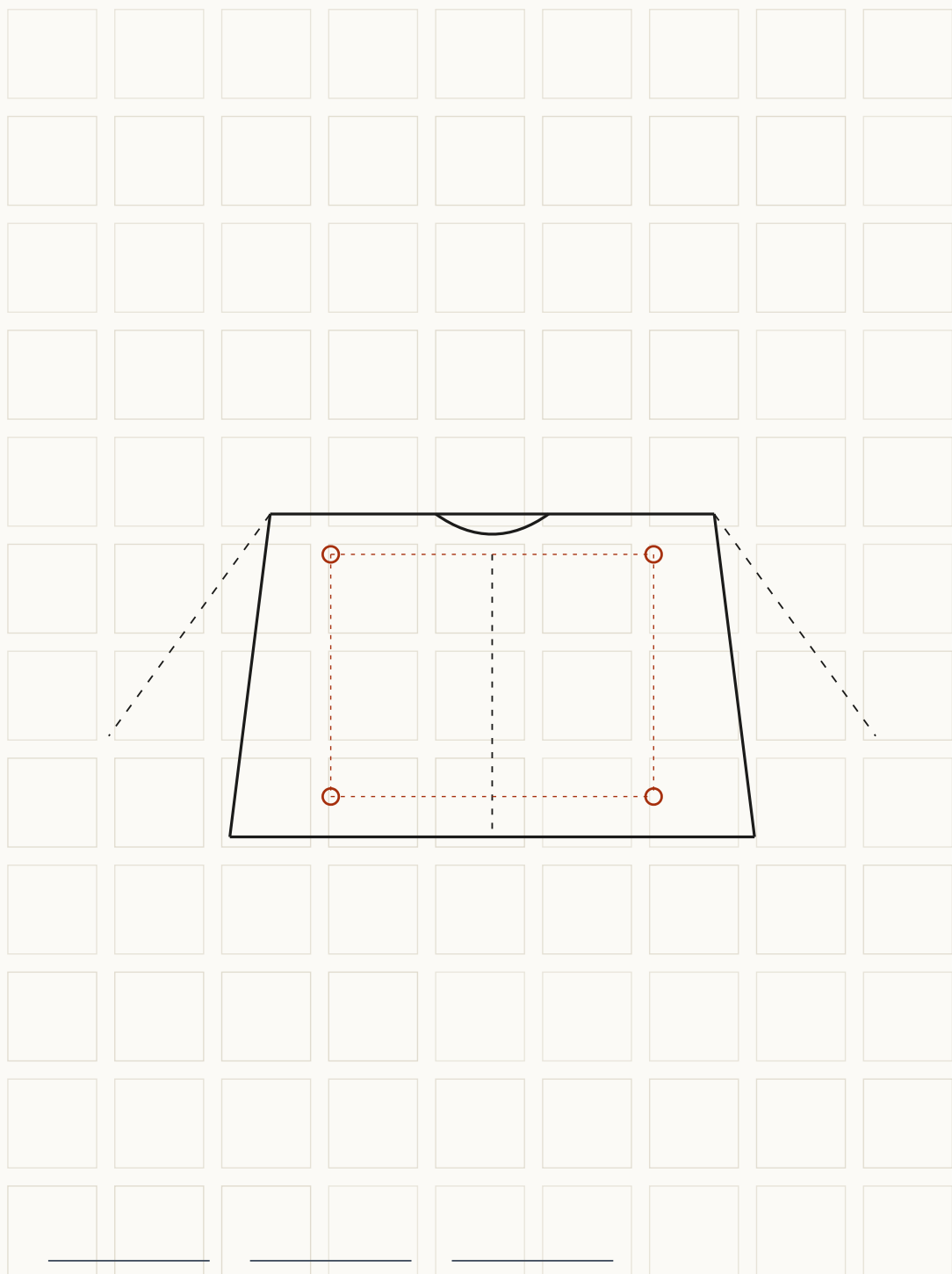




## 目录

1. [01行业现状：30% 退货率与扩散模型崛起](#)
2. [02FASHN VTON v1.5：把试穿流程砍到一行](#)
3. [2.1为什么「无掩模」是一道分水岭](#)

