

商场随手拍 · 购物小票 + 半货架背景

多模态系列 · 第 1 章

多模态大模型是什么

从外挂到统一的演进逻辑

先看这张图 —— 就这么一张小票，我连问你三个问题，**难度根本不在一个量级**。

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究



商场随手拍 · 购物小票 + 半货架背景

同一张图，三个问题 难度递增

1 看见 · 感知 ★☆☆

这张图里有什么？

2 看清 · 细粒度 ★★☆

小票上第三行商品的单价是多少？

3 想明白 · 推理 ★★★

买的东西够不够做一顿晚餐，为什么？

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究



商场随手拍 · 购物小票 + 半货架背景

同一张图，三个问题，
难度不在一个量级

看见	看清	想明白
★☆☆	★★☆	★★★

这一讲，顺着这三个问题，把一件事讲透：

- 多模态大模型到底是什么
- 它又为什么会长成今天这个样子

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

回到那张小票 · 三个问题 → 三种能力

① 图里有什么? ★

② 第三行商品的单价? ★★

③ 够做一顿晚餐吗? ★★★

看见 感知

认出小票、货架、大概场景

第 1 级

看清 细粒度理解

定位第三行 · 读出十几像素的数字

第 2 级

想明白 推理

读出所有商品 · 组合推断用途 · 给理由

第 3 级

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

看见 → 看清 → 想明白

难度 · 台阶式往上走 ↗



唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

看见 → 看清 → 想明白 · 难度台阶式往上走

把这三个时间点对着说一遍，你就有体感了

看见 感知

认出小票、货架、大概场景

2023 已答得不错

第 1 级

看清 细粒度理解

定位第三行 · 读出十几像素的数字

等高分辨率技术成熟

第 2 级

想明白 推理

读出所有商品 · 组合推断用途 · 给理由

学会思考+看清细节+调用工具 一起跟上才慢慢改善

第 3 级

⚠ 不是单点突破一步到位

- 学会思考
- 看清细节
- 调用工具

→ 一起跟上 才慢慢改善

唐国梁Tommy丨TGLTommy · 原创出品 · 侵权必究



到今天，还得靠人核验

不能全交给它

看见 ✓ 看清 ✓ 想明白 ✓ → 仍需人把关

唐国梁Tommy丨TGLTommy 原创出品 · 侵权必究

📚 整门课 · 主线

模型怎么一步步学会
看见 · 看清 · 想明白

📍 这一讲 · 开篇

先回答最底层的问题：
这类模型是什么、
为什么演化成今天的形态

看见
感知
→
看清
细粒度
→
想明白
推理

唐国梁Tommy | TGLTommy
原创出品 · 侵权必究

多模态大模型

MLLM Multimodal Large Language Model

宽口径定义

≥ 2 种模态联合处理 + 大规模预训练 → 理解 / 生成

宽定义下面有好几条不同的实现路线

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

这门课重点讲 最主流的一条路线

交叉注意力 | 原生统一 | ★ 主流路线 | 连续视觉表征

① 脑 · 认知 / 生成中枢

以大语言模型
为认知 / 生成中枢

② 眼 · 非文本模态

接入图像 / 视频等
非文本模态的表征

③ 桥 · 统一序列

统一 token 序列上
完成理解 / 生成

① 脑 · ② 眼 · ③ 桥 —— 一副三件套的雏形

↩ 其中一部分模型还能反过来生成非文本输出

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

主流路线 = 三件套 + 统一 token 序列 （最常见的实现）

⚠️ 不是 MLLM 的必要定义

「三件套 + 统一序列」只是这条主流路线最常见的实现，而非硬性定义

这些路子并不完全照它来

≠ 交叉注意力

≠ 原生统一模型

≠ 连续视觉表征

这条主线里有三个词 —— 逐个讲清楚 →

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

