第1章 识图引擎：二手电商全能估价引擎

上传一幅照片，让 AI 看懂物品、查清行情、报出价格、写好文案

本章导读

家里那副吃灰的二手耳机，到底能卖多少钱？这几乎是每个耳机发烧友都遇到过的小难题。本章将带着大家，用**一句话需求、全程零代码**的方式，做出一个真正能用的**二手物品全能估价引擎**：用户上传一幅物品照片，它就能看懂这是什么、查到当前二手行情、给出合理的价格区间，并顺手写好一段可以直接发布的转卖文案。

更重要的是，这一章会把贯穿全书的两个**看家本领**交到读者手里——把需求“说清楚”的**七要素指令法**，以及多模态智能体最核心的工作范式**模态接力**。读完本章，读者不仅能做出一个估价工具，还能掌握一套可以无限复用的开发方法。

1.1 从一个生活难题说起

先请读者设想一个再普通不过的周末：你整理房间，翻出一副用了一年多、成色还不错的降噪耳机，想挂到闲鱼上换点零花钱。这时，第一个让人犯难的问题就来了：**它到底值多少钱？**

价格定高了，没人问津；价格定低了，自己心疼。于是你打开购物平台查新品价格，再去二手平台一个个比对同款成交记录，还要斟酌自己这副耳机的磨损程度该打几折。一通操作下来，半小时过去了，最后写商品描述时又卡壳：成色怎么如实又好听地说？卖点怎么提炼？这套流程，笔者称之为**二手转卖的隐形门槛**——不难，但烦琐到很多人宁愿让闲置物品继续落灰。

笔者观察

二手转卖的真正成本，从来不是定价本身，而是“识别→查行情→估价→写文案”这一连串琐碎动作累加起来的精力消耗。任何一步卡住，转卖意愿就会断在那里。

那么，能不能把这一连串动作交给 AI，让它一次性全部办好？答案是肯定的。本章要做的，

正是这样一个智能体：用户只需要**上传一幅照片**，剩下的识别、查价、估价、写文案，全部由它自动完成。先来看一眼最终要做出来的智能体长什么样（见图 1–1），有了这个“剧透”，后面的每一步就都有了奔头。



图 1–1 估价引擎的部分成品页面：识别结果卡与建议估价区间卡

用户上传一幅耳机照片，它会清清楚楚地告诉你：品类是耳机、品牌是 BOSE、成色评分为 7 分，并给出捡漏价、合理价、最低价三档参考。这就是本章的目标。接下来，笔者将带领读者一步步把它做出来。

1.2 先认识主角：扣子编程与“模态接力”

1.2.1 什么是扣子编程（Vibe Coding）

看到“开发”“代码”这些词，大家别紧张。本书自始至终使用的工具叫**扣子**（Vibe Coding），它最大的特点只有一句话：**你用大白话把需求说清楚，它替你把完整的代码全部写好。**

传统开发中，做一个能上传图片、调用AI、联网搜索、渲染出漂亮界面的应用，至少要懂前端、后端、接口调用三套技术，没有几周下不来。而在扣子编程里，读者要做的是**把一段需求描述**

清楚——它会自己规划任务、初始化工程、编写前后端代码、自测验证，最后交给你一个能直接预览、一键上线的成品。这就是本书反复强调的**零代码**的真正含义：门槛不在于你会不会写代码，而在于你能不能把想要的东西讲明白。

一句话理解

过去：想法 → 学技术 → 写代码 → 调试 → 成品（以周计）

现在：想法 → 说清楚需求 → 成品（以分钟计）

1.2.2　全书的核心范式：“模态接力”

所谓 AI 多模态，指的是 AI 能同时处理图像、文字、语音、视频等多种信息形态。而**模态接力**，则是指把一种能力的产出当作下一种能力的输入，像接力赛一样一棒接一棒地传递下去。

本章的估价引擎，就是一次最朴素的模态接力：**图像理解**先把照片“翻译”成结构化文字（品类、品牌、成色），再将这段文字“喂”给**联网检索**去查行情，最后所有信息汇总给**语言模型**进行综合估价、文案生成。三种能力首尾相接，缺一不可——这正是多模态智能体区别于单一工具的地方：它不是几个功能的拼盘，而是一条有**上下文、会传递的生产线**。估价引擎的“模态接力”链路（也叫模态接力链）如表 1–1 所示。

表 1–1　估价引擎的“模态接力”链路

接力棒	调用的能力	它做的事	交给下一棒的产出
第 1 棒	视觉理解（LLM Vision）	看懂照片：识别品类、品牌、型号、可见成色	一段结构化的物品描述
第 2 棒	联网检索（Web Search）	拿着物品描述在闲鱼 / 转转中查近期成交价	当前二手市场行情区间
第 3 棒	综合生成（LLM）	结合成色与行情，给出估价区间并撰写文案	三张结果卡片（最终交付）

把表 1–1 记牢，后续每一章解锁的能力（文生图、图生视频、语音合成……），本质上都是在这条接力链上增添新的“接力棒”。

1.3　七要素指令法：把需求“说清楚”的模板

既然零代码开发的关键是“把需求说清楚”，那么怎样才算“清楚”？凭感觉随便写几句，AI 给出的结果往往南辕北辙。为此，笔者总结了一套可以套用在任何项目上的模板——**七要素指令法**。它把一段需求拆成七个必答项，只要逐项填好，AI 就能精准识别你的意图。以本章的估价引擎为例，七要素指令法详解如表 1–2 所示。

表 1–2 七要素指令法详解（以本章的估价引擎为例）

序	要素	作用	本例怎么写
1	目标	一句话讲清要做什么	做一个“二手物品全能估价”智能体
2	输入	说明用户提供什么	用户上传一幅物品照片
3	处理流程	拆解 AI 要执行的步骤	①视觉识别物品；②联网查行情；③综合估价
4	输出结构	固定产出的内容与格式	固定三块卡片：识别结果 / 估价区间 / 转卖文案
5	界面交互	描述长什么样、怎么用	顶部大号上传区，下方三卡纵向；估价用大字；文案带复制按钮
6	边界约束	规定不能做、不确定怎么办	识别不确定时主动提示补拍细节，不要瞎猜
7	风格基调	界定语气与视觉气质	专业、可信赖的“鉴定室”质感（本例由 AI 自主把握）

★ 为什么是表 1–2 所示的这七个要素

大家可以这样记：前四要素（目标、输入、处理流程、输出结构）回答“做什么”，决定智能体的骨架；后三要素（界面交互、边界约束、风格基调）回答“怎么做得好”，决定智能体的体验与可靠性。一句需求描述只要七要素齐全，就几乎不会让 AI 跑偏。

这套方法会贯穿全书每一个案例。笔者建议大家准备一张便签，把这七个要素抄下来贴在显示器旁——往后无论做什么智能体，先按七要素把需求过一遍，开发就成功了一大半。

1.4 动手开发：十分钟造一个估价引擎

本节动手开发一个估价引擎，全程只需十分钟。

1.4.1 写下你的第一条指令

打开扣子编程，新建一个项目（笔者把它命名为“二手物品全能估价”，作为需求写入提示词）。接下来，把上一节用七要素整理好的需求，原原本本地写进对话框。下面是笔者实际输入的完整指令，读者可以直接对照使用：

做一个「二手物品全能估价」智能体。用户上传一张物品照片，你就要完成估价并给出转卖建议。

处理流程：①调用视觉理解能力识别物品的品类、品牌、型号、可见成色（划痕 / 磨损 / 配件齐全度）；②联网检索该型号当前二手行情（闲鱼 / 转转同款近期成交价区间）；③综合成色给出本机的合理估价区间。

输出固定为三块卡片：1）识别结果（品类 · 品牌 · 型号 · 成色评分 1–10）；2）建议估价区间（最低价 / 合理价 / 捡漏价，并注明依据）；3）一键生成的二手转卖文案（含卖点、成色实话实说、可直接复制发布）。

界面：顶部一个大号图片上传区，下方结果用三张卡片纵向展示，估价区间用醒目大字，文案块右上角带「复制」按钮。

识别不确定时主动提示用户补拍细节图，不要瞎猜。

提　示

在作者给出的提示词中，需要保留「二手物品全能估价」中的标点、一“张”图片中的量词的习惯用法，不能改成“二手物品全能估价”或者“一幅图片”。

指令发送之后，扣子编程便接管了一切。页面左侧实时展示思考与操作过程，右侧同步预览正在生成出来的界面。图 1–2 所示是输入需求后的扣子编程页面：左侧是我们输入的指令，右侧已经渲染出一个标着“二手物品全能估价”的、带大号上传区的页面雏形。



图 1–2　在扣子编程左侧输入需求指令，右侧实时预览界面雏形

★ 别慌

看不懂页面左侧的英文和代码，完全不用慌张。那是 AI 的“工作草稿”，我们只需要像验收装修一样盯着页面右侧的预览区，看效果对不对即可。这正是零代码开发的底气所在。



1.4.2 看 AI 如何规划与搭建

一个有经验的工程师接到开发任务，不会一上来就敲代码，而是先列计划。扣子编程也是如此。它会先把大任务拆成若干小任务，逐项推进。如图 1–3 所示，页面左侧就是它自动生成的“更新计划”，页面右侧则是它初始化出来的部分项目文件结构。

