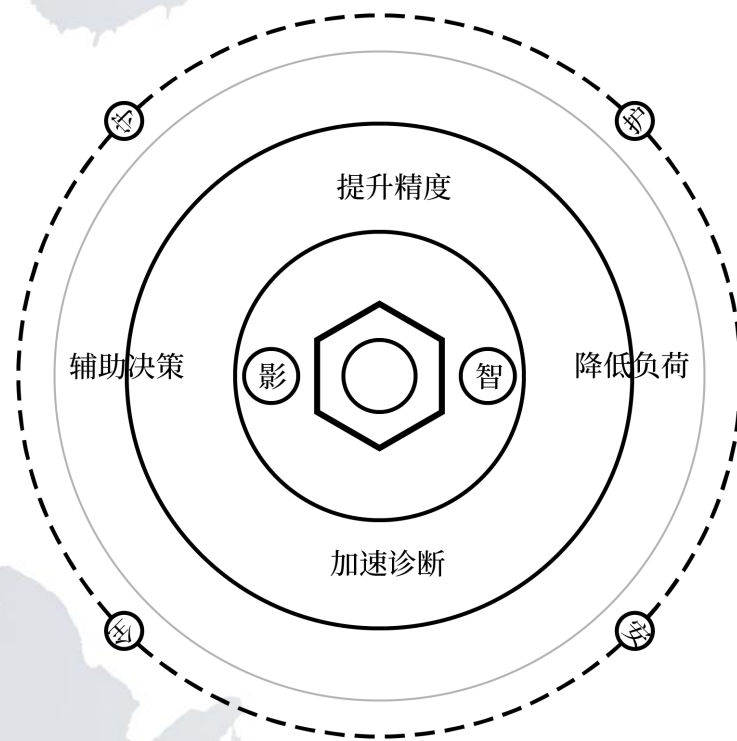
1. $A \wedge \neg B \rightarrow C$ 为什么要转型？
技术能力 $\neq$ 商业价值
流量优势 $\neq$ 变现能力
(技术优势 $\wedge$ 商业化失败) $\Rightarrow$ 寻求转型
2. $C \rightarrow D$ 案例佐证：
 - Stability AI: 技术困境 $\rightarrow$ 视效领域融合
 - Leonardo.ai: 独立运营 $\rightarrow$ 平台整合
3. $\forall$ (成功转型案例) $\rightarrow \exists$ (产业链整合 $\vee$ 场景深耕)
 - 原生技术 $\rightarrow$ 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. $A \rightarrow$ 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. $B \wedge C \rightarrow$ 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 $\rightarrow$ 临床验证 $\rightarrow$ 商业应用
 $\forall$ (成功医疗AI) $\rightarrow \exists$ (专业性 $\wedge$ 实用性 $\wedge$ 可靠性)
3. $D \rightarrow$ 发展趋势
领域专家认可 (Hinton、吴恩达等) $\Rightarrow$ 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 **视频生成** 实时生成