第一个词 · ①

以大语言模型为认知中枢



👁️ 装上眼睛　👂 装上耳朵

那个大脑，是现成的

已在海量文本上练出——

语言　知识　推理

整条技术路线最重要的前提

不重新造一个大脑，
而是给现成大脑装上眼睛和耳朵。

唐国梁Tommy丨TGLTommy 原创出品 · 侵权必究

第二个词 · ②

非文本输入



图像

最成熟

>



视频

依次跟进

>



音频

依次跟进

图像是最成熟的一路，视频和音频依次跟进。

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

第三个词 · ③

统一的 token 序列 万物皆 token

图块 · 图像



文字



声波 · 音频



切块

统一 token 序列 · 跟文字排进同一串



图像 token　文字 token　音频 token

这是本课主讲路线的 **典型实现** ，不是所有多模态模型的硬性要求。

唐国梁Tommy丨TGLTommy 原创出品 · 侵权必究

能力全景 · 覆盖六个面

按能力类型分，一个成熟的多模态大模型大概能做这些

完整能力表列给你，扫一眼有个全貌就行

① 💬

看图说话 · 图片问答

VQA · 看图描述与回答

② 📄

读图里的文字

OCR · 文档理解

③ 📊

读图表 · 解数学题

图表理解 · 数学推理

④ 🎯

视觉定位 Grounding



指出某个东西
在图里的
「哪个位置」

⑤ 🎞

跨多图 · 视频时序

多图比较 · 视频时序理解

⑥ 🎨 生成侧

生成图像 / 音频

反向生成 · 生成侧

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

本课主线 · 只走理解侧 = 看见 → 看清 → 想明白

① 看图说话 · 图片问答

VQA · 看图描述与回答

② 读图里的文字

OCR · 文档理解

③ 读图表 · 解数学题

图表理解 · 数学推理

④ 视觉定位 Grounding



指出某个东西
在图里的
「哪个位置」

⑤ 跨多图 · 视频时序

多图比较 · 视频时序理解

⑥ 生成侧

生成图像 / 音频

反向生成 · 生成侧

↑ ⑥ 生成侧：只在必要处点一下 · 值得单独一门新的精品课来讲

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

这类模型，为什么长成今天这个样子？

不讲编年史，讲驱动演进的稳定逻辑

① 编年史式讲法

罗列几十个模型 + 发布日期

Qwen3-VL	2025.09
Gemini 3	2025.11
GLM-4.6V	2025.12
MiniCPM-o 4.5	2026.02
Gemini 3.1 Pro	2026.02
Qwen3.5	2026.02
GLM-5.1	2026.04
Seedance 2.0	2026.04
Lance	2026.05
LLaVA-OneVision-2.0	2026.05
GLM-5	2026.06
Gemini 3.6 Flash	2026.07
更多，还在变长…	

✗ 名单会一直变长

换框架

② 瓶颈 → 解法框架

把整段演进史压成三次转折

转折一 23 — 24

瓶颈：从零训练太贵 → 解法：复用组件 · 搭桥

转折二 24 — 25

瓶颈：吃不下真实输入 → 解法：原生分辨率 · 联训

转折三 25 — 26

瓶颈：想不深 · 两套系统 → 解法：RL 推理 · 统一表征

驱动演进的逻辑是稳定的 · 压成三次转折

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

这类模型，为什么长成今天这个样子？

不讲编年史，讲驱动演进的稳定逻辑

① 编年史式讲法

罗列几十个模型 + 发布日期

Qwen3-VL	2025.09
Gemini 3	2025.11
GLM-4.6V	2025.12
MiniCPM-o 4.5	2026.02
Gemini 3.1 Pro	2026.02
Qwen3.5	2026.02
GLM-5.1	2026.04
Seedance 2.0	2026.04
Lance	2026.05
LLaVA-OneVision-2.0	2026.05
GLM-5.2	2026.06
Gemini 3.6 Flash	2026.07
更多，还在变长…	