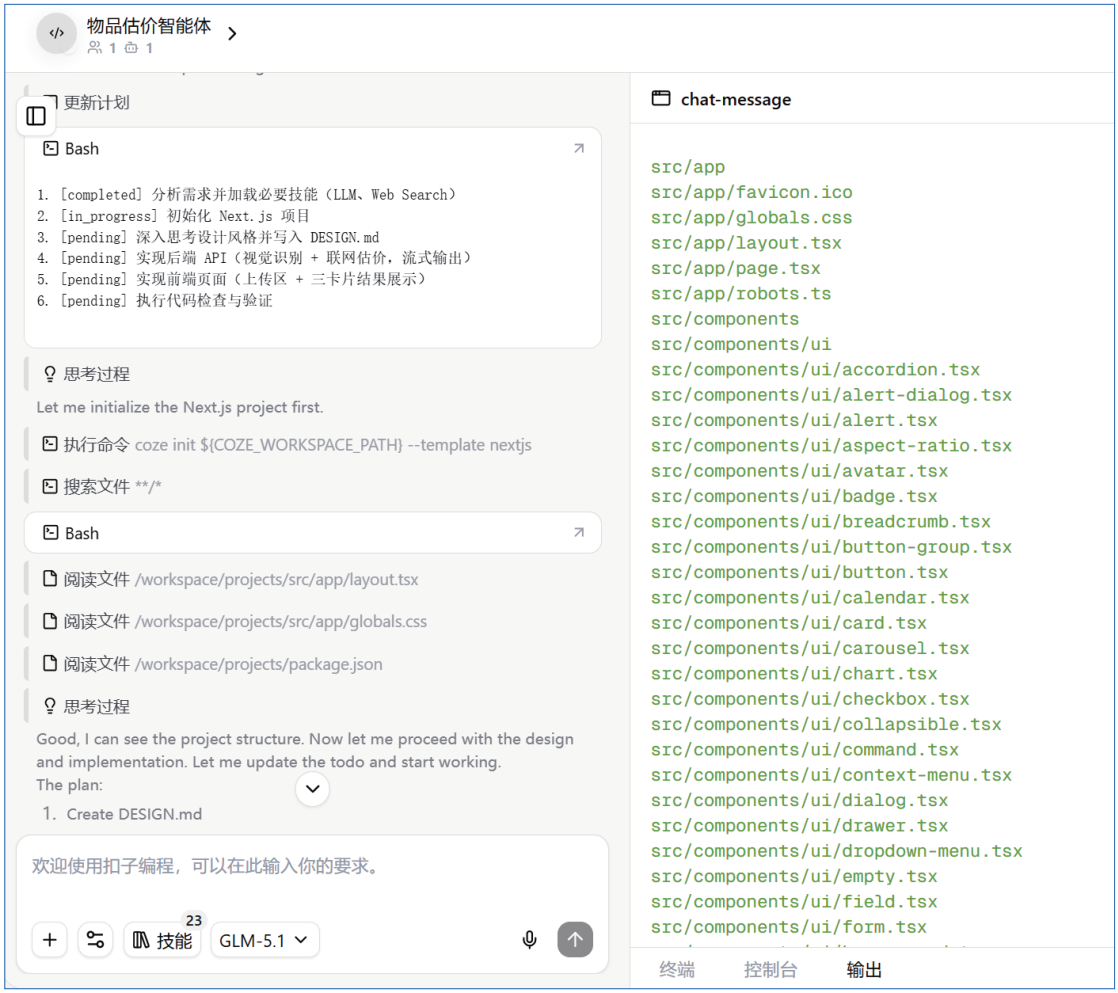


图 1–3 AI 自动生成的任务计划（左）与初始化的项目文件结构（右）

我们把图 1–3 所示页面左侧的计划清单翻译成自然语言，就能明白 AI 在替我们计划哪些事了，具体如表 1–3 所示。

表 1–3　AI 自动拆解的开发任务清单

步骤	AI 的任务	通俗解释
①	分析需求并加载必要技能（LLM、Web Search）	先想清楚要用到哪些 AI 本领
②	初始化 Next.js 项目	搭好应用的地基框架
③	深入设计风格并写入 DESIGN.md	定下整体视觉风格基调
④	实现后端 API（视觉识别 + 联网估价，流式输出）	搭建幕后的“智能流水线”
⑤	实现前端页面（上传区 + 三卡片结果展示）	做出用户看得见、点得动的界面
⑥	执行代码检查与验证	自己检查作业，确保能跑通

注意第 ③ 步主动撰写设计文档和第 ④ 步中的“流式输出”——这些都是 AI 自发补上的工程细节，笔者的指令中甚至没有明说。这说明一个好工具不只是“听话”，更会“补位”。下一节，笔者就带大家看懂这条流水线背后的门道。

1.5　拆解技术方案：三步模态接力是怎么跑起来的

这一节稍微“掀开引擎盖”看一眼内部，把原理讲透——理解了原理，往后调优才有方向。整套方案可以归结为一句话：**后端把三步模态接力串成一条流水线，再用流水的方式把进度实时推给前端**。

1.5.1　三步分别用了什么

估价引擎三步流水线的技术构成如表 1–4 所示。

表 1–4　估价引擎三步流水线的技术构成

步骤	使用的技术	作用与要点
视觉识别	doubao–seed–2–0–pro 视觉模型	读懂照片，输出品类、品牌、型号与 1~10 分的成色评分
联网估价	Web Search 联网检索	搜索闲鱼 / 转转等平台同款近期成交价，把“行情”实时拉进来
综合生成	语言模型（LLM）综合	把成色与行情糅在一起，算出三档价格并写出转卖文案

联网估价这一步至关重要，如果跳过它，AI 只能凭训练时的“老记忆”瞎猜价格，而二手行情是天天变化的。**接入实时联网检索，估价才有了“今天的”依据**，这也是这个工具敢叫“全能估价”的底气。

1.5.2 为什么要做成“流式”体验

三步走完需要几秒到十几秒。如果界面在这段时间里一片空白、毫无反应，用户会以为程序卡死了。为此，采用了一种叫 SSE（Server-Sent Events，服务器推送事件）的技术，让后端每完成一步就立刻把进度推送到前端，在界面上实时显示**识别中→搜索中→估价中**的三步进度。如图 1-4 所示，抓拍到的正是它正在综合估价并生成文案的那一刻。



图 1-4 SSE 流式进度反馈：界面实时显示“正在综合估价并生成文案……”

> **★ 体验即专业**
>
> 进度反馈看似只是一个小动画，实则是产品专业度的分水岭。它把“漫长的等待”变成了“看得见的进展”，用户的焦虑被安抚，信任随之建立。

1.6 实测验收：给一副 BOSE 耳机估价

代码写完不算完，能用、好用才算数。本节笔者将按照“先验接口、再验体验”的顺序，带大家完整跑一遍验收。

1.6.1　接口冒烟测试：一个“报错”反而证明它对了

“冒烟测试”是工程行话，意思是先做一个最小化的快速测试，看看会不会“冒烟”（出大问题）。扣子编程用一幅 1×1 像素的极小图片去调接口，结果如图 1–5 所示，系统返回了“图片尺寸过小”的错误提示。



图 1–5　接口冒烟测试：1×1 像素的测试图触发了“图片尺寸过小”提示

> **★ 如何看待这个“报错”**
>
> 这个报错恰恰是好消息。它证明：接口成功收到了图片、正确启动了识别流程，并且在发现图片太小、无法识别时，把错误如实地抛了出来，而不是乱猜一个结果。同时，进度事件也正常工作。一个会“老实认错”的系统，远比一个“什么都敢答”的系统可靠。

1.6.2　上传真实照片，看它怎么干活

笔者随手拍了一副放在桌上的二手 BOSE 头戴耳机的照片，如图 1–6 所示。照片中的光线一般，背景也有些杂乱——刻意如此，就是要看看“二手物品全能估价”智能体在真实场景下的表现，而不是用“摆拍好的样板图”自欺欺人。



图 1–6 用于实测的二手 BOSE 头戴耳机（真实随手拍）

把这幅照片拖进上传区，三步进度依次走完，几秒钟后，三张结果卡片（识别结果、建议估价区间、转卖文案）便依次呈现。

1.6.3 解读三张结果卡片

识别结果卡与建议估价区间卡如图 1–7 所示。最上方有一条醒目的橙色提示——“识别信息不确定，请补拍耳机完整的头梁标识、产品铭牌或包装盒以确认具体型号”。这正是我们在七要素里写下的“不确定不瞎猜”约束起了作用：**它识别出了耳机品牌 BOSE，但因为识别不出型号铭牌，便诚实地把型号标为“未知”**，并主动引导用户补拍，而不是随便编一个型号糊弄过去。



图 1–7 识别结果卡与建议估价区间卡（含“识别信息不确定”主动提示）

识别结果卡给出了：品类“耳机”、品牌“BOSE”、成色评分“7 分”，并附一句“外观有轻微使用痕迹，耳罩无明显破损，未见配件”。估价区间卡则用醒目大字给出三档参考价，并在下方注明估价依据——“因 BOSE 该型号暂无明确行情，参考主流头戴式耳机二手价估算，7 成新仅轻微使用痕迹、无配件，价格相应下修”。估价区间三档定价的含义如表 1–5 所示。

