M2 · 第 5 讲

结构化输出与批量任务

让一百条结果长得一样

阅读约 35 分钟

练习约 30 分钟

产出：一张完整分类表

本节目录

开场：200 条评论，一条一条问要问到明天

本课目标

一、为什么格式会崩

二、三件套：让输出稳定

三、表格还是 JSON

四、批量处理：分批 + 编号

五、失败处理：让模型自己报错

六、30 分钟练习

七、四个常见坑

八、检查清单

九、延伸阅读

开场：200 条评论，一条一条问要问到明天

你导出了 200 条用户评论，想让 AI 分成“功能诉求、体验问题、价格异议、好评”四类。你试了三条，格式就开始飘：有的输出一段话，有的多给解释，有的把两条合成一条。

问题不在模型笨，而在于你只给了任务，没给**结构**。这一讲解决两件事：让输出稳定成形，让批量处理可控。

本课目标

- 用“字段 + 示例 + 校验”三件套固定输出；
- 判断什么时候用表格、什么时候用 JSON；
- 把 200 条任务拆成可重试的小批；
- 建一个简单的失败处理规则。

一、为什么格式会崩

三个原因，按出现频率排：

1. **没定义字段**：你说“分类”，没说输出哪几列；
2. **没给示例**：模型不知道你要的颗粒度；
3. **一次给太多**：输入太长，它开始省略、合并、编造。

对应三个动作：定义 schema、给一个样例、分批处理。

二、三件套：让输出稳定

1. 字段定义

> 输出四个字段：id（原文编号）、category（四选一：功能/体验/价格/好评）、reason（20 字以内）、confidence（高/中/低）。

2. 一个样例

> 示例输入：3号，“导出老是失败，气死了”。 示例输出：{"id":3,"category":"体验","reason":"导出功能失败","confidence":"高"}。

3. 校验规则

> 只输出 JSON 数组，不要解释；分类只能是四个值之一；无法判断时 category 填“待定”，confidence 填“低”。

三件套齐全，格式基本不会跑偏。

三、表格还是 JSON

用途	选表格	选 JSON
给人看	是	否
给程序用	否	是
字段少、结构固定	是	可以
需要嵌套（一条多个标签）	勉强	是

一个经验：**给人看的用表格，给下一步程序用的用 JSON。**

四、批量处理：分批 + 编号

不要一次丢 200 条。按 20–30 条一批，并给每条编号：

> 下面是第 1–20 条，已编号。请逐条处理，输出条数必须等于输入条数。缺哪条就在结果里标“缺失”，不要补写。

三个关键约束：

- **条数对齐**：输出条数 = 输入条数；
- **保留编号**：方便回填与核对；
- **禁止补写**：信息不足就标“待定”。

处理完一批，抽查 3 条，再继续下一批。不要等 200 条全跑完才发现分类标准错了。

> 批处理分两种：Batch API 约按同步价 5 折、异步执行，有独立限额；把 20–30 条放进普通请求循环，仍按标准价并受 RPM/TPM 限制（2026-11-20 查询，按平台当期规则）。

五、失败处理：让模型自己报错

批处理一定会遇到异常：空文本、外语、广告、重复。提前约定：

> 如果某条无法处理，输出 {"id":n,"category":"待定","reason":"原因","confidence":"低"}，不要跳过，也不要猜测。

这样你拿到的是一张完整的表，而不是一个需要手工对齐的残缺结果。

六、30 分钟练习

1. 找 20 条真实评论/反馈（5 分钟）；
2. 定义 3–4 个字段，写一个样例（10 分钟）；
3. 分两批让 AI 处理，检查条数与编号（10 分钟）；
4. 抽查 3 条，修正分类标准后再跑一遍（5 分钟）。

七、四个常见坑

1. **只给分类名，不给判断标准**：同类会被分到不同类；
2. **一次丢整份文档**：模型开始摘要，不再逐条；
3. **不检查条数**：少了 5 条你可能一周后才发现；
4. **把置信度当装饰**：低置信度的条目必须人工复核。

八、检查清单

- [ ] 字段名、取值范围、长度都写清；

- [ ] 至少给了一个输入—输出样例；
- [ ] 单批不超过 30 条；
- [ ] 输出条数与输入一致；
- [ ] 低置信度条目已人工复核。

九、延伸阅读

- 知识库 1.2 Prompts（提示词）
- 知识库 1.3 AI Agents（智能体）
- 知识库 2.5 数据：AI 产品数据

> 下一讲：图像与视频提示词。你会用一张分镜表，把文字想法变成可拍的镜头。