✗ 名单会一直变长

换框架

② 瓶颈 → 解法框架

把整段演进史压成三次转折

转折一 23 – 24

瓶颈：从零训练太贵 → 解法：复用组件 · 搭桥

转折二 24 – 25

瓶颈：吃不下真实输入 → 解法：原生分辨率 · 联训

转折三 25 – 26

瓶颈：想不深 · 两套系统 → 解法：RL 推理 · 统一表征

驱动演进的逻辑是稳定的 · 压成三次转折

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

三次转折之前，必须先发生的一件事

CLIP · 把图文对比学习彻底规模化

4.0 亿组 图文对

2021 规模化 · zero-shot 迁移

同一个语义空间 · 图像与文字共用一套坐标



唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

一张狗的照片 ≈ 「一只狗」这句话 — 向量空间里靠得很近



zero-shot 迁移 · 无需再训练

不同概念 · d ≈ 476

距离小 · d ≈ 198

「一只狗」

文本 · 一句话

图像 · 狗照片

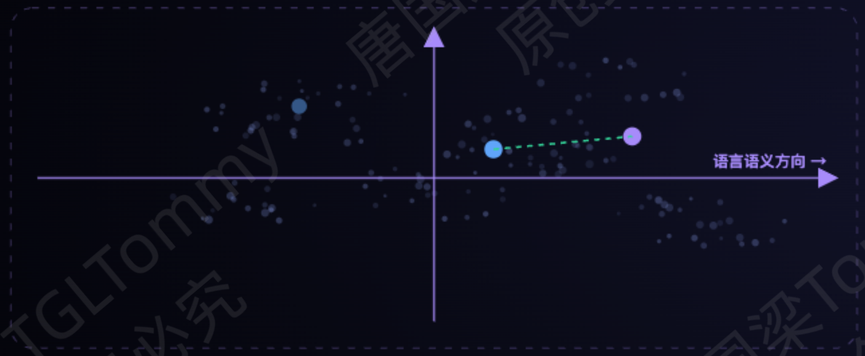
把图文映射到同一空间 并非 CLIP 首创；它真正的分量 ——

把这条路线做到规模 + 证明迁移有价值 · 一发布即受关注，当时价值主要在 开放词汇的视觉表征

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

又过了两年，人们想给语言模型 装眼睛 时才发现

特征空间 = 语言坐标系



CLIP 视觉编码器输出的特征，天生沿语言语义排布

特征天生带着语言坐标系

→ 成为给 大语言模型接视觉输入的重要底座

视觉编码器 = 眼睛

下一节：三件套登场

* CLIP 的完整故事，本课第 2 章会专门讲

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

第一次转折 2023 — 2024

瓶颈 🔴

从零训练多模态，贵得离谱 💰

解法 🟢

复用两个成熟组件，中间搭桥

最贵的做法：数据、算力、时间，全部从零烧起

➜ 两个都别动，只搭中间那座桥

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

第一次转折 · 外挂拼接

搭桥立起「三件套」



两端都冻结 · 只训中间那座桥 → 训练成本骤降

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

两个代表作，放一起对比着看

同一套三件套 · 只换中间那座桥

BLIP-2 2023 初

两端冻结

视觉编码器 → Q-Former（32 token） → 大语言模型

整张图 ➜ 压成 32 个 token



LLaVA 2023.4

冻结 · 更新

CLIP 视觉编码器 → 线性投影（一层直连） → Vicuna

更简的路：一层线性投影直连

第二阶段：大脑一起更新 并非两端都冻结

2023 初 · 两端都冻结 · 32 个可学习 query 把整张图压成 32 token · 训练成本极低

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

两个代表作，放一起对比着看

同一套三件套 · 只换中间那座桥

BLIP-2 2023 初

两端冻结

视觉编码器 → Q-Former（32 token） → 大语言模型

整张图 ➜ 压成 32 个 token



LLaVA 2023.4

冻结 · 更新