4. [2.2部署与最低推理门槛](#)
5. [2.3两类服装图：model 与 flat-lay](#)
6. [03阿里 Tstars-Tryon 1.0：电商级多品类试衣的工业解法](#)
7. [3.1与 SHEIN / 京东 / 淘宝的功能定位差异](#)
8. [3.2从商品详情页到内容场](#)
9. [04技术路线演进：从 GANs 到 LDMs，从单图到视频](#)
10. [4.1GANs 的瓶颈：褶皱、光影与人体的「塑料感」](#)
11. [4.2LDMs 的解法：从噪声里重建](#)
12. [4.3视频换装：从静态到动态的最后一公里](#)
13. [0530% 退货率背后：AI 试穿如何变成可量化的转化收益](#)
14. [5.1从「减退货」到「增搭配」](#)
15. [5.2数字孪生分身的真实形态](#)
16. [5.3几个经常被忽略的隐性成本](#)
17. [06趋势与展望：原生多模态 API、视频换装和数字分身](#)
18. [6.11 · 原生多模态试穿 API 进入主流](#)
19. [6.22 · 视频换装从研究走向产品](#)
20. [6.33 · 数字孪生成为用户的随身资产](#)
21. [6.4回到无掩模这条主线](#)

Tech Briefing · 2026 / 07 / 01

## AI 虚拟换装：进入无掩模时代

FASHN VTON v1.5、阿里 Tstars-Tryon 商用大模型与扩散模型技术路线的最新进展

主题：AI · 虚拟换装类型：技术日报日期：2026-07-01

服装电商的退货率长期高悬在 30% 以上，而基于扩散模型的 AI 虚拟换装（TryOn）技术正以两条截然不同的路径同时向前：一边是 FASHN VTON v1.5 用「无掩模」思路把试穿从流水线里解放出来，另一边是阿里 Tstars-Tryon 1.0 把多品类 复杂搭配做到「随手拍也能识别」的工业级鲁棒性。本期资讯回顾这两条线 在 2026 年中的最新交汇。

### 本期要点

- FASHN VTON v1.5 是首个明确放弃人像分割掩模、直接在像素空间生成试穿图的开源项目
- 阿里 Tstars-Tryon 1.0、SHEIN、淘宝、京东的电商级落地推动 AI 试穿进入工业可用阶段
- 扩散模型（LDM）替代 GANs 成为虚拟换装的事实标准路线，OOTDiffusion、ViViD 持续刷新质量
- 30% 的服装退货率正被 AI 试穿从系统性挑战改写为可量化的转化收益
- OpenAI 推出的原生多模态试衣 API（延迟 <800ms）把行业竞争从「能不能做」推向「做得多快」

01

## 行业现状：30% 退货率与扩散模型崛起

服装品类长期是电商退货率的"重灾区"。多个电商平台和行业报告都引用过相近的 数字——服装类商品退货率常年在 **30% 以上**， 排在所有品类的最前列。其中"买家秀与卖家秀不一致"、"版型不合预期"、"细节差异大"几乎是每一条商品页评论里都会出现的痛点。 多年来行业尝试过两种传统方案：一是加大详情页的真人模特拍摄投入， 二是给 SKU 做更细的尺码表与参数标注；但这两条路都只能"减痛"， 没有真正解决用户在线上无法"试穿"的体验断层。

2024 到 2026 这一轮 AI 虚拟换装（Virtual Try-On，简称 TryOn）的浪潮， 终于让这件事从"减痛"走到了"可替代"。核心驱动来自两条几乎独立 但方向交汇的技术线：

| 路线 A：研究 / 开源社区                                                                 | 路线 B：商业平台                                                                     |
|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| OOTDiffusion、TryOnDiffusion、ViViD、FASHN VTON v1.5 等项目把质量推到工业可用， 同时不断下探算力与部署门槛。 | 阿里 Tstars-Tryon 1.0、SHEIN 的 AI 试衣、京东"虚拟衣间" 把同一批技术封装成面向消费者的产品， 直接接入商品详情页与购物链路。 |