镜头语言 运镜控制

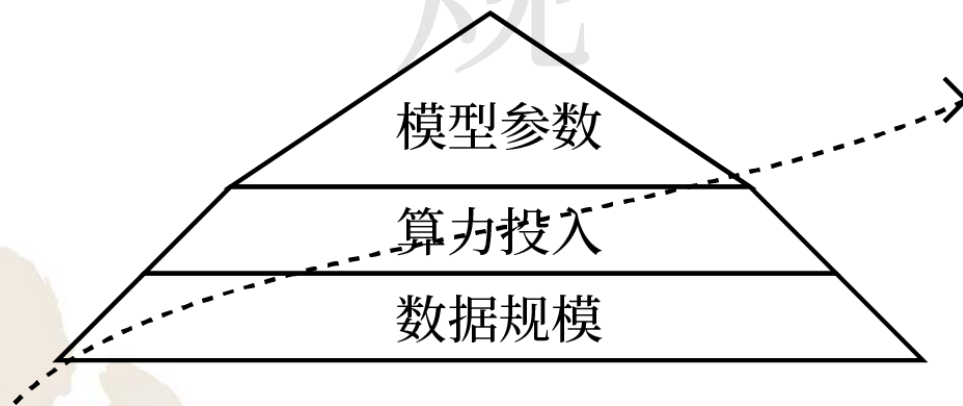
分辨率 生成式游戏 生成速度
视觉编辑 多模态融合

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

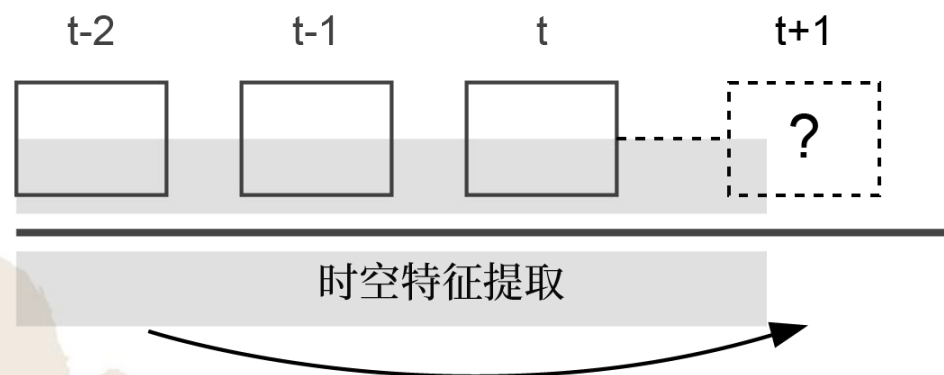
1. 问题难度跃升 (A)
视频生成 / 图像生成 $\approx$ 百倍复杂度：大量连续帧、时序连贯性、主体一致性
2. 技术路线探索 (B)
自回归Transformer方案 or 扩散模型 $\rightarrow$ 规模化训练
3. 技术突破与统一 (B $\rightarrow$ C $\rightarrow$ D)
DiT架构整合：
 - 融合Transformer与扩散模型优势
 - 引入时空块编码创新
 - 通过规模化训练实现性能突破

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

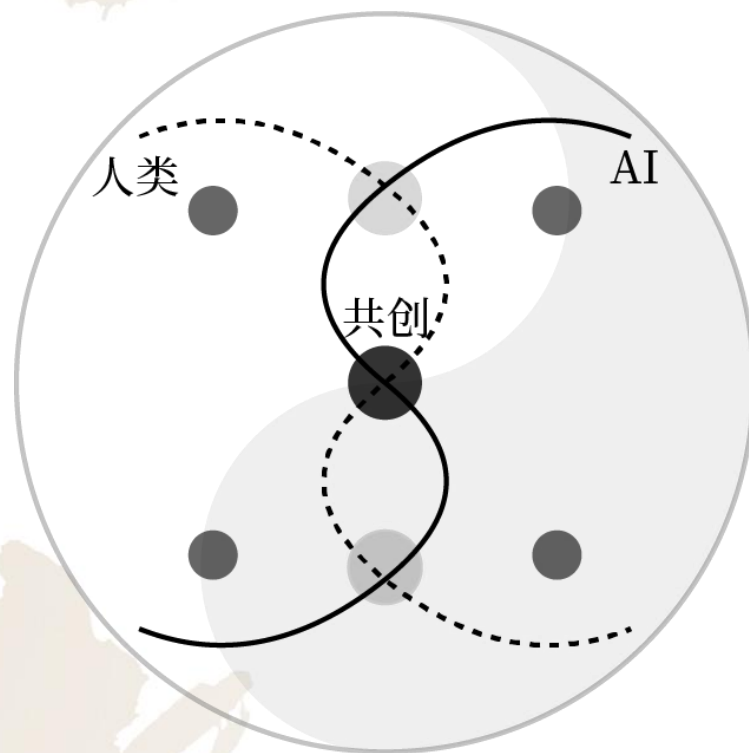
1. 技术路线分化 (A)
 - 扩散模型：噪声迭代 $\rightarrow$ 帧序列生成
 - 自回归Transformer模型：token预测 $\rightarrow$ 帧序列构建
2. 实现机制对比 (B $\wedge$ C)
 - DiT方案：整体扩散过程、时空特征同步建模
 - 自回归Transformer方案：视频token化、序列化预测
3. 预测能力 (D)
 - 连续性保证：时序特征学习 + 运动规律理解 + 状态迁移预测