CLIP 视觉编码器 → 线性投影（一层直连） → Vicuna

更简的路：一层线性投影直连

第二阶段：大脑一起更新 并非两端都冻结

2023.4 · 一层线性投影就搭好桥 · 第二阶段连大脑一起更新，并不是两端都冻结

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

LLaVA-1.5 · 配置演化

开创视觉指令微调



336px	两层 MLP	576
像素输入	替换线性投影	patch token

这几个数字，是 LLaVA-1.5 的配置

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

⚠ 视觉指令微调 · 一个常被讲错的点

~~直接看图 → 不是让 GPT 直接看图，自己生成对话数据~~

真实做法



再拿这些数据，去教模型听懂指令

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

外挂拼接时代的遗产 · 三件套 & 三个短板



视觉编码器 + 连接器 + 大语言模型 = 三件套

这个时代最重要的遗产 · 一直沿用到今天 · 下一节专门讲它

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

外挂拼接时代的遗产 · 三件套 & 三个短板

模型能力上限 = 编码器强弱框住



① 信息瓶颈

32 token 装不下一张大图

碰 OCR / 视觉定位就吃力



② 看不清

只吃固定低分辨率 224 / 336px

文档缩到这尺寸 · 字全糊



③ 适配受限

冻结压成本，视觉表征难进化

编码器强弱 · 框住整个模型上限



唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

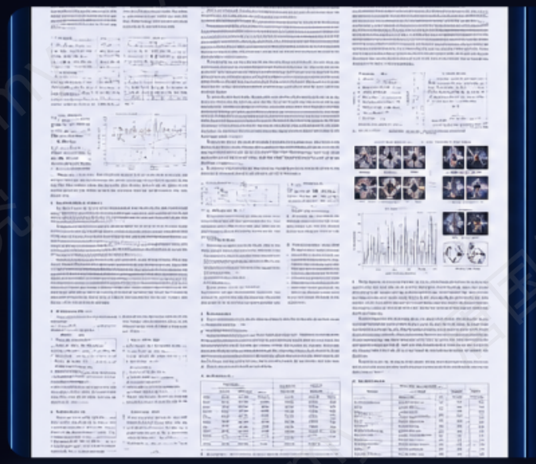
第二次转折 2024 — 2025

瓶颈 🔴

看不清、吃不下真实世界输入

🌍 真实世界的输入

✗ 上一代外挂拼接 · 基本处理不了







📄 4K 文档扫描件

🖥️ 超长网页截图

📦 几十分钟视频

唐国梁Tommy丨TGLTommy · 原创出品 · 侵权必究

第二次转折 2024 – 2025

瓶颈 🔴
看不清、吃不下真实世界输入

→

解法 🟢
从 2024 起 · 沿三条线同时突破

整个行业 · 沿三条线同时突破

① 分辨率
看清 4K · 超长截图
↳ 动态切块 / 原生分辨率

② 冻结 → 联合
编码器解冻 · 联合微调
↳ 为大脑量身进化

③ 模态拓宽
音频进场 · omni 全模态
↳ GPT-4o 全模态

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

第一条线 · 主攻分辨率

旧方案：固定低分辨率

4K 文档 / 大图



固定低分辨率

强制压成 336 × 336



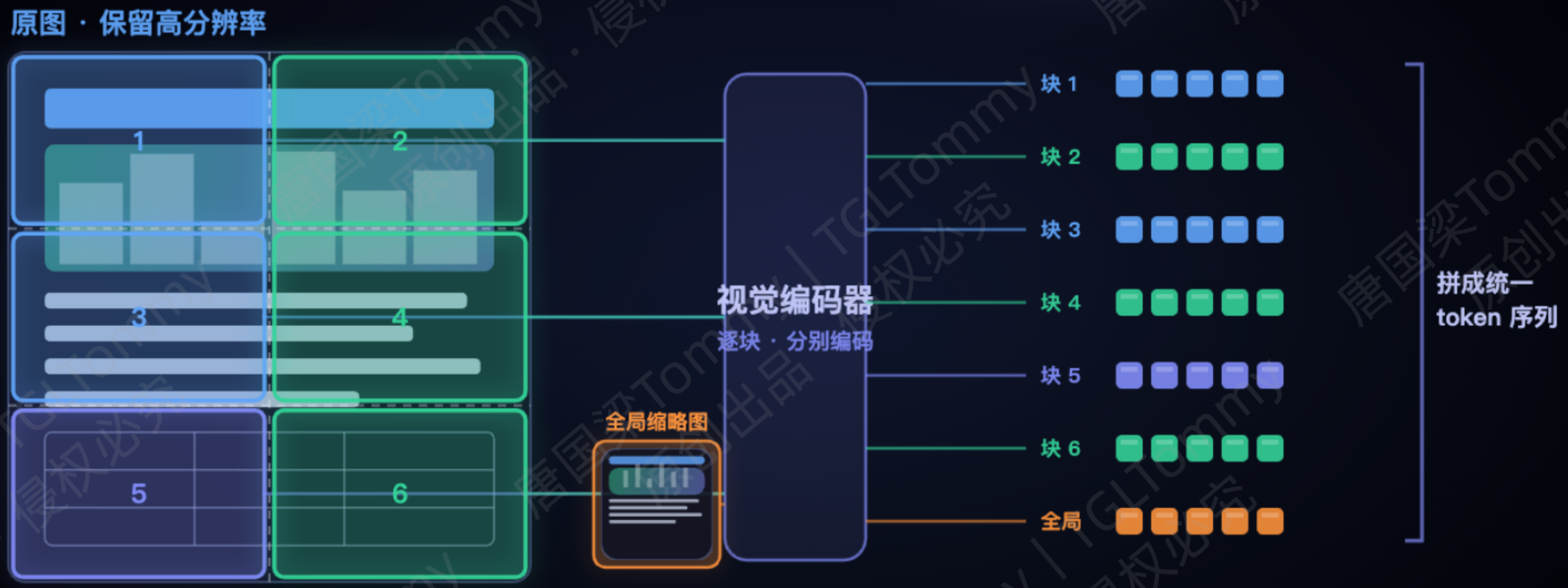
字全糊了 ✗

⚠ 一张 4K 文档缩到 336^2 · 细节丢失、小字读不出 —— 这正是要解决的问题

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

解法 · 动态切块

把大图切成小块，分别编码



表：LLaVA-NeXT · InternVL-1.5 › 切块思路至今仍在用（InternVL 系列） › 紧接着 → Qwen2-VL · NaViT › Naive Dynamic Resolution：按原尺寸动态处理 →下一