表 1–5　估价区间三档定价的含义

档位	本例	适用场景
捡漏价	¥120	急用钱、想最快出手时的“秒杀价”，买家视角的“捡漏”
合理价	¥200	综合成色与行情后最推荐的挂牌价，买卖双方都易接受
最低价	¥150	可接受的底线价，议价时心里有数、不至于贱卖

转卖文案卡如图 1–8 所示，右上角带一个“复制”按钮，单击一下就能拿走文案直接发布。它写道：“BOSE 二手耳机，成色 7 新，正常使用痕迹，耳罩完好无破损，功能全正常，裸机无配件……二手物品售出非功能问题不退不换哦～”短短几句，卖点、成色实话、交易提醒一应俱全，比许多人自己花半天写出来的还要周到。

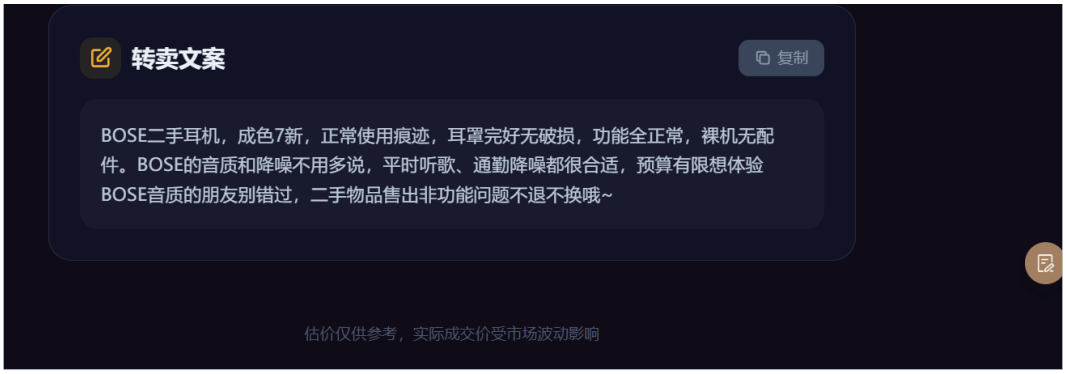


图 1–8　一键生成的转卖文案卡，右上角带“复制”按钮

★ 回看七要素

至此读者不妨回头对照表 1–2：从三张卡片的输出结构，到大字估价、文案复制按钮的界面要求，再到“不确定就提示补拍”的边界约束——我们当初写下的每一条需求，都在成品里被一一兑现。这就是七要素指令法的威力。

1.7 本章小结与延伸

用一句话总结本章内容：我们靠**一段清楚的需求描述**，零代码做出了一个能识图、查价、估价、写文案的二手估价引擎。然而，比成品更值钱的是，大家随手装进了三个可以反复使用的工具。

> ★ **本章三个核心收获**
>
> （1）**七要素指令法**——根据目标、输入、处理流程、输出结构、界面交互、边界约束、风格基调，把任何需求过一遍，开发就成功一大半。
>
> （2）**模态接力**——上一种能力的产出，喂给下一种能力。这是贯穿全书的生产线思维。
>
> （3）**不确定性处理**——宁可说“不知道、请补拍”，也不瞎猜。可靠，是 AI 产品的生命线。

商用延伸：这台引擎还能怎么改造

本章成品不只是一个小玩具，把需求稍作调整，它就能长成真正能赚钱的工具：

- **二手回收商：**把成色评分细则改成自家收货标准，店员拍照即报收购价，统一定价、提升效率。
- **电商运营：**把“转卖文案”换成“商品卖点文案”，上传产品图即出一套标题与卖点，告别绞尽脑汁写详情。
- **闲置社群 / 个人：**直接封装成网页或小程序分享出去，朋友拍照即得估价，顺手就是一个有流量的小服务。

本章我们让 AI 学会了**识图**。下一章，笔者将带大家走向接力链的下一棒——**生成图片**：只用一句话，生成专业级的营销主视觉海报。识图与出图打通，多模态创作的大门才算真正向大家敞开。

第2章 出图引擎：实体店开店物料一键生成

几句描述加一幅产品照，让 AI 产出门头、产品美图与开业海报一整套视觉物料

本章导读

上一章我们让 AI 学会了**看懂一幅图**——识别物品、查清行情、报出价格。这一章，笔者要带大家走向接力链上更激动人心的一棒：**生成一幅图**。我们的案例升级为一个能落地经营的小工具——**实体店开店物料智能体**：店主只用几句话介绍自己的小店，再上传一幅招牌产品的照片，它就能一口气产出门头招牌效果图、招牌产品展示美图，以及一幅带文字的开业促销海报。

本章会把两项全新的看家本领交到读者手里——用文字凭空造图的**文生图**，以及让 AI 照着照片再创作的**图生图**。同时，我们还会解决一个让无数人头疼的老问题：**怎样让 AI 生成的图一个字都不出错**。读完本章，读者就能打通多模态创作中“识图”与“出图”的任督二脉。

2.1 从一个开店难题说起

先请读者设想一个场景。你的朋友小林盘下了小区门口一个小铺面，准备开一家社区咖啡馆。装修快好了，咖啡豆也选好了，可临到开业，一连串“物料”却把她难住了——**门头招牌该长什么样？产品照片怎么拍才有质感？开业海报上的店名、日期、优惠又该怎么排？**

这些活儿，样样都得花钱找设计师，还得来回沟通改稿。一套门头效果图加几幅产品图、一幅开业海报，少则三五天，多则一两周，预算轻松上千元。对于一个还没开业、处处要花钱的小店主来说，这是一笔不小的负担，更是一道让人迟迟不敢迈步的**隐形门槛**。

笔者观察

小本经营的真正难处，往往不在咖啡好不好喝，而在“开张前那一堆看似不起眼、却样样要花钱和时间的物料”。设计门槛一旦卡住，开业的心气儿就容易泄掉一半。

那么，能不能让小林只花几分钟、不花一分设计费，就拿到一整套像模像样的开店物料？答案是能。本章要做的，正是这样一个智能体：店主**用几句话描述小店、上传一幅招牌产品照**，剩下的风格设定、门头设计、产品出图、海报排版，全部交给它一键完成。我们先看一眼最终要做出的成品长什么样，如图 2–1 所示。



图 2–1　开店物料成品速览：AI 自动排版生成的带字开业海报

一幅构图专业、店名醒目、日期与优惠一字不差、连二维码位置都留好的开业海报，这就是本章的目标之一。而它仅仅是 AI 一次性产出的四块物料中的一块。接下来，笔者将带领读者把这套“开店物料工坊”一步步造出来。

2.2　先认识新本领：文生图与图生图

在动手之前，先认识本章的两位新主角——文生图（Text–to–Image）与图生图

（Image-to-Image）。它们都属于 AI 的**图像生成**能力，但分工不同。理解了它们的区别，后面用起来才得心应手。

2.2.1　文生图：一句话造一幅图

文生图，顾名思义，就是你提供一段文字描述，AI 凭空“画”出对应的图片。比如，你写“一家复古木质风格的社区咖啡馆门头，暖色灯光，招牌上写着‘拐角’二字”，AI 就能生成一幅符合描述的门头效果图。本章讲解的**门头招牌效果图**和**开业海报**，都是用文生图造出来的。

文生图最考验人的是描述的精准性。描述越具体（风格、配色、构图、氛围、要出现的文字），出图就越贴近预期。这也正是为什么要先让 AI 想清楚“风格关键词”和“主辅色”——它们最终都会变成喂给文生图的精确指令。

2.2.2　图生图：让 AI“照着画”

图生图则更进一步：你不只给文字，还给一幅**参考图**，AI 会在保留参考图主体的基础上重新创作。本章的**招牌产品展示图**就用到了它——店主上传的产品原照可能光线一般、背景杂乱，图生图能在**保留这杯咖啡真实卖相**的前提下，把它“搬”进窗边洒满阳光的桌面或是温馨的店内场景，生成两幅精美的展示图。

一句话理解

文生图：无中生有，凭文字描述创造出全新的图（门头、海报）。

图生图：有中再造，以你的照片为底，换个场景与质感重新出图（产品美图）。

2.2.3　接力链的延伸：从“看懂”到“造出”

还记得第 1 章那条“模态接力”链路吗？本章的智能体，正是在它的基础上又接上了新的一棒。这一次的接力，比上一章更长，也更精彩，如表 2-1 所示。

表 2-1　本章开店物料智能体的“模态接力”链路

接力棒	调用的能力	它做的事	交给下一棒的产出
第 1 棒	视觉理解（LLM Vision）	看懂上传的招牌产品：识别品类、卖相、特色	一段结构化的产品描述
第 2 棒	风格决策（LLM）	结合店铺描述与产品，定下视觉风格、主辅色、宣传语	一张完整的“门店风格卡”
第 3 棒	文生图（Seedream 5.0）	据风格卡造出门头招牌图、带字开业海报	两幅全新的视觉物料图
第 4 棒	图生图（Seedream 5.0）	以产品原照为底，生成两个场景的展示美图	保真又精美的产品图