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

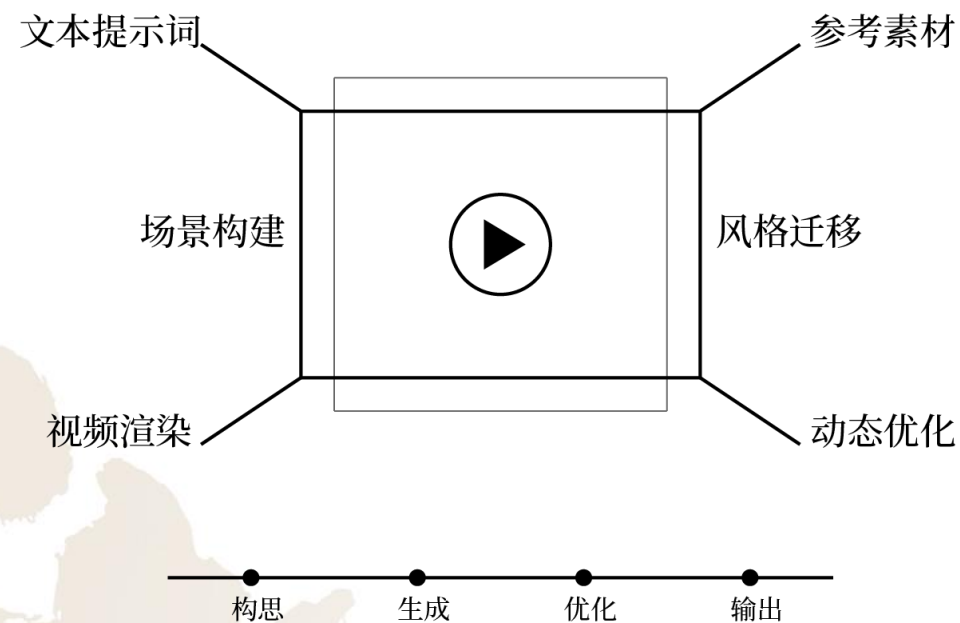
- 生态构建路径 ($A \wedge B$)
 - 打造工具产品 → 运营社区
 - 举办比赛活动 → 扩大影响
- 艺术家价值链 ($C \rightarrow D$)
 - 前端：优质作品展示
 - 中端：专业反馈收集
 - 后端：训练数据优化
- 闭环形成 ($A \wedge B \wedge C \rightarrow D$)
 - 工具应用 → 社区运营 → 艺术家合作 → 数据反馈 → 模型优化

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

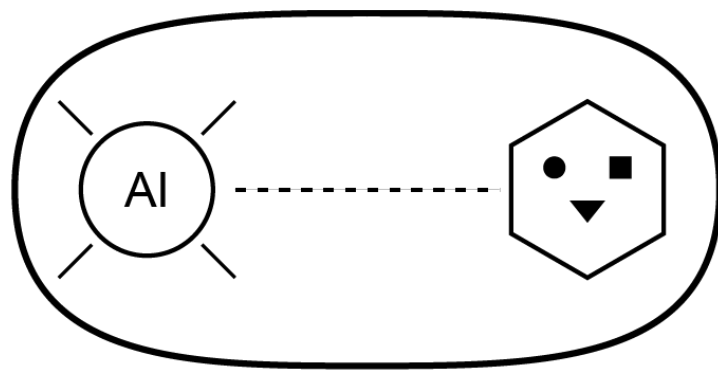
- 交互使用难度差异 (A): 文本模型(简单) → 视频模型(复杂)
- 能力扩展 (A → B)
文生视频 → 图生视频 → 首尾帧控制 → 视频生视频
- 控制增强 (B → C)
精细化控制 = 运镜控制 + 运动笔刷 + 主体选定 + ...
- 范式升级 (C → D)
传统创作工具 → AI原生创作平台 (多工具集成+ workflow 适配+专业功能对标)