分辨率演进 · 动态分辨率

Qwen2-VL　NaViT　Naive Dynamic Resolution



✓ 保持宽高比　✓ 像素预算 + 网格约束下等比对齐　✓ 不再压成固定方块

token 数量随图像大小 自然浮动 ↕

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

① 固定低分辨率 › ② 动态切块 › ③ 动态分辨率

同一机制，token 数跨度极大



已成旗舰 VLM 常见配置 固定单一小方图不再是唯一主流 · 第 2 章专门展开

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

第二次转折 · 原生融合

第二条线：冻结 → 解冻

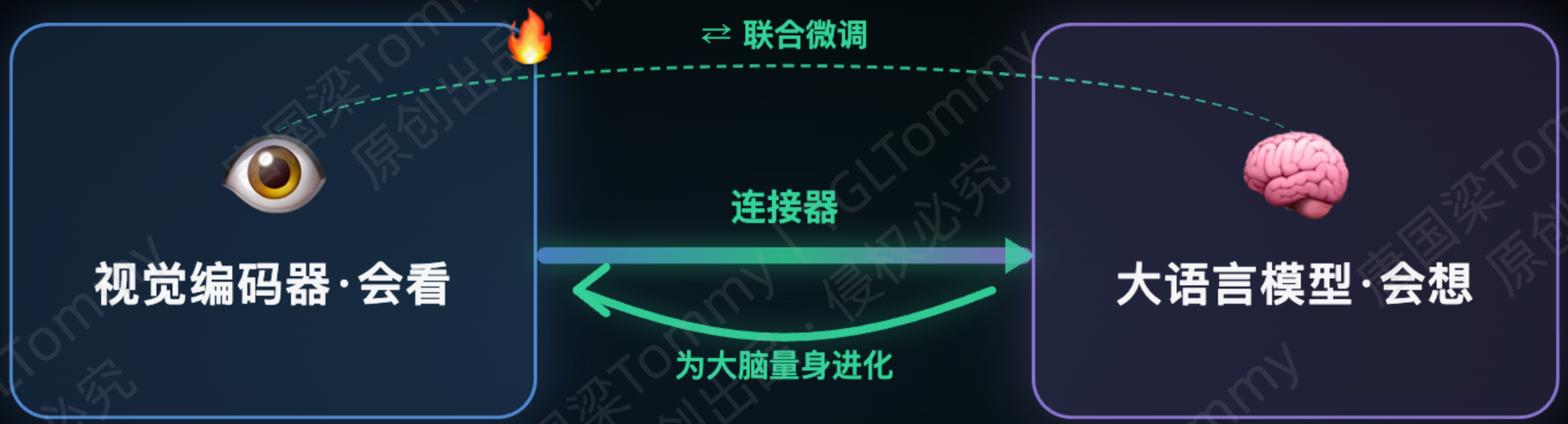


🔥 视觉编码器训练后期普遍解冻 · 不再是冻着不动的前处理插件

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

第二次转折 · 原生融合

第二条线：联合微调

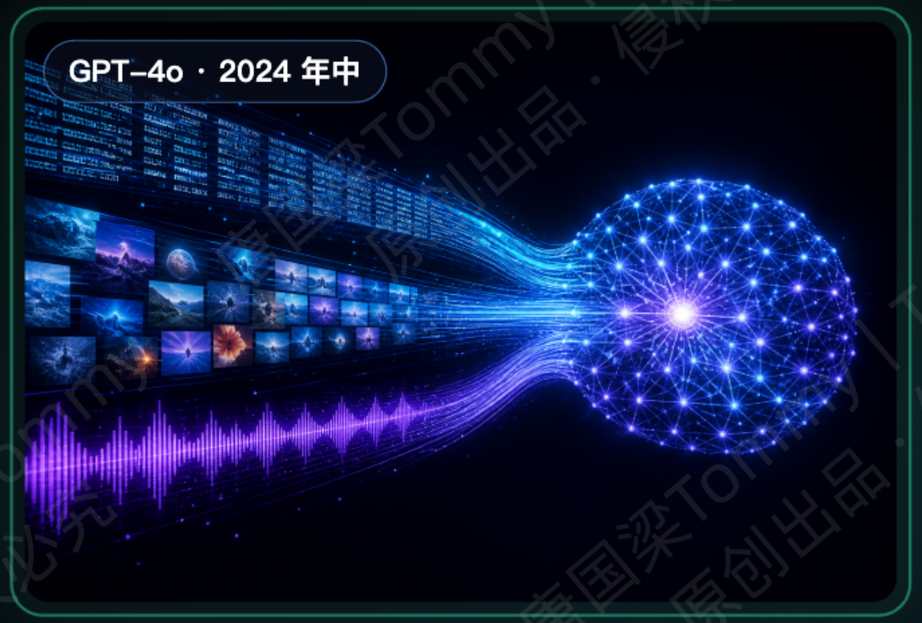


⇄ 编码器与大脑一起联合微调 · 为它服务的大语言模型量身进化

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

第三条线 · 模态拓宽

omni 全模态 = 新旗舰形态



文本 →

视觉 →

音频 输入 + 输出 新进 →

端到端训练进 · 同一个神经网络

omni · 全模态 = 新的旗舰形态

开源阵营陆续跟进（各家方法不一，仅作趋势）· 多图 / 视频数据占比明显上升

💡 多模态不再是语言模型的外设 · 而是从输入层就原生设计进去的能力

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

第三次转折

2025 — 2026

瓶颈

能看清了，但想不深 · 理解/生成两套系统

解法

R1-Zero：纯 RL，学会推理

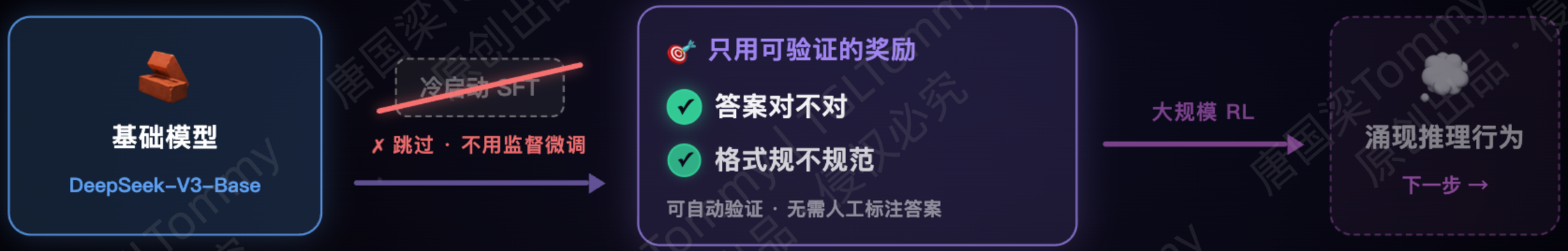
看得清 ≠ 想得深；理解与生成，仍是各管一段的两套系统

➜ 转机来自纯文本领域的一次实验：2025.1 · DeepSeek R1-Zero

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

第三次转折 · 推理与统一

R1-Zero：跳过冷启动 SFT，纯用可验证奖励做 RL



纯文本领域的实验：从基础模型直接开跑强化学习

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