对照第 1 章的三棒接力会发现：**第 1 棒“视觉理解”是两章共用的起点**，差别在后面——上一章接的是联网检索——“查行情、算价格”，这一章接的是风格决策——“定风格、造图片”。这正是模态接力的魅力：**同样一项能力的产出，接到不同的下一棒，就能长出完全不同的应用**。把这条链记牢，本章就成功了一半。

2.3 用七要素重写需求

第 1 章中的**七要素指令法**，是可以无限复用的。本节笔者就带读者再走一遍流程，把开店物料这个新需求，逐项填进七要素指令法的模板里，如表 2–2 所示。读者会越来越体会到：**只要七项填齐，再复杂的智能体需求，也能说得清清楚楚**。

表 2–2 七要素详解（以开店物料智能体为例）

序	要素	本例怎么写
1	目标	做一个“实体店开店物料”智能体
2	输入	用户用几句话介绍小店，并上传一幅招牌产品照片
3	处理流程	①视觉识别招牌产品；②定风格主辅色与宣传语；③文生图 + 图生图产出物料
4	输出结构	固定四块卡片：风格卡 / 门头图 / 产品图（两场景）/ 开业海报
5	界面交互	顶部店铺描述框 + 产品上传区，下方四卡纵向；海报与门头带“下载”“再生成”
6	边界约束	门头与海报文字须与输入完全一致、无错别字；产品图保真不夸大；信息不足主动追问，不编造地址或优惠
7	风格基调	温暖务实的实体店气质（本例由 AI 把握，落地为琥珀暖橙 + 石板深灰 + 米白）

★ 边界约束这次为何格外重要

本章的第 6 项边界约束，比上一章更“吃紧”。因为门头和海报上要**直接出现店名、开业日期、优惠文字**——这些字只要错一个，物料就报废了。所以我们必须在指令里把“文字与输入完全一致、无错别字”这条红线固定。2.5 节会专门讲，AI 是用什么巧办法兜住这条红线的。

2.4 动手开发：搭一个开店物料工坊

本节动手搭一个开店物料工坊。

2.4.1 写下你的第一条指令

打开扣子编程，新建项目（笔者命名为“实体店物料生成网站”）。把上一节用七要素整理好的需求，原原本本写进对话框，如图 2–2 所示。

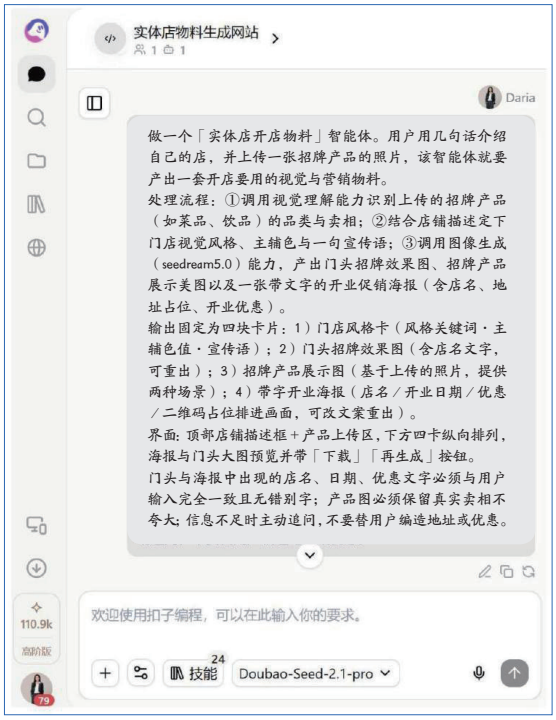


图 2–2　在扣子编程中输入按七要素整理的开店物料需求指令

下面是笔者实际输入的完整指令，读者可直接对照使用：

★ 本章实际输入的需求指令

做一个「实体店开店物料」智能体。用户用几句话介绍自己的店，并上传一张招牌产品的照片，该智能体就要产出一套开店要用的视觉与营销物料。

处理流程：①调用视觉理解能力识别上传的招牌产品（如菜品、饮品）的品类与卖相；②结合店铺描述定下门店视觉风格、主辅色与一句宣传语；③调用图像生成（seedream5.0）能力，产出门头招牌效果图、招牌产品展示美图以及一张带文字的开业促销海报（含店名、地址占位、开业优惠）。

输出固定为四块卡片：①门店风格卡（风格关键词·主辅色值·宣传语）；②门头招牌效果图（含店名文字，可重出）；③招牌产品展示图（基于上传的照片，提供两种场景）；④带字开业海报（店名／开业日期／优惠／二维码占位排进画面，可改文案重出）。

界面：顶部店铺描述框＋产品上传区，下方四卡纵向排列，海报与门头大图预览并带「下载」「再生成」按钮。

门头与海报中出现的店名、日期、优惠文字必须与用户输入完全一致且无错别字；产品图必须保留真实卖相不夸大；信息不足时主动追问，不要替用户编造地址或优惠。

> ★ 别慌
>
> 和第 1 章一样，发送指令之后，扣子编程就接管了全部开发工作。页面左侧那些飞速滚动的英文与代码，是 AI 的“工作草稿”，我们完全不用去读。只需像验收装修一样，盯着页面右侧预览对话框看效果对不对即可。



2.4.2 看 AI 如何规划与搭建

扣子编程依旧先列计划，再分步施工。它把这个比上一章更复杂的需求，拆成了一条清晰的流水线，并自动补齐了许多笔者没提到的工程细节。图 2–3 所示是扣子编程宣告“已完成”时给出的功能清单——一幅图就把整个项目的骨架交代得明明白白。

实现功能

完整处理流程：用户输入店铺描述 + 上传招牌产品照片 → AI 自动产出全套开店物料

5 个后端 API：

1. `/api/upload` — 产品图片上传至对象存储（S3Storage）
2. `/api/analyze` — LLM 视觉理解产品 + 风格定位（doubao-seed-2-0-pro 多模态模型）
3. `/api/generate-signboard` — 门头招牌效果图生成（SeeDream 5.0）
4. `/api/generate-product` — 产品展示图生成（2 个场景，基于上传图片做图生图）
5. `/api/generate-poster` — 带字开业海报生成（店名 / 日期 / 优惠 / 二维码占位）

前端四卡片输出：

6. **门店风格卡** — 风格关键词 · 主辅色值(色块 +Hex) · 宣传语 · 产品识别结果
7. **门头招牌效果图** — 含店名文字，大图预览 + 下载 / 再生成按钮
8. **招牌产品展示图** — 2 种场景对比展示，各自可下载
9. **带字开业海报** — 店名 / 开业日期 / 优惠 / 二维码占位，可再生成改文案

设计规范：琥珀暖橙主色 + 石板深灰辅色 + 米白背景，温暖务实的实体店气质

约束保障

- 门头 / 海报中的店名、日期、优惠文字与用户输入一致（prompt 中用双引号包裹确保文字准确）
- 产品图保留真实卖相，prompt 明确 "preserving real appearance without exaggeration"
- 信息不足时接口返回明确的验证错误提示，不替用户编造

所有检查都通过了。现在我来创建 `AGENTS.md`：

All checks pass now! The logs show that:

1. All APIs work correctly
2. The analyze endpoint successfully returned style data

图 2–3 AI 完成搭建后的功能清单：五个后端接口、四块前端卡片与约束保障

把图 2–3 所示的清单翻译成自然语言，如表 2–3 所示。这一次最值得留意的，是它把任务拆成了**五个各司其职的后端接口**——这正是本章技术方案的核心。

表 2–3　AI 自动拆解的 5 个后端接口

接口	它负责的事	通俗解释
/api/upload	把产品照片上传到对象存储	先把店主的照片保存到云端，得到一个网址
/api/analyze	LLM 视觉理解产品 + 定风格主辅色	看懂照片、定下整店的视觉调性
/api/generate–signboard	文生图生成门头招牌效果图	凭风格描述造出门头大图
/api/generate–product	图生图生成产品展示图（两场景）	照着原照，换两个场景重新出图
/api/generate–poster	文生图生成带字开业海报	把店名 / 日期 / 优惠 / 二维码排成一幅海报

★ 工具不只会“听话”，更会“补位”

注意 AI 主动做的那些事：把产品照先传到对象存储再处理、为图生图保留原图、在信息缺失时返回明确的验证提示，甚至自动写好项目协作文档 AGENTS.md——这些在笔者的指令里都没明说。一个好工具的价值，正在于它不仅照单全收，还会替你想到那些“你以为理所当然、其实容易翻车”的细节。

2.5　拆解技术方案：四棒接力是怎么跑起来的

整套方案可归结为一句话：**前端把店主填的信息和上传的照片，分发给五个后端接口；后端按“识别→定风格→出图”的顺序，把四棒模态接力依次跑完，再把四块物料回传给前端展示。**下面挑三个最关键的门道讲透。

2.5.1　五个接口如何分工

在五个接口里，upload 负责存图，analyze 负责“动脑”，三个 generate 负责“动手”出图。它们的技术构成如表 2–4 所示。

表 2–4　开店物料流水线的技术构成

步骤	使用的技术	作用与要点
产品上传	对象存储（S3Storage）	把店主上传的照片保存到云端，换成一个稳定可访问的网址，供后续识别与图生图使用
识别 + 定风格	doubao–seed 多模态模型	看懂产品卖相，并结合店铺描述输出风格关键词、主辅色 Hex 值与一句宣传语
门头与海报	SeeDream 5.0 文生图	根据风格卡生成门头大图与带字海报，文字直接“画”进图里
产品展示图	SeeDream 5.0 图生图	以上传的原照为参考，保留真实卖相，换两个场景生成展示美图

2.5.2 让文字“一个字都不出错”的关键

用过 AI 画图的人都知道：在 **AI 生成的图里，文字常常缺胳膊少腿、张冠李戴**。门头要写对店名、海报要写对日期和优惠，错一个字，物料就白做了。AI 是怎么兜住这条红线的?