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)

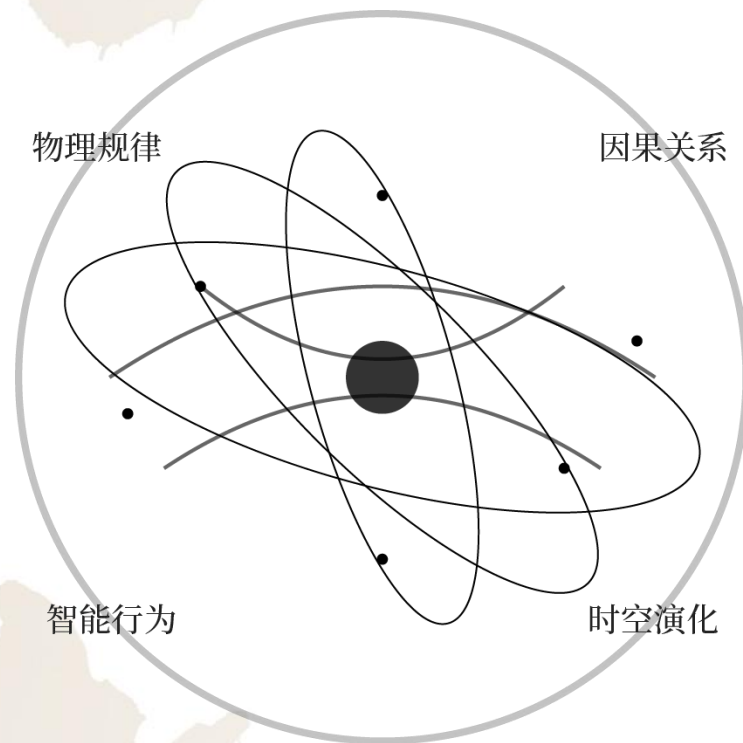
现实世界 = 复杂度指数级增长 + 数据收集成本激增 + 状态空间爆炸

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

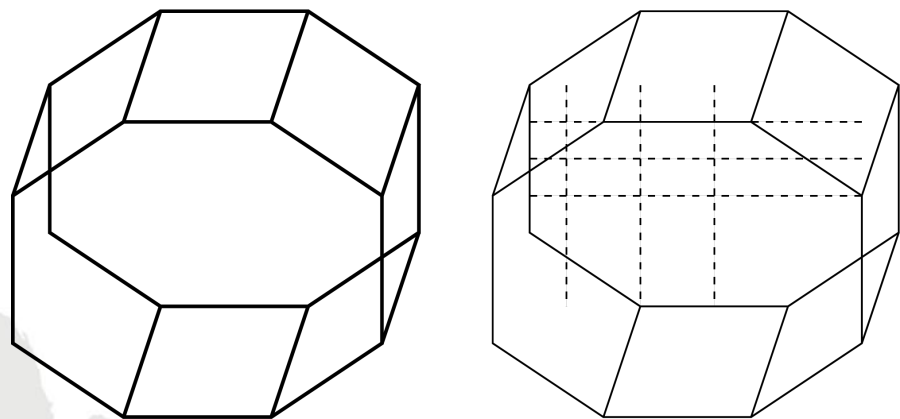
深度估计 3D生成 纹理质量
游戏场景 几何精度
实时渲染

虚拟世界 高斯泼溅 物理碰撞 AI元宇宙
角色生成 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

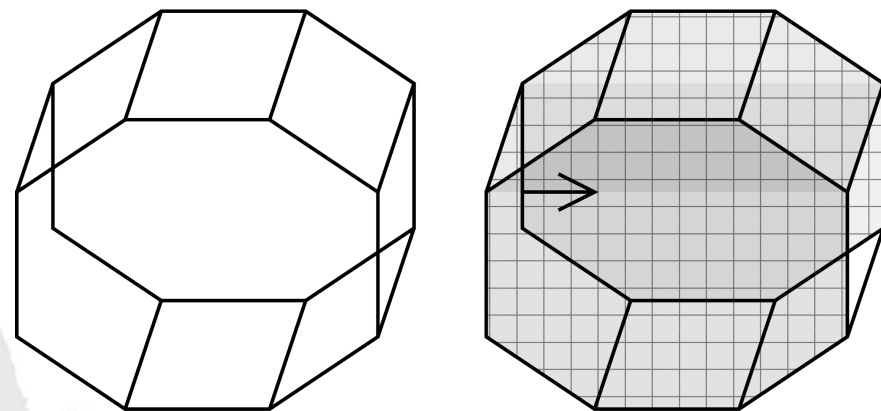
1. $A \rightarrow$ 主流技术方案
 - Tripo 2.0: DiT+U-Net架构
 - GRM: 四视图transformer重建
 - Unique3D: 多视图及法线扩散
2. $B \wedge C \rightarrow$ 应用价值