✅ 可验证奖励 ▶ 🔥 大规模 RL ▶ 🧠 长出推理行为

🔗 长链推理

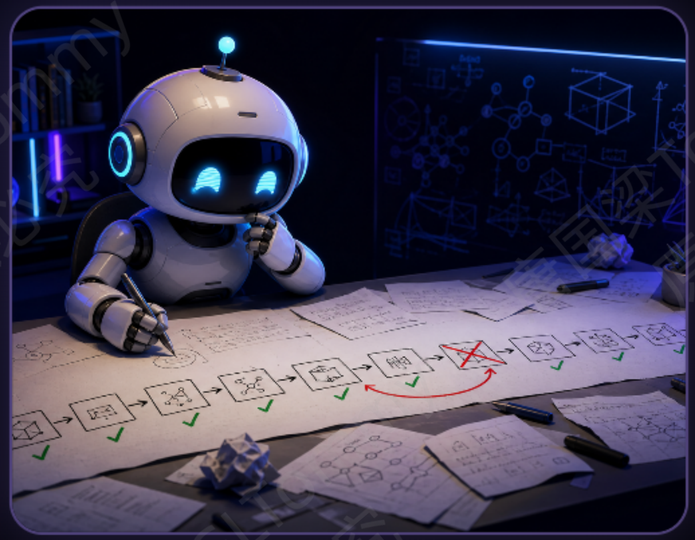
推理链条越拉越长

🔍 自我检查

中途回看、验证自己

↩️ 回溯

走错就退回、换条路



模型在草稿纸上，自己想、自己检查、自己回头

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

一个必须分清的细节

R1-Zero 和 DeepSeek-R1，不是一回事

R1-Zero

展示的结论

基础模型

DeepSeek-V3-Base

✗ 不用冷启动 SFT

直接开跑纯 RL

只靠可验证奖励驱动

✓ 纯 RL 也能强化推理

≠

两条路线

DeepSeek-R1

工程化版本

同一套纯 RL

左边证明可行的那套核心

冷启动 SFT

先用监督微调打底

多轮打磨

对齐 · 迭代调优

✓ 做成一个真能用的模型

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

一个必须分清的细节

R1-Zero 和 DeepSeek-R1，不是一回事

R1-Zero

展示的结论

基础模型

DeepSeek-V3-Base

✗ 不用冷启动 SFT

直接开跑纯 RL

只靠可验证奖励驱动

✓ 纯 RL 也能强化推理

≠

两条路线

DeepSeek-R1

工程化版本

同一套纯 RL

左边证明可行的那套核心

\+ **冷启动 SFT**

先用监督微调打底

\+ **多轮打磨**

对齐 · 迭代调优

✓ 做成一个真能用的模型

完整是四段配方 · 训练那节课再完整梳理

⚠ 别混为一谈：结论 vs 工程配方

R1-Zero 证明「纯 RL 能行」，R1 才是把它做成产品的配方

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

旗舰 内置思考模式

🧠 视觉推理 · R1 风格复现井喷

<think> 推理过程（思考）

答案 ✅

最终答案

短

长 · 思考长度可调



模型草稿纸 · 自我推演

先输出一段 推理过程 → 再给 答案 ， 思考的长度还能调

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

RL 后训练 · 成越来越常见的后训练阶段

2026 起

常规阶段

① 📚

预训练

Pre-training

海量数据打底

常规阶段

② 🧭

指令微调

SFT

学会听懂指令

NEW · 新增

③ 🎯

RL 后训练

RL Post-train

可验证奖励 · 强化推理