答案藏在一个朴素的小技巧里：**在喂给文生图的指令中，把必须精确出现的文字用双引号严格包裹起来**。比如，在门头指令里写明“招牌上写着‘拐角’”，在海报指令里写明优惠是“拐角开业 首杯半价”。双引号像一道“免改令”，告诉模型这几个字必须原样照搬、不得改写。同时，这些文字的来源始终是**店主在输入框里亲手填写的内容**，AI 全程不自行改动——这就从源头上杜绝了错字。

> **★ 文字准确，是物料的生命线**
>
> 估价工具答错一个数字，用户还能自己复核；而门头招牌把店名印错一个字，几乎没有补救余地。所以本章把“文字与输入完全一致”当作不可逾越的红线——双引号包裹只是手段，**让用户的原始输入一路不走样地抵达“成图”**才是目的。

2.5.3 图生图如何“保真不吹牛”

产品展示图用的是图生图，这里也有讲究。营销图很容易“用力过猛”——把普通的咖啡修成不真实的样子，顾客到店一对比就会有落差感。所以，要在图生图的指令里明确写下一句英文约束：preserving real appearance without exaggeration（保留真实卖相、不夸大）。

这句约束的意思是：AI 可以换更好看的背景、补更专业的光线，**但那杯咖啡本身的颜色、份量、卖相，必须和店主上传的原照保持一致**。这与第 1 章“成色实话实说、不瞎猜”的精神一脉相承——**可信，永远是 AI 产品的底线**。

2.6 实测验收：给“拐角”咖啡馆出一套物料

方案讲透了，照例进入“先验接口、再验体验”的验收环节。笔者就以朋友小林的“拐角”社区咖啡馆为例，完整跑一遍。

2.6.1 接口冒烟测试：几个“失败”反而证明它对了

和上一章一样，扣子编程会先做冒烟测试，用最小数据快速探一探五个接口是否可用。结果初看有点“吓人”：upload 返回成功（200），而其余几个接口要么报“参数缺失”（400），要么报“服务出错”（500），要么干脆超时无响应（000）。乍一看像是失败了，细究之下才发现——**这些“失败”，恰恰是系统正常工作的证据**，如表 2–5 所示。

表 2–5　冒烟测试结果解读：为什么“失败”反而是好事

接口	测试结果	为什么这其实是对的
/api/upload	200 通过	用真实文件上传，正常成功
/api/analyze 等	400 参数缺失	测试故意不传参数，接口规规矩矩地拒绝并报错——这正是该有的校验行为
generate–product	500 服务出错	测试用的是无法访问的占位图网址，AI 拿不到有效图片才报错，换真实图便正常
generate–signboard / poster	000 超时	AI 出图本就要十几秒，而测试限定 10 秒必须返回，超时是预期之内的

★ 如何看待这些“失败”

把这三类结果对号入座就明白了：**400** 说明接口的“门卫”尽职——没带齐证件（参数）就不放行；**500** 是因为我们传递了一幅假地址的图片，AI 当然获取不到；**000 超时**则是因为 AI 画图确实费时，而测试掐了 10 秒的表。一个在“数据不全、图片无效、时间不够”时都老老实实报错、绝不硬编一个结果糊弄的系统，远比一个“什么都敢答”的系统可靠。这与第 1 章那幅 1×1 像素图触发“尺寸过小”提示，是同一个道理。

2.6.2　填好店铺信息，一键生成

接口探明白了，接下来上传真东西。笔者把小林的店情如实填进界面：店铺描述写下“我开了一家社区小咖啡馆，在小区门口，主打手冲和现烤可颂，氛围想松弛、邻里、有人情味。客人主要是周边上班族和遛娃家庭。风格要复古暖调、木质感，配色用焦糖棕和米白。本周六开业，想做开业首杯半价的活动”，上传一幅手冲咖啡的产品照，再补上开业日期、店铺地址和开业优惠，如图 2–4 所示。



图 2–4　填写店铺信息的输入界面：描述框、产品上传区与日期 / 地址 / 优惠

值得一提的是界面的细节：店铺地址、开业日期、开业优惠都是**独立的输入项**，由店主亲手填写。这正是 2.5 节强调的“文字源头”——所有要进画面的关键文字，都从这里出发，AI 绝不替用户编造。填好之后，单击那个醒目的“一键生成开店物料”按钮，几十秒后，四张卡片便依次呈现。

2.6.3 解读四张物料卡片

第一张是门店风格卡，如图 2–5 所示。AI 先看懂了那杯手冲咖啡，又读懂了“复古暖调、木质感”的描述，于是定下风格关键词“复古暖调 · 木质松弛 · 邻里温情”，主色焦糖棕 #8B4513、辅色奶霜米白 #F5F0E6，并写了一句宣传语“拐角有好咖啡”。卡片底部还附上了它对产品的识别结果——品类、特征、卖点、适用场景一应俱全。这张卡，就是后面三幅图的“设计总纲”。



图 2–5 门店风格卡：风格关键词、主辅色值、宣传语与产品识别结果

第二幅是门头招牌效果图，如图 2–6 所示，由文生图生成。复古木质的门面、暖色的灯光、橱窗里温馨的店内陈设——风格卡里的关键词被精准地“画”了出来。最关键的是招牌上那两个大字“拐角”，**与店主输入完全一致、清晰无误**，这正是双引号“免改令”起了作用。卡片右上角带着“再生成”和“下载”按钮，若对效果图不满意，可以一键重出。

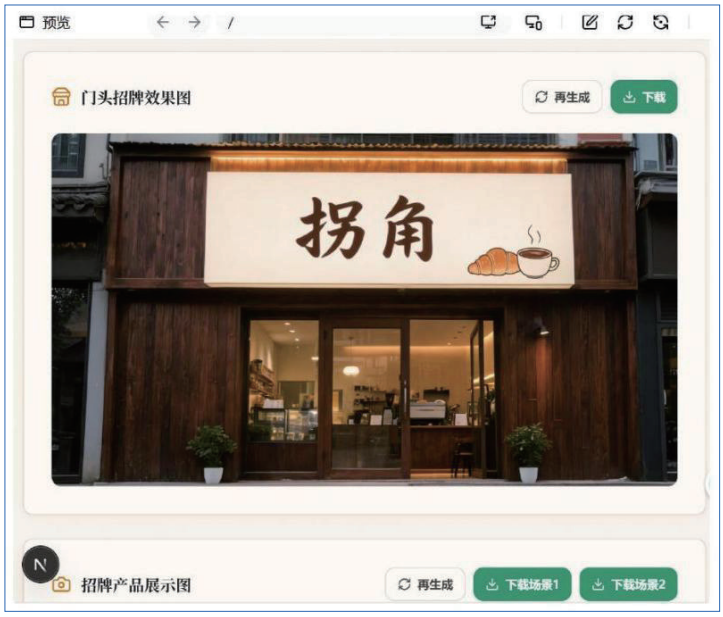


图 2–6　门头招牌效果图：店名“拐角”准确嵌入，风格与描述高度吻合

第三幅是招牌产品展示图，如图 2–7 所示，由图生图生成，一次给出两个场景：场景 1 把咖啡放在洒满晨光的窗边木桌、配上可颂（一种酥脆的法式羊角面包）；场景 2 则把它搬进有顾客围坐的温馨店内。两幅图背景不同、氛围各异，但那杯手冲咖啡的卖相始终如一——**保真而不夸大**的约束在这里清晰可见。两个场景可分别下载。



图 2–7　招牌产品展示图：基于上传原照，图生图产出两种场景且卖相保真

第四幅就是开篇剧透过的带字开业海报，如图 2–1 所示。店名“拐角”、宣传语“拐角有好咖啡”、主打活动“拐角开业 首杯半价”、开业日期“2026.07.04”、地址“上海黄浦区中山南路市民新村”，还有右下角留好的二维码占位与“扫码了解”——所有文字与店主的输入分毫不差，排版疏密得当，俨然一幅可以直接打印张贴的成品海报。若对文案不满意，还能改了重出。

> **★ 回看七要素**
>
> 此时不妨回头对照表 2–2：从四卡片的输出结构，到“下载”“再生成”按钮的界面要求，再到“文字与输入完全一致、产品保真不夸大”的边界约束——我们当初写下的每一条需求，都在这套物料里被一一兑现。七要素指令法，又一次精准地把意图变成了成品。