 - $\forall(\text{形态还原}) \rightarrow \exists(\text{速度} \wedge \text{质量} \wedge \text{效率})$ 广泛应用:
 - 游戏开发: 场景建模 - 影视制作: 特效场景
 - 工业设计: 产品迭代 - VR构建: 虚拟环境
3. $D \rightarrow$ 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

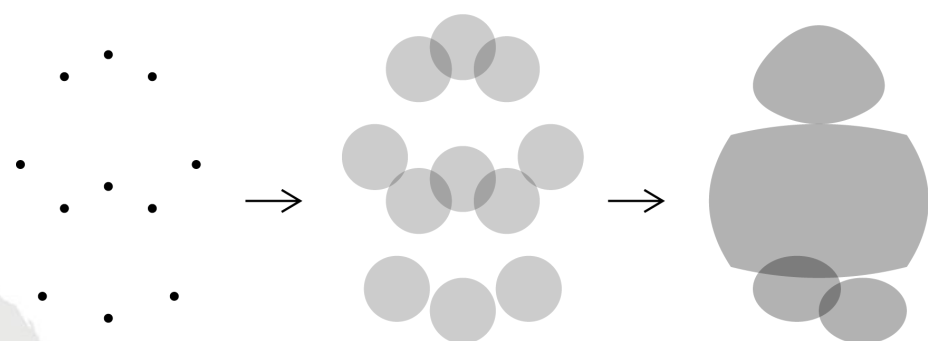
1. A → 基础方法体系 生成路径:
 - 基于图像: 纹理映射/无缝拼接
 - 基于模型: 细节增强/PBR渲染
 - 基于深度学习: GAN/CNN架构
2. B → 关键技术 技术要素:
 - UV展开与优化 - 程序化生成算法
 - PBR材质系统 - 纹理合成修复
3. C ∧ D → 发展趋势
算法突破 → AI赋能 → 应用拓展
AI应用: - 自动生成/转换 - 风格迁移/增强
 $\forall$ (纹理生成) → $\exists$ (自动化 ∧ 真实感)