本期技术日报将围绕这两条主线展开：第 2 节看 FASHN VTON v1.5 怎样用"无掩模"思路把试穿流程砍到极致；第 3 节看阿里 Tstars-Tryon 与 SHEIN/京东如何把同一思路工程化、产品化；第 4 节回顾扩散模型 替代 GANs 成为虚拟换装事实标准的技术原因；第 5 节拆解"30% 退货率" 这个商业痛点的具体数字与可验证的改善幅度；第 6 节回到未来—— 多模态原生 API、视频换装与个性化数字分身。

本期的判定原则  
本文所有"商业可用"判断都基于两条标准：(1) 项目处于可用版本号（非 0.x / 非测试分支）且提供公开 demo / 公开权重；(2) 至少有一家具有可比 MAU 量级的电商或内容平台在自家 APP 或小程序中默认开启该能力。

02

## FASHN VTON v1.5：把试穿流程砍到一行

在 Open Source 这一侧，2026 上半年最具范式意义的一次更新来自 **FASHN VTON v1.5**。它的关键宣称只有一句话——**不需要人像分割掩模，直接在像素空间里生成试穿结果**。听上去简单，但这意味着它打破了 2018 年以来几乎所有主流 TryOn 模型的"标准前处理"：人像分割、衣服区域抠图、掩模生成、姿态关键点 标定.....这些步骤的开销与技术门槛，把很多想自建试穿功能的中小团队 挡在了门外。

2.1

### 为什么「无掩模」是一道分水岭

传统流程需要在推理前把人像的"哪里是头部、哪里是躯干、哪里是手臂"、 "这件衣服的边界在哪里"显式标注出来，再把这些信息作为条件送进 生成模型。一旦遇到遮挡、叠穿、复杂姿势、模糊

边界，分割掩模 的失败就会直接污染最终的试穿图——这是工业场景里最常见的"翻车" 来源之一。

FASHN VTON v1.5 的设计选择是放弃显式掩模，让扩散模型在像素空间 直接对齐服装参考图与人体姿态。这意味着：

- 前处理被砍到几乎为零，普通开发者上传两张图（人物 + 服装）即可生成试穿；
- 对极端姿势和叠穿的鲁棒性显著提升，因为模型不再依赖前置掩模的精度；
- 上传"穿在人身上的模特图"也能作为服装参考——这贴合电商真实素材形态。

不需要分割掩模，直接就在像素空间里生成逼真的试穿结果。

FASHN VTON v1.5 官方示例文档，[fashn-AI/fashn-vton-1.5 GitHub](#)

## 2.2

### 部署与最低推理门槛

项目本身对自建算力非常友好：模型权重大约 2GB，仓库仅提供最小推理代码， GPU 可用时自动启用；CPU 也能跑（只是更慢）。对中小团队来说， 这个体量已经可以放进一台普通的双卡推理服务器里。

三类别支持

FASHN VTON v1.5 把服装分为 tops（上身）、bottoms（下装）、one-pieces（连衣裙 / 连体裤）三类，分别对应 三组训练得到的专属权重。推理时必须显式指明 `category=` 这一参数，否则会按"通用"分支处理，生成质量 通常不如专门类别稳。

## 2.3

### 两类服装图：model 与 flat-lay

工程上另一项值得注意的小细节：FASHN VTON v1.5 同时支持 **flat-lay**（白底平铺服装图）和 **model**（模特已经穿着服装的图）两种服装参考输入。 后者尤其贴合中小品牌的现实素材条件—— 很多时候根本没有准备"白底商品图"，只有从 Instagram 或小红书扒来的模特上身图。这一点把试穿功能从"商品详情页专用" 拓宽到了"社媒营销 / 二手买家秀还原"等多个工作流。

## 03

### 阿里 Tstars-Tryon 1.0：电商级多品类试衣的工业解法

开源一线的灵活和工程一线的可控，往往是两条相互平行的产品哲学。 阿里在 2026 年中段公开披露的 Tstars-Tryon 1.0是这条商业化主线最值得对照的样本。它的官方宣称里有两条最关键的约束：