2.7 本章小结与延伸

用一句话总结本章内容：我们靠**一段清楚的需求描述**，零代码做出了一个能识图、定风格、文生图、图生图的开店物料工坊，让小店主几分钟、零设计费就获得一整套开业物料。

> **★ 本章三个核心收获**
>
> **（1）文生图与图生图**——前者凭文字无中生有，后者照参考图有中再造。这是多模态创作里最基础也最强大的两支笔。
>
> **（2）让图中的文字不出错**——用双引号严格包裹关键文字，并让用户输入全程不走样地抵达成图。这是 AI 出图能落地商用的关键。
>
> **（3）保真即可信**——无论估价还是出图，AI 都要守住“不夸大、不编造”的底线。

商用延伸：这座工坊还能怎么改造

把需求稍作调整，这座“开店物料工坊”就能服务更多生意：

- **餐饮连锁：**把输出物料换成菜品海报、外卖封面、桌牌立牌，新店开业或上新菜品时，根据一幅照片就能获得一整套物料，全国门店视觉还能保持统一。
- **电商卖家：**把“门头招牌”换成“店招 / 主图”，把“开业海报”换成“促销主图”，上传产品照即出一套带卖点文字的电商视觉产品，告别天天找设计师改图。
- **本地生活服务：**美甲店、宠物店、健身房……任何需要开业物料的小微商户，都能拍照即得整套素材，然后直接封装成网页分享出去，顺手就是一门小生意。

本章我们让 AI 学会了**生成一幅图**，但静止的画面，还不足以承载全部的表达。下一章，笔者将带领读者走向接力链的又一棒——**让图“动”起来**：用图生视频的能力，制作一个有镜头、有节奏的短视频。当“出图”进化到“出片”时，多模态创作的想象空间将再次被打开。

第3章 视频引擎：小猫打工 vlog 一键成片

上传几幅小猫照片加一段文案，让 AI 导演写分镜、控成本、一次性交付 20 秒治愈成片

本章导读

前两章，我们都在用扣子编程“造工具”——第 1 章造了一个会看图的估价引擎，第 2 章造了一座会出图的物料工坊。这一章，笔者要换一种打法：**不造工具，直接出成品**。我们的目标是制作一条可以直接发布的短视频——**小猫打工 vlog**：上传几幅自家猫咪的照片，配上一段日常文案，让 AI 导演写分镜、定风格、逐镜生成画面，最后自动合成一条 20 秒的治愈短片。

完成这个目标，靠的是本章的新主角**扣子视频**——一位随叫随到的“AI 导演”。本章会把两个新本领交到读者手里：让静态照片动起来的**图生视频**，以及贯穿全书的一个关键判断——**什么时候该用扣子编程“造工具”，什么时候该用扣子视频“直接创作”**。同时，这也是全书第一次把图像、文字、声音、视频四种模态装进同一个作品里，为后面更复杂的全模态创作打好地基。

3.1 从一个“想给猫拍条短片”的念头说起

先说说本章故事的起点。笔者养了一只黑白花的奶牛猫，名唤嘟嘟，它的鼻子看起来很有趣，睡姿也千奇百怪。刷短视频时看到别人家的“猫咪打工 vlog”——又好笑又治愈，动辄几十万点赞量。于是心里一动：**我家猫咪这么可爱，凭什么不能有一条自己的 vlog？**

可将念头落地就遇到各种困难。真拍？猫不听指挥，“骑车通勤”“地铁发呆”这些画面根本无从拍起。请人做动画？一条 20 秒的定制动画，报价四位数起步，工期以周计。用剪辑软件东拼西凑？素材不够，风格也难统一。这个需求就这样卡住了——**它太“小”了，小到不值得为它学一门手艺、花一笔预算；又太“具体”了，具体到通用模板根本套不上**。

笔者观察

生活里大量的创作冲动，都“死”于同一个尴尬区间：认真做，成本太高；随便做，效果太差。这类需求有一个共同特点——**它们是一次性的**。片子做完、发出去、收获一波点赞，任务就结束了，没有下一次。为这样的需求专门开发一个应用，就像为了喝一杯奶茶去盘一家奶茶店。

“一次性”这三个字，正是本章的题眼。前两章的估价、出图，都是会**反复发生**的需求，所以我们造了工具；而这一次，我们需要的不是工具，而是**一位可以直接对话、直接交片的创作者**。扣子恰好为这两种需求准备了两个入口。先看一眼最终的成片长什么样，如图 3–1 所示：一条 20 秒、竖屏 9:16 的写实短片，主角就是笔者的猫咪，从被闹钟拍醒到瘫在沙发上，配有第一人称的慵懒旁白——所有画面，全部基于笔者上传的猫咪日常生活照片，由 AI 生成。

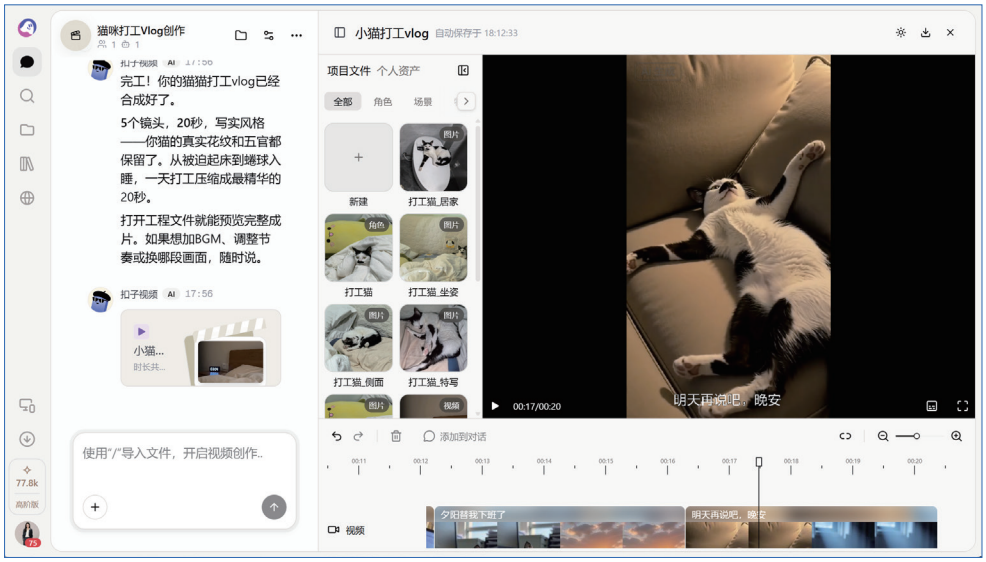


图 3–1　成品速览：合成完毕的 20 秒小猫打工 vlog，可在工程编辑器中继续精修

3.2　扣子视频与图生视频

本节先来认识两位新主角——扣子视频与图生视频。

3.2.1　扣子视频：一位随叫随到的“AI 导演”

扣子视频是扣子家族中专管视频创作的智能体。如果说扣子编程像一位“全栈工程师”，接到需求先搭框架、写代码，再交付一个应用，那么扣子视频更像一位**驻场导演**：你把素材和

想法交给它，它负责理解创意、确认主角、撰写分镜脚本、逐个镜头生成画面，最后把所有片段剪成一条完整的成片。全程都在对话框里推进，你说一句，它做一步，做完还会回头问你“这样行不行”。

在扣子首页单击右上角的“+”号，就能看到“新建视频项目”的入口，如图 3–2 所示。注意，它和“新建编程项目”并排出现，这个并排本身就在提醒我们：**创作与开发，是两条并行的路**。



图 3–2 扣子的新建菜单：“新建视频项目”与“新建编程项目”并列

3.2.2 图生视频与“垫图”：让主角永远是你的猫

本章的核心新本领叫**图生视频**（Image–to–Video）：给 AI 一幅或几幅参考图，再加一段镜头描述，它就能生成一段几秒钟的动态视频。行业里，把参考图输入视频模型这个动作叫**垫图**——图片就像垫在画面下的“底稿”，模型在它的基础上让画面动起来。

垫图解决的是 AI 视频生成**主体一致性**问题。如果只靠文字描述“一只黑白花的猫”，十个镜头会生成十只长相各异的猫，观众一眼出戏。而把猫的真实照片垫进每一个镜头，猫脸上的黑色“八字胡”斑纹、背上的花色分布，就能在骑车、敲键盘、瘫沙发等完全不同的场景中保持一致——这正是这条 vlog 能让“铲屎官”认领“这就是我家猫”的技术底气。

一句话理解

图生图（第 2 章）：照着你的照片，再画一幅**静止的图**。

图生视频（本章）：照着你的照片，生成一段**会动的短片**。垫图，就是把“长什么样”这个问题交给照片回答，把“怎么动”交给文字描述。

3.2.3　关键抉择：直接创作，还是搭建应用

认识了新主角，就要回答本章最重要的问题：**同样是做视频，什么时候用扣子视频，什么时候用扣子编程**？判断标准只有一条——看这个需求会不会**重复发生**，如表 3–1 所示。