强化学习后训练 → 排在 预训练 、 指令微调 之后，越来越常见

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

架构层面的另一条线 · 收敛

理解 与 生成 · 共享同一套表征

共享表征

🧠 理解 ⊕ 生成 🎨

反直觉 · 两种能力不见得打架

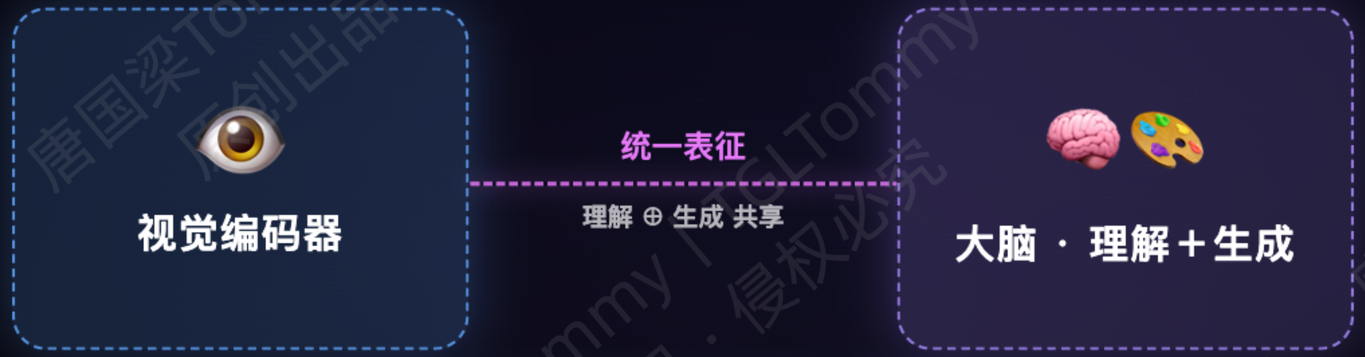
⇄ 观察到互补收益

⚠ 部分研究在各自设置下的结果 · 尚非已收敛的定论

唐国梁Tommy丨TGLTommy 原创出品 · 侵权必究

统一架构 · 从论文走进旗舰实践

三件套的边界 · 开始松动



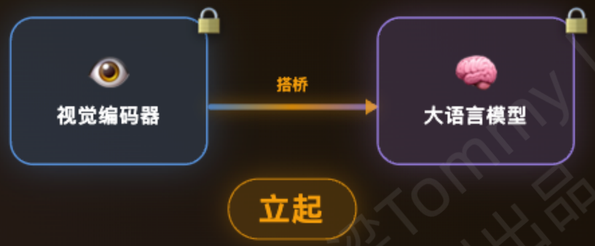
❓ 连「要不要保留独立视觉编码器」都被重新拿出来讨论

竞争焦点 看得清 ↑ 想得深 + 合得拢

唐国梁Tommy丨TGLTommy · 原创出品 · 侵权必究

回看三次变革 · 每次都在做同一个动作： 先指瓶颈 → 再给解法

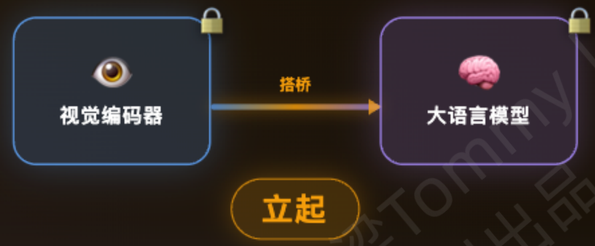
① 转折一 · 外挂拼接　复用两个成熟组件、中间搭桥 → 三件套立起

转折	瓶颈	解法	三件套状态
转折一 2023 – 2024 外挂拼接	从零训练太贵	复用两组件 · 中间搭桥	
转折二 2024 – 2025 原生融合	看不清 · 吃不下真实输入	原生分辨率 · 解冻联训 · 模态拓宽	
转折三 2025 – 2026 推理与统一	想不深 · 两套系统	可验证奖励 RL · 理解生成共享表征	

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

回看三次变革 · 每次都在做同一个动作： 先指瓶颈 → 再给解法

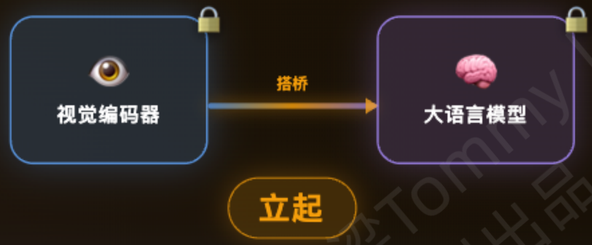
② 转折二 · 原生融合　原生分辨率 · 解冻联训 · 模态拓宽 → 深度融合

转折	瓶颈	解法	三件套状态
转折一 2023 – 2024 外挂拼接	从零训练太贵	复用两组件 · 中间搭桥	
转折二 2024 – 2025 原生融合	看不清 · 吃不下真实输入	原生分辨率 · 解冻联训 · 模态拓宽	
转折三 2025 – 2026 推理与统一	想不深 · 两套系统	可验证奖励 RL · 理解生成共享表征	

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

回看三次变革 · 每次都在做同一个动作： 先指瓶颈 → 再给解法

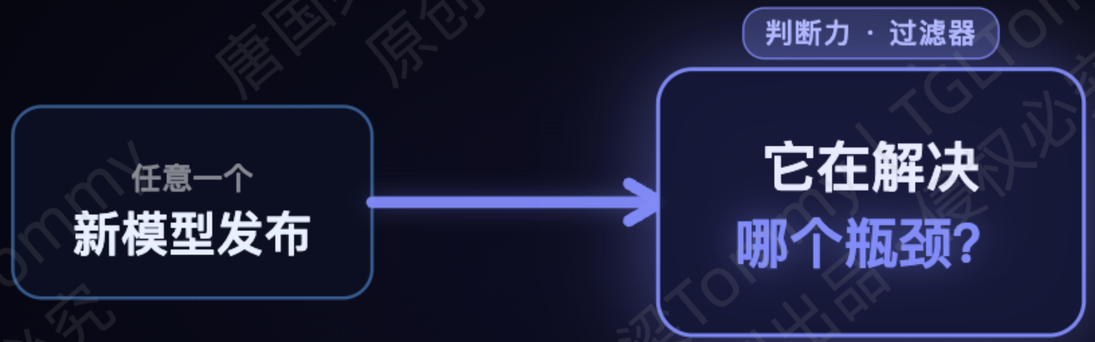
③ 转折三 · 推理与统一　可验证奖励 RL · 理解生成共享表征 → 边界松动