- **多件服装复杂搭配**——同时换多件（连衣裙外套 + 腰带 + 鞋子），不是只换一件单品；

- **极端鲁棒性**——用户上传的照片不需要"摆拍"，随手拍的 侧脸、逆光、镜面自拍也能直接进入生成。

这两条任一项单独成立都已经达到行业第一梯队的水准，Tstars-Tryon 1.0 把它们放到了同一个模型上。这从根上区分了它和面向开发者的开源项目——后者通常为追求"最清爽的 demo 效果"做优化，前者必须直面几亿月活用户随手拍出的非标准人像。

3.1

与 SHEIN / 京东 / 淘宝的功能定位差异

横向看几家头部电商平台 2025-2026 上线的 AI 试穿功能，它们各自选择了不同的路径：

表 1 · 头部电商 AI 试穿能力对比（2025-2026）

| 平台    | 试穿能力                           | 差异化卖点                               |
|-------|--------------------------------|-------------------------------------|
| 淘宝    | 基于 Tstars-Tryon 系列模型的「AI 试穿助手」 | 用户上传一张背景简洁的全身正面照，秒级生成专属虚拟形象，再试穿商品   |
| 京东    | 「虚拟衣间」与数字孪生分身方案                | 强调用户数字资产可复用，跨店铺、跨品类持续试穿             |
| 阿里通义  | Tstars-Tryon 1.0 商用级多品类试衣大模型   | 多件搭配 + 极端鲁棒性，主打开放 API 与商家自接入        |
| SHEIN | 自研端到端试穿流水线                     | 覆盖全球 200+ 市场、与柔性供应链打通，试穿数据回流反哺选款与补货 |

从「能换」到「能搭」  
单件换装是上一代试穿模型的标配上限，多件搭配与数字孪生才是 2026 年电商级方案真正拉开差距的地方。"能搭"意味着用户可以一次验证"周末出去玩这一身到底协不协调"，也意味着平台可以借此把"试穿"内嵌到内容流（逛逛、小红书笔记等）。

3.2

从商品详情页到内容场

商业视角下更有意思的，是这些能力的"装配位置"。最早期上线时，AI 试穿往往被作为商品详情页的二级功能——藏在模特图旁的一个"试试这件"的折叠按钮里；但 2026 上半年，几个头部平台开始把它推向内容场——短视频里的穿搭挑战、直播间的连麦试穿、以及"逛逛"内容流里的"用你的虚拟形象一键搭"卡片。入口位置的不同会带来完全不同的留存与转化曲线，而内容场几乎必然成为下一阶段的主战场。

04

技术路线演进：从 GANs 到 LDMs，从单图到视频

把虚拟换装放到一个更长的技术时间轴上看，过去三年发生了几件事：**GANs 退场、潜在扩散模型 (LDM) 上位、单图→视频**。这三件事彼此并不孤立，它们共同回答了一个问题——"当用户上传的不是白底模特图，而是一段三秒种的走路视频时，模型还能生成自然的试穿吗？"

#### 4.1

### GANs 的瓶颈：褶皱、光影与人体的「塑料感」

2018–2022 年间，基于生成对抗网络 (GANs) 的虚拟换装方案占据了研究侧的主流。但即便在最好调参的状态下，GANs 也常常在三类细节上"翻车"：

1. **褶皱**：静态裁剪的服装在动态人体上呈现的拉伸 / 压缩不够自然；
2. **光影**：服装材质的高光、阴影与人体的方向不一致，产生"塑料贴图"般的违和感；
3. **手部与配件**：手部、首饰、包袋与服装边界的衔接常常出现断裂。

GANs 的"对抗训练"机制天然对上述细节敏感——判别器很容易把"光影 错位"判定为真实分布的一部分，因此这类瑕疵反而被掩盖在分布里。

#### 4.2

### LDMs 的解法：从噪声里重建

2023 年起，OOTDiffusion、TryOnDiffusion 等基于潜在扩散模型 (LDM) 的方案开始主导研究侧。它们的共同思路是：把服装试穿拆成两个可学习的物理过程——形变建模（衣服怎么贴合身体）和渲染合成（光线怎么照亮布料）——分别学习，最后在去噪 UNet 的自注意力层里完成对齐。

TryOnDiffusion 绕开了 SMPL + 网格变形的旧路径，直接用 UNet 编码器提取人体姿态热图和服装掩码作为条件输入，让网络自己学习"哪里该拉伸、哪里该压缩"。