表 3–1　扣子视频与扣子编程的分工：一次性创作与可复用流水线

维度	扣子视频（直接创作）	扣子编程（搭建应用）
适用场景	一次性、个性化的成片需求：给猫做条 vlog、给婚礼剪个开场、给年会做个短片	会反复发生的批量需求：每周上新都要产品视频、每个客户都要同款模板片
交付物	一个**可直接发布的成片**（附可继续编辑的工程文件）	一个**可反复使用的应用**，谁都能用、每次输入不同素材
交互方式	和 AI 导演**对话**，边聊边改，所改即所得	用七要素**描述需求**，AI 替你写代码、搭界面
上手成本	零门槛，发照片、说想法即可	需要把流程、输出、边界一次性说清楚
一句话判断	“这片子我就做**这一次**”	“这种片子我以后**每周都要出**”

这条判断标准是全书的一道分水岭，值得读者抄在便签上。扣子视频和扣子编程也可以组合使用：先用扣子视频探索一次性创作，摸清了套路、验证了效果，再用扣子编程把成熟的流程固化成应用——**创作探路，开发固化**。下一章，我们就会亲眼看到这套组合拳的威力。

本章的模态接力链如表 3–2 所示。对照前两章会发现，接力链第一次出现了“视频”这一棒，而且棒与棒之间开始出现**来回传递**——导演写完分镜要等你确认，改完风格再回炉重写。接力赛升级成了排练厅。

表 3–2　本章小猫 vlog 的“模态接力”链路

接力棒	调用的能力	它做的事	交给下一棒的产出
第 1 棒	视觉理解（LLM Vision）	看懂 6 幅猫片：认下主角的花纹、体态、神情	入库的“主角形象”资产
第 2 棒	创意与分镜（LLM）	把 9 段日常文案改编成分镜脚本，逐镜写提示词	一份完整的视频脚本文件
第 3 棒	图生视频（Seedance 2.0）	每个镜头垫 3 幅猫片，生成写实视频片段	5 段各 4 秒的视频素材
第 4 棒	工程合成	按时间轴把片段拼装、加旁白字幕轨	20 秒竖屏成片 + 工程文件

3.3　把素材和想法交给导演：需求怎么提

对话式创作不需要写第 1 章那样滴水不漏的七要素长指令，但**七要素的思维依然管用**，只是形态更松弛——目标、输入、风格、边界，可以拆开来在对话里陆续补齐。笔者的第一条消息只交代了三样东西：**一句话目标**（“创作图片中的小猫的拟人化的第一人称视角的 vlog”）、**一段参考文案**（九段猫咪打工日常，从“被迫起床营业”到“打工是假的，摆烂才是真的”，连话题标签都准备好了）、**六幅猫片**（正脸特写、全身、侧面、睡姿各个角度），

如图 3–3 所示。

原始提示词如下：

创作图片中的小猫的拟人化的第一人称视角的 vlog
参考内容如下，你可以进行优化补充

一只猫的打工日常，比你还真实。
早上伸个懒腰，被迫起床营业
洗脸甩水，精神状态：已上线（但不多）
骑车通勤 → 地铁发呆刷手机
到工位第一件事：踩键盘（谁懂啊）
午饭时间，小鱼干才是灵魂续命
下午统一症状：犯困、摆烂、灵魂出走
下班骑车回家，夕阳都在替我下班
回家直接瘫倒，今日 KPI：活着完成
最后蜷成一团——明天再说吧
打工是假的，摆烂才是真的。
小猫 vlog # 打工日常 # 治愈视频 # 猫咪日常 # 摆烂文学 # 可爱暴击 # 真实的一天



图 3–3 在对话框中输入需求

笔者在提示词中特意加了一句“你可以进行优化补充”——这句授权很关键，等于告诉导演：文案是参考，不是准则，专业的事你可以拿主意。

发送之后，扣子视频立刻进入工作状态（见图 3–4）：它先加载创作技能，然后逐一查看猫片、把主角形象入库，接着自己定下创意方向——温暖治愈 3D 动画风格、竖屏 9:16、第一人称与第三人称混切——并开始撰写分镜。

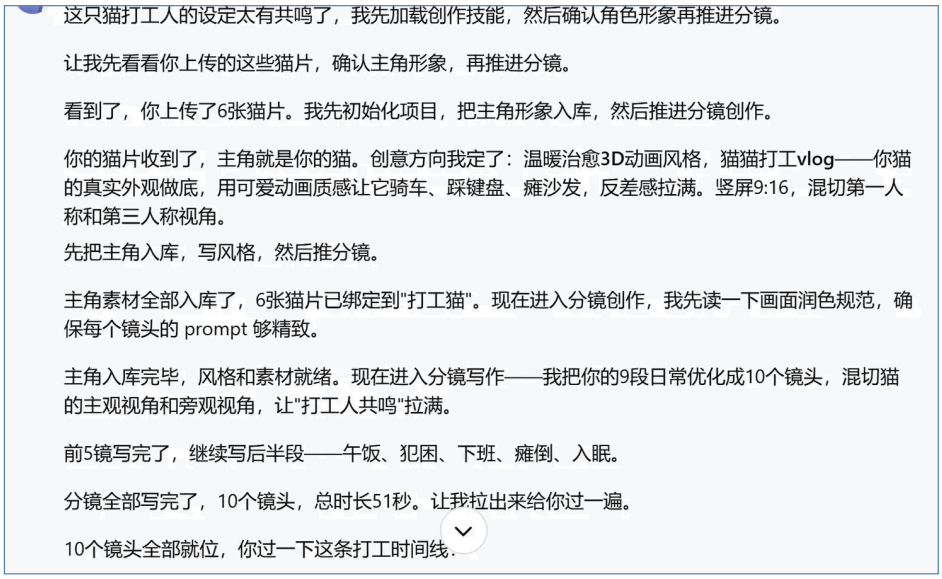


图 3–4　发出需求后，扣子视频开始阅读流程、分析猫片，把主角形象入库

> **★ 别慌**
>
> 整个过程中，扣子视频会像工作汇报一样不断“说话”：主角入库了、风格定了、前 5 镜写完了、正在写后半段……这些播报不是废话，而是**给你留出的插话窗口**——任何一步觉得不对，随时打断它。对话式创作的安全感，正来自这种全程透明。



3.4　三轮“剧本会”：和 AI 导演讨价还价

接下来发生的事，是本章最精彩的部分。笔者和 AI 导演开了三轮“剧本会”，每一轮都真实展示了扣子视频的一项看家本领。

3.4.1　第一轮：初版方案与 10 镜分镜表

几分钟后，导演交出了初版方案：**10 个镜头、总时长 51 秒**，并把分镜表完整拉了出来（见图 3–5）——每一镜都标了时间、场景内容，还有一栏“导演备注”来解释创作意图：“低角

度猫眼看天花板，沉浸感开场”“瞳孔缩小的放空特写，打工人共鸣场面”“安静温暖的收尾，明天再说留余韵”。它甚至主动给出了自己的节奏判断：前半段用可爱反差造笑点，后半段用夕阳和瘫倒过渡到共情。最后问了一句：“确认方向没问题的话，我就推进生成。想调哪里也可以说。”



扣子视频 AI 06-08 17:16

这只猫打工人的设定太有共鸣了，我先加载创作技能，然后确认角色形象再推进分镜。

让我先看看你上传的这些猫片，确认主角形象，再推进分镜。

看到了，你上传了6张猫片。我先初始化项目，把主角形象入库，然后推进分镜创作。

你的猫片收到了，主角就是你的猫。创意方向我定了：**温暖治愈3D动画风格，猫猫打工vlog**——你猫的真实外观做底，用可爱动画质感让它骑车、踩键盘、瘫沙发，反差感拉满。竖屏9:16，混切第一人称和第三人称视角。
先把主角入库，写风格，然后推分镜。

主角素材全部入库了，6张猫片已绑定到“打工猫”。现在进入分镜创作，我先读一下画面润色规范，确保每个镜头的prompt 够精致。

主角入库完毕，风格和素材就绪。现在进入分镜写作——我把你的9段日常优化成10个镜头，混切猫的主观视角和旁观视角，让“打工人共鸣”拉满。

前5镜写完了，继续写后半段——午饭、犯困、下班、瘫倒、入眠。

分镜全部写完了，10个镜头，总时长51秒。让我拉出来给你过一遍。

10个镜头全部就位，你过一下这条打工时间线：

表格

#	时间	场景	导演备注
1	0-5s	被闹钟叫醒，猫爪拍闹钟伸懒腰	低角度猫眼看天花板，沉浸感开场
2	5-9s	洗脸甩水，水珠慢动作飞溅	全片最“精神”的一镜，反差笑点
3	9-14s	骑迷你自行车晨间通勤	耳朵被风吹倒的细节很戳，治愈感拉满
4	14-19s	地铁发呆刷手机，灵魂出走	打工人共名场面，瞳孔缩小的放空特写
5	19-24s	跳上键盘踩出一串乱码	全片最搞笑的镜头，“谁懂啊”的精准视觉化
6	24-29s	小鱼干午餐，腮帮子鼓鼓	唯一一镜眼睛亮了，反差证明猫猫本质还是
7	29-35s	下午犯困，脑袋磕键盘流口水	灵魂离体的社畜巅峰时刻
8	35-40s	夕阳下骑车回家，影子拖长	情绪转折点，暖光收住一天的丧
9	40-45s	瘫成猫饼，KPI：活着完成	每个打工人回家的真实状态
10	45-51s	蜷成球入睡，月光洒进小窝	安静温暖的收尾，“明天再说”留余韵