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

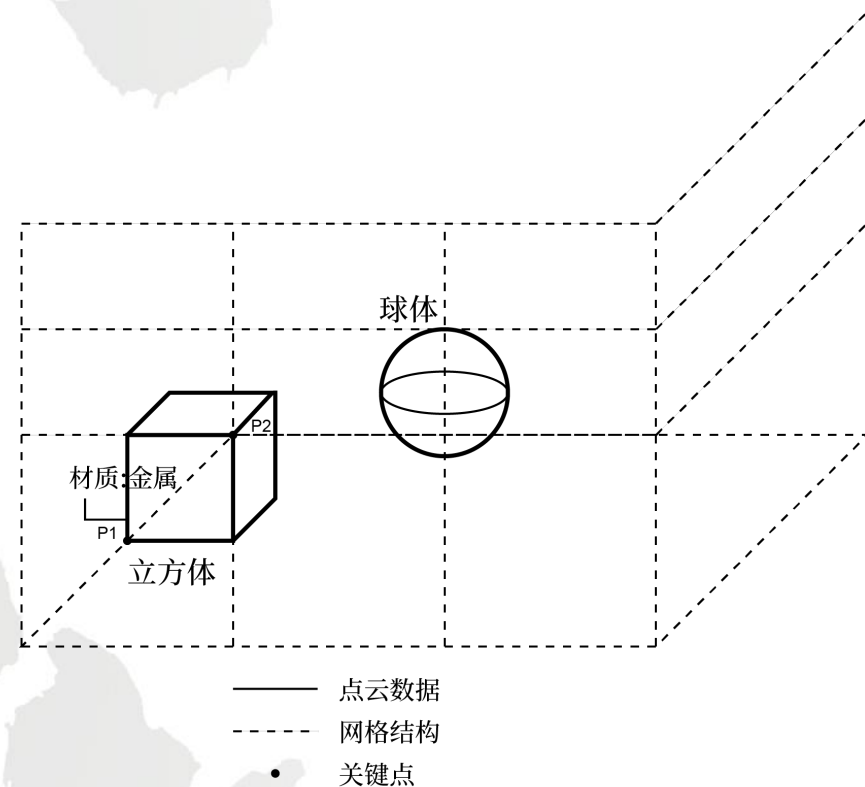
1. $A \wedge B \rightarrow$ 技术原理
光栅化 $\wedge$ 高斯函数描述 $\rightarrow$ 精确场景表达
属性: 位置 $\wedge$ 协方差 $\wedge$ 颜色 $\wedge$ 不透明度
2. $C \rightarrow$ 性能三角:
 - 高品质渲染($> \text{NeRF}$)
 - 实时性能($> 100\text{fps}@1080\text{p}$)
 - 训练效率($< 1\text{h}$)创新价值: - 数据采集简化 - 优化机制灵活 - 实时渲染提升
3. $D \rightarrow$ 应用价值
 $\forall (\text{高斯泼溅}) \rightarrow \exists (\text{效率} \wedge \text{质量} \wedge \text{易用性})$
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

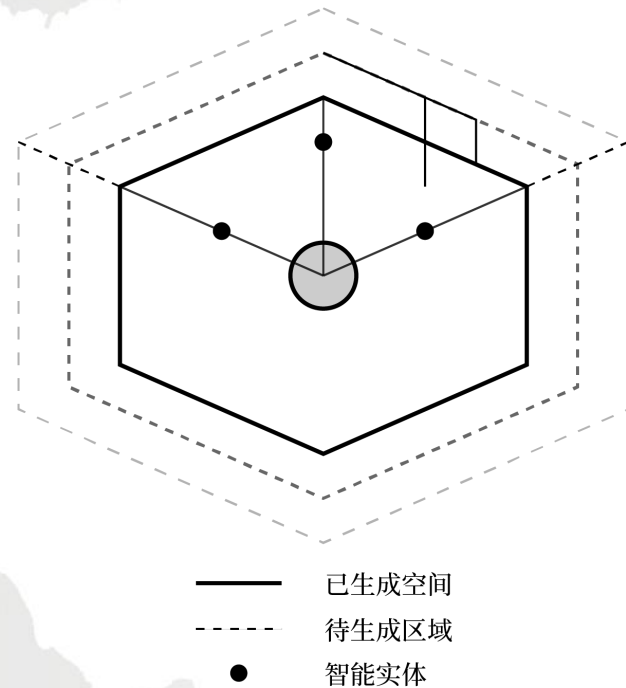
1. $A \wedge B \rightarrow$ 困境闭环，3D训练数据的稀缺性影响：
 - 高成本限制应用 - 小规模制约投入 - 低投入加剧稀缺
2. $C \rightarrow$ 破局路径，尝试中的技术演进：
 - 合成数据(Bootstrap3D) - 多模态融合(ULIP)
 - 领域适应(Swin3D++) - 单图生成(VFusion3D)构建正向循环：
技术创新 $\rightarrow$ 成本下降 $\rightarrow$ 规模扩大 $\rightarrow$ 持续优化
3. $D \rightarrow$ 应用牵引，明确的场景需求：
 - 自动驾驶(感知安全)
 -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