转折	瓶颈	解法	三件套状态
转折一 2023 – 2024 外挂拼接	从零训练太贵	复用两组件 · 中间搭桥	
转折二 2024 – 2025 原生融合	看不清 · 吃不下真实输入	原生分辨率 · 解冻联训 · 模态拓宽	
转折三 2025 – 2026 推理与统一	想不深 · 两套系统	可验证奖励 RL · 理解生成共享表征	

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

这个框架真正值钱的地方

不止解释过去 · 更帮你消化未来



🧭 见到任何新模型发布，先问这一句 —— 这一个问题，就是你的判断力

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

判断力分流

同一个问题，把新模型分成两类



模型名单会一直变长，但 你这份判断力不会过期

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

整门课 · 照三次转折的演进逻辑铺 · 前三章打底



三章学完 ➜ 看懂任何一个三件套模型的设计

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

整门课 · 12 章地图

前 3 章打底 · 后 9 章实战复现

📍 你在这里

第 1 章
全景

第 2 章
眼睛

第 3 章
桥

第 4 章
标本

第 5 章
标本

第 6 章
标本

第 7 章
标本

第 8 章
标本

第 9 章
标本

第 10 章
标本

第 11 章
标本

第 12 章
标本

第 4–12 章 实战：一个一个讲透 · 动手复现代表模型

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

这一节 · 带走三样东西就够了

第一样

一张图

三次转折 · 演进图

第二样

一个判断

瓶颈推着架构走

第三样

一个方法

下一节揭晓 →

转折一

外挂拼接

从零训练太贵 → 复用两组件·搭桥

转折二

原生融合

看不清·吃不下 → 原生分辨率·解冻联训

转折三

推理与统一

想不深·两套系统 → 可验证奖励·共享表征

阶段名记不住没关系 —— 但每次转折被哪个瓶颈逼出来、拿什么解，这条 **瓶颈 → 解法** 的线，**要记牢**

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

这一节 · 带走三样东西就够了

第一样 ✓

一张图

三次转折 · 演进图

第二样

一个判断

瓶颈推着架构走

第三样

一个方法

下一节揭晓 →

一个判断 · 架构演进被瓶颈推着走

真瓶颈 · 推动

每一次范式更替背后，几乎都站着一个具体的、拦住过很多人的工程瓶颈

不解决瓶颈 · 失效

不真正解决瓶颈的所谓创新，往往热闹一阵就过去，很难留下持续的影响

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

收尾 · 带走的第三样

一个方法

③

一个方法 · 判断力（呼应前面「它在解决哪个瓶颈」）

见任何新模型，先别问它多强，先问它在解决哪个瓶颈

1

转折一

外挂拼接

2023 — 2024

复用两组件 · 中间搭桥

2

转折二

原生融合

2024 — 2025

原生分辨率 · 解冻联训 · 模态拓宽

3

转折三

推理与统一

2025 — 2026

可验证奖励 RL · 理解生成合流

阶段名记不住没关系 —— 记牢「被哪个瓶颈逼出来、拿什么解」这条线

唐国梁Tommy | TGLTommy · 原创出品 · 侵权必究

下一节预告

把三件套拆开 · 黑箱变白箱



开场的三个问题，这一讲都回答了

- 看见 · 感知
- 看清 · 细粒度
- 想明白 · 推理

拆开三件套 · 看内部到底发生了什么

一张图喂进去 → 一段话吐出来

~~黑箱~~ → 白箱

拉三个真实旗舰模型当标本 · 当场分析

标本 A　标本 B　标本 C

同样一副三件套，各家到底 差在哪儿

好，我们下一节继续 →

唐国梁Tommy | TGLTommy 原创出品 · 侵权必究

版权声明

唐国梁 Tommy 原创出品｜熵智未来（深圳）科技有限公司版权所有。保留所有权利。

本课程及配套资料，包括但不限于课程视频、PDF 文档、课件、代码、图表、示例项目、知识库文档及其他学习材料，仅供购买本课程的用户本人学习使用。未经版权所有者书面许可，任何个人或机构不得以任何形式对本课程及配套资料进行录制、复制、传播、上传、公开分享、分发、转售、改编、翻译、汇编，或用于商业培训、企业内训、商业化产品、知识库建设、数据集构建、人工智能模型训练、微调及其他商业用途。如发现侵犯本课程及配套资料相关权益的行为，权利人有权依法要求停止侵害、删除下架、赔偿损失，并追究相应责任。

本课程及配套资料中涉及的第三方开源软件、模型、代码、数据集、商标、图片及其他素材，仅在相关许可证允许范围内进行教学展示，其相关权利归原权利人所有。

侵权举报 / 授权合作：legal@tgltommyai.com

微信公众号：唐国梁TGLTommy｜官方网站：TGLTommy.com（国内访问需科学上网）