总时长51秒，竖屏9:16，3D动画质感。每镜都有打工猫的旁白金句，配合画面节奏不会赶。

我的判断：这个节奏对治愈vlog来说刚好——前半段用可爱反差造笑点，后半段用夕阳和瘫倒把情绪从搞笑过渡到共情，最后安静收住不煽情。

确认方向没问题的话，我就推进生成。想调哪里也可以说。

图 3–5 初版方案：10 镜 51 秒的分镜表，每镜附时间、场景与导演备注

3.4.2　第二轮：一句话改风格——“我要写实”

笔者看完前述扣子导演的创意，发现它把成片风格定为 3D，但笔者希望就使用嘟嘟本体，因此提了修改意见：“我希望是写实的风格，尽量还原我的猫猫真实图片，毕竟它这么可爱。”这句话推翻的是导演已经定稿的整体视频美术方向。换成人类团队，这叫“客户临时改风格”，需要设计师加一晚上班。而扣子视频的反应是先讲道理再动手：“写实风格完全合理——猫本身就够可爱了，3D 动画反而会稀释她的魅力。风格我直接改。”随后它把全片风格改成了**超写实摄影**，并承诺猫的真实毛发和五官细节都会保留，如图 3–6 所示。

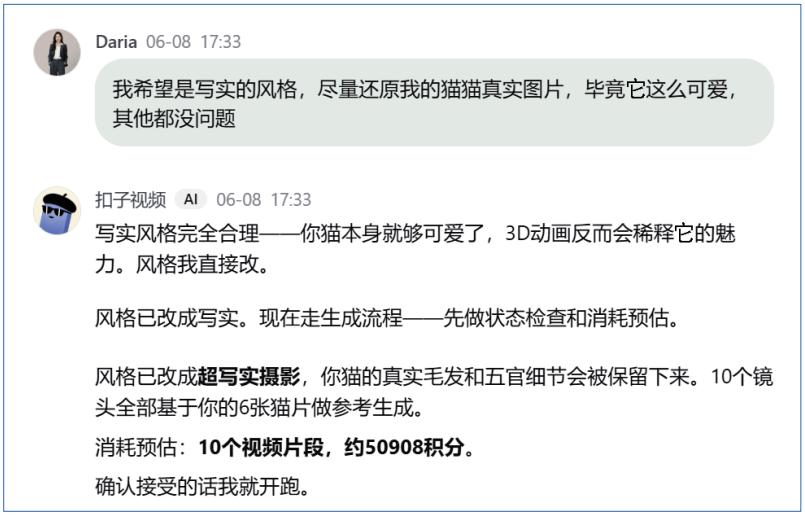


图 3–6　第二轮修改：一句话把全片改成超写实摄影风格

改完风格，它没有直接开跑，而是先做了**状态检查和消耗预估**：10 个视频片段，约需 50908 积分。这是扣子视频另一项值得点赞的设计——**先报价，再干活**。AI 生成视频是真金白银的算力消耗，把账单预估摆在开工之前，用户才有踩刹车的机会。

3.4.3　第三轮：多图垫图与压缩成本

这个报价让笔者踩了刹车——在翻看了脚本细节之后，笔者附上两个修改要求：**第一**，“脚本里每个镜头垫三幅参考图，还原猫咪全身的花纹”；**第二**，“为了节省积分，你压缩一下时长，精选最出彩的片段，时长压缩到 20 秒”。

导演的处理：它先复述确认——“两个调整都很关键：多图参考还原花纹，精选镜头压缩到 20 秒”；然后动手重构——**10 镜合并成 5 镜**，每镜内部用分时序完成两个节拍（比如“闹钟叫醒 + 洗脸甩水”合成晨起一幕）；**每镜挂上 3 幅不同角度的猫片**做参考，全身花纹、面部细节、侧面体态都有了着落；发现台词与新时长不匹配时，主动把每句旁白压缩到 4 秒内能承载的字数，保证“金句不丢”，如图 3–7 所示。

★ 三轮剧本会，三项看家本领

（1）**会解释，不盲从**——改风格时先说清“为什么合理”，专业判断力和执行力并存。

（2）**先报价，再干活**——每次生成前给出积分消耗预估，成本透明可控。

（3）**改一处，顺全局**——压缩时长牵动台词节奏，它自动重新排片，不用你逐项交代。



图 3–7 第三轮修改：10 镜并 5 镜、每镜 3 幅参考图，总时长压缩到 20 秒

3.5 看懂“施工图”：一份视频脚本文件里的门道

三轮讨论的最终成果，沉淀为一份视频脚本，这是全片的“施工图”，值得点开细看，如图3–8所示。文件头部声明了全局参数：**超写实摄影风格、真实光影质感、电影级调色；画幅 9:16、清晰度 720p、共 5 个镜头、模型 Seedance 2.0、总时长 20 秒**。

再看单个镜头的提示词，门道更多。以镜头 1 为例：它先声明“参考输入：图片 1= 打工猫；图片 2= 打工猫 _ 特写；图片 3= 打工猫 _ 居家”——**给每幅垫图起名字、派角色**；然后用镜头语言描述画面：“【低角度仰拍，镜头微微晃动】温暖但昏暗的卧室，从猫的视角看向天花板，手机闹钟屏幕亮着……”**旁白的声线**也都写得明明白白：“女声，青年音色，音调中低偏软，音色质感微哑带鼻音，发音方式慵懒拖沓，气息绵长偶尔叹气，吐字含糊可爱”；末尾还有两条否定约束：“不添加字幕，不添加 BGM”——把干净的素材留给后期，字幕和配乐在合成阶段统一处理。



图 3–8　视频脚本文件：全局风格声明 + 逐镜提示词 + 每镜绑定的 3 幅参考猫片

> **★ 提示词的三层结构，值得学习**
>
> 一条专业的视频提示词 = **画面层**（景别、运镜、光线、动作）+ 声音层（声线、语气、台词）+ **约束层**（不要什么）。哪怕将来换用别的视频工具，这个三层框架照样通用——它本质上是把导演、录音师和剪辑师的要求写进了同一段文字。

3.6　一键开跑：从片段生成到自动成片

笔者查看了扣子视频的修改，直接回复：“非常棒！我很喜欢，开始创作吧。”导演照例先做状态检查和消耗预估：**5 个视频片段，约 19964 积分**——比初版的 50908 积分省了六成还多。随后 5 个片段陆续生成，对话里能看到每段的缩略图，如图 3–9 所示。

片段全部就位后，扣子视频自动推进**工程合成**：把 5 段素材按分镜顺序装配到时间轴上，对齐旁白节奏，合成一条完整的 20 秒成片，如图 3–9 所示。它的交付总结言简意赅：“5 个镜头，20 秒，写实风格——你猫的真实花纹和五官都保留了。从被迫起床到蜷球入睡，一天打工压缩成最精华的 20 秒。”

尤其要注意成片的交付形态：**不是一个只能播放的视频文件，而是一个可以继续编辑的工程**。打开工程编辑器（见图 3–1），左侧是全部素材资产——猫片按“居家、坐姿、侧面、特写”分类入库，5 段生成视频整齐排列；下方是带字幕轨的时间轴，“被迫营业勉强上线”“谁懂啊！小鱼干续命”“夕阳替我下班了”逐段就位。想换 BGM、调节奏、重新生成某一段画面，随时可行。**成片不是终点，而是下一轮创作的起点**——这正是“工程交付”比“文件交付”高明的地方。



图 3–9 确认开跑：19964 积分的最终预估，5 个写实片段陆续生成

3.7 本章小结与延伸

用一句话总结本章内容：我们没有写一行需求文档，只靠**几幅照片、一段文案、三轮对话**，就从 AI 导演手里拿到了一条主角保真、节奏完整、成本减半的 20 秒成片。比成片更值钱的，依旧是装进工具箱的三个本领。

★ 本章三个核心收获

（1）一次性创作找导演，重复需求造工具。扣子视频管“这一次”，扣子编程管“每一次”。判断只看需求会不会重复发生。

（2）垫图保主角。参考图是主体一致性的锚，多角度多图垫入，猫还是那只猫。这一招在后续一切图生视频场景中通用。

（3）成本要谈判。生成视频前先看消耗预估，用合并镜头、压缩时长节省成本。会省积分，也是一种创作力。

商用延伸：这位 AI 导演还能接什么活

- **宠物与母婴博主：**固定几幅主角照片入库，每周用新文案出一条 vlog，人设与画风长期统一，账号再也不断更。

- **电商卖家：**把“猫片”换成产品图，垫图生成产品使用场景短视频，详情页从图文时代直接跳进视频时代。
- **本地商家与活动策划：**开业宣传、节日祝福、年会开场，一次性的片子交给 AI 导演，当天想法当天出片。

本章我们体验了“直接向 AI 要一条成片”的爽快。但如果同样的图每周都要出几十幅呢？下一章，笔者将带大家用回扣子编程，把生图能力锻造成一条**可复用的工业级流水线**——不仅上传产品照片，让 AI 输出带零件标注的**爆炸结构图**，还要让 AI 检查自己画的图。“创作探路，开发固化”的后半式，即将登场。