OODDiffusion 论文与配套博客

LDM 路线有两个立刻显现的工程优势：(1) 可通过 `guidance scale` 显式调节生成图像中服装特征的强度（值越高服装越"像参考图"，值越低越像"自然穿着感"）；(2) 通过 `classifier-free guidance` 类技巧训练"有条件"与"无条件"两路分支，方便推理阶段插入风格 / 季节 / 用户偏好等多路控制信号。

#### 4.3

### 视频换装：从静态到动态的最后一公里

用户最终消耗的不是一张静态试穿图，而是一段短视频——TikTok、小红书、淘宝逛逛、京东发现好货都已经习惯把穿搭内容以 5–15 秒视频形式呈现。2025 年底登场的 **ViViD** (Video Virtual

Try-on using Diffusion Models) 代表着一次工程 跳跃：把图像扩散的能力延展到视频帧间一致性， 让一段走秀视频里的人物穿上任意服装， 而且前后帧不会出现"衣服突然变形"的跳变。

视频换装的两条主流路线

- 1. 逐帧独立生成 + 时间一致性损失（早期 ViViD 等）；
- 2. 视频潜在扩散一次成型（2026 年开始出现的 类 Sora 思路，把整个短视频视为一段噪声潜在表示）。

05

## 30% 退货率背后：AI 试穿如何变成可量化的转化收益

服装电商 30% 以上的退货率，长期是行业"无法直视"的核心 KPI。它不仅吞噬毛利，还通过逆向物流迅速消耗运营能力。多家平台披露，2025-2026 上半年接入 AI 试穿后，退货率出现了一位 数百分点 区间内的下降（行业报告通常给出 "约一成"这个保守估计），同时商品详情页的加购率出现了两位数的提升。

数字怎么读

退货率下降 "约一成" 指的是**相对值**：把基期退货率 从 30% 量级降到 27% 量级，绝对值下降 3 个百分点，对应数十亿 GMV 的电商意味着每年数亿元的运营成本节省。不同平台、不同品类的具体数字会有出入，但方向是一致的。

5.1

### 从「减退货」到「增搭配」

AI 试穿的价值已经不只是"减少不合适退货"。配合多件搭配能力， 它还带来几条过去想做但做不了的转化路径：

- **单品 → 搭配**：用户在试穿一件连衣裙时， 系统主动推荐配套腰带、外套、鞋子，一次结算 3-4 件；
- **SKU 反向被推荐**：从一句"我不知道自己适合什么" 的搜索，转变到"基于我的体型和过往订单，AI 主动为我配一身"；
- **直播连麦试用**：观众在直播间用 AI 试穿主播 推荐款，点击直达下单页。

5.2

### 数字孪生分身的真实形态

把"用户专属虚拟形象"作为资产长期沉淀下来，是 2026 年阿里 虚拟衣间为代表的商用方案最值得单独讲清楚的一件事。它在 工程上对应三步：

- 1. **形象生成**：用户上传一张背景简洁、四肢无遮挡 的全身正面照，模型生成用户的专属虚拟形象；

2. **参数估计**：自动估算身高、肩宽、腰围、臀围、惯用手侧等关键参数，存为用户档案的虚拟属性；
3. **跨店铺复用**：用户进入任意接入虚拟衣间协议的 店铺，虚拟形象与参数都自动可用。

支持"多件服装复杂搭配"并具备"极端鲁棒性"，可处理"随手拍"的野生图片。

阿里 Tstars-Tryon 系列公开介绍，面向多品类与数字分身

一旦数字分身真正成为跨店铺可用的"用户资产"，它的角色就不仅限于 试衣——它将和会员体系、推荐算法形成正反馈：模型从用户的穿搭选择里学到偏好，偏好又反过来决定下次推送什么。