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



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

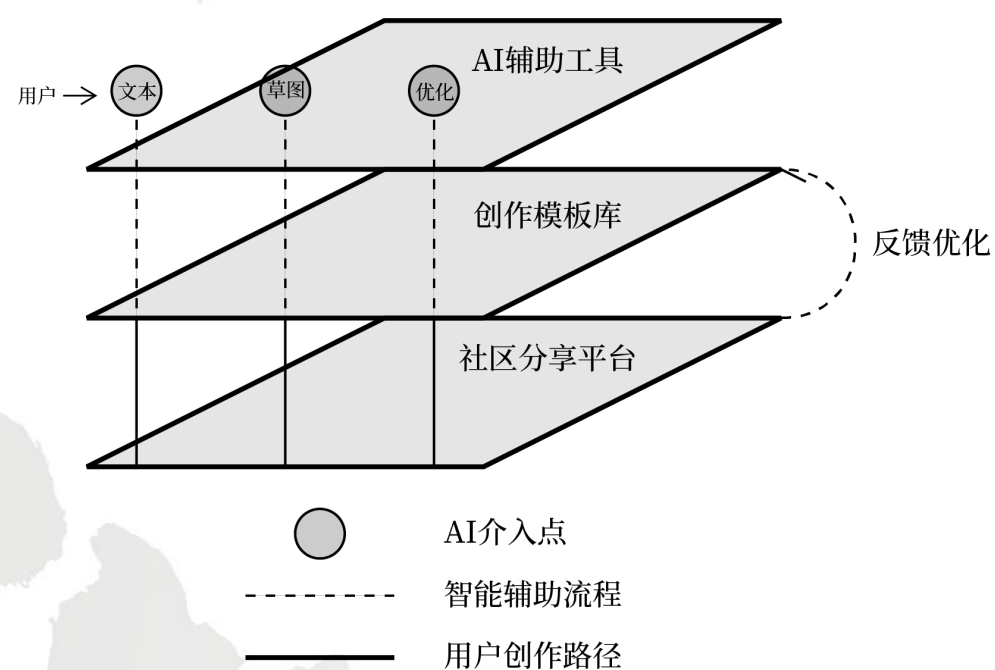
1. A → 内容生态重构
生产模式转变: 人工创作 → AI生成
供给特征: 内容稀缺 → 注意力稀缺
创作门槛: 专业技能 → 提示工程
2. B → 交互深度演进
功能 → 情感 → 价值, 每一维度都需要特定AI能力支撑
一阶交互: 功能响应 ⇒ 二阶交互: 情感链接 ⇒ 三阶交互: 价值共创
- 场景生成-角色互动 - 个性适配-情绪共鸣 - 协同创作-价值交换
3. C ∧ D → 系统性升级
基础设施: - 智能优化(资源分配) - 自适应调节(负载均衡)
价值重构: 内容价值 → 交互价值 → 网络价值

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

1. 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
2. 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
3. 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

1. 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)
2. 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

全栈生成

</

云端沙

#

服务机器人