5.3

## 几个经常被忽略的隐性成本

在大力推进 AI 试穿的同时，也要提到几个容易被忽视的隐性成本，它们决定了一个方案能否真正从 demo 转化为长期能力：

- **合规与肖像**：用户上传个人照片是否构成"生物识别 数据"，在哪里存储、是否脱敏、是否跨境传输，都直接影响方案设计；
- **商品图版权**：以"他人发布的模特上身图"作为 服装参考，是否侵犯相应摄影作品 / 商家版权——开源 vs 商业版的边界反而在这里发生冲突；
- **算力与单位成本**：一张图片 vs 一段 10 秒视频，推理算力消耗差距可达两个数量级，规模化的电商需要持续压单位成本；
- **结果一致性**：同一商品不同次生成的造型是否一致，关系到商家页面能否稳定呈现"我的虚拟形象穿上这件商品"的可信度。

06

## 趋势与展望：原生多模态 API、视频换装和数字分身

把 2026 年中的节点与更远的轨迹放在一起看，三件事几乎确定会在 接下来 18 个月里同时发生：

6.1

### 1 · 原生多模态试穿 API 进入主流

OpenAI 在 2026 年 1 月发布的原生多模态试穿 API 把端到端 延迟压到 800 毫秒以下，并允许用自然语言描述服装风格（"慵懒的法式通勤感"、"赛博朋克金属配件"），模型实时生成穿着效果。

这意味着 TryOn 不再是一个"独立功能", 而是被并入了更广义的多模态 生成接口——图片、视频、3D、试穿共享同一套输入约定与基础设施。

从「能不能做」到「做得多快」

当延迟被压到秒级以下, 行业的竞争点会从模型效果转向工程效果——单位算力成本、冷启动一致性、安全过滤、合规审计。小团队很难再靠"我模型跑得比你快"获得长期优势。

6.2

## 2 · 视频换装从研究走向产品

ViViD 等项目证明: 基于扩散模型的视频虚拟试穿, 在帧间一致性上已经达到 **可以直视** 的水准。接下来一年的关键工程问题是把推理速度压到可商用区间 ——一段 10 秒视频通常需要数十秒生成时间, 这对实时直播连麦来说依然太慢。主流路径是 (a) 把扩散步数砍到 5-10 步的 Consistency Model / Latent Consistency Model 变体, (b) 借助视频 Tokenizer 把 10 秒视频压缩成几百个 token 再分块生成。

6.3

## 3 · 数字孪生成为用户的随身资产

当用户的虚拟形象跨店铺、跨平台、跨年度都保持稳定, 它就具备了真实的"用户资产"属性——可被用来试穿、可被用来生成内容、可被推荐算法直接调用。这会反过来重塑三件事:

1. **推荐系统**: 从基于 SKU 属性的协同过滤, 转向基于"我的虚拟形象穿上身"的相似度;
2. **内容生产**: 用户自拍替换为"数字分身看秀 / 数字分身约会"等 UGC 模板;
3. **隐私边界**: 数字分身是用户私密资产, 离开用户授权, 平台不得用于训练行业"通用分身"。

6.4

## 回到无掩模这条主线

把"无掩模"放回到最后一起回顾, 是因为它会左右上述三件事的工程取舍。当模型不再依赖显式分割时, 单位算力会被进一步压缩、鲁棒性会进一步上升、创意工作流(叠穿、混搭、二次元化)会被系统性打开。这一类基础性简化往往是研究侧最容易被低估、却最终决定商业侧能不能"全民铺开"的能力——这正是 FASHN VTON 系列之于本期的最大意义。

结论: 从'能用'到'全民铺开'还差什么

用一句话概括本期想留下的判断: 2026 年中的 AI 虚拟换装, 已经走完"实验室到头部电商试点"这一程; 接下来的 12-18 个月, 决定胜负的将是**无掩模思路带来的工程门槛下降**、以及**多件搭配 + 数字分身**共同支撑的产品厚度。

把这条线压回到具体可执行的几条评估项:

- 把 TryOn 当成产品功能时，优先选择支持'无掩模'或多类别的工程方案，前者降低接入成本，后者降低维护成本
- 评估一个商用方案时，关注三项硬指标：多件搭配能力、跨店铺数字分身、可处理的姿态复杂度
- 对开源方案（FASHN VTON v1.5 / OOTDiffusion / ViViD）保持持续关注，但接入时要预留商业 license 与合规审核空间
- 对团队内部，把'单位算力成本 × 用户行为分布'作为长期监控指标，延迟 800ms 是一个清晰的工程阈值