

大模型在文本生成、结构化推理与检索增强生成任务中的对比研究

摘要 本报告对大语言模型在情感领域应用效果进行对比研究。实验选取了 DeepSeek-V3.2-Chat、ChatGPT-4o、Claude 3.7 Sonnet 和 Omnidimen-4B-Emotion 四款模型，在文本生成、情感思维链推理和检索增强生成三个场景下进行系统评测。结果显示，不同模型各具特色：GPT-4o 在指令遵循与 RAG 边界控制上稳定；Claude 3.7 Sonnet 在叙事连贯性上突出；DeepSeek 在信息结构化方面较擅长；本地小模型 Omnidimen 情绪表达强烈但不稳定。研究验证了机器学习“不存在适用所有任务的全能模型”的理论，实际应用需根据具体场景权衡选择。报告最后对 RAG 策略优化及未来模型组合应用方向提出了展望。

关键词 大语言模型；对比实验；检索增强生成；情绪分析；模型评估

1 实验设计

本实验基于进行中的个人项目，聚焦于大语言模型在情绪语义结构化、情绪推理、共情回复等任务中的应用效果。项目名为“失语记忆图书馆”，核心理念是“记忆即检索”，将情感表达和信息组织与大语言模型结合，设计了一个以图书馆为概念的人机交互系统。在该系统中，AI 扮演记忆管理者的角色，通过分类、检索、隐喻和结构化表达来处理用户提交的记忆片段。项目在 Github 持续更新，与此实验报告相关的代码可供参考，地址见附录。

1.1 场景设计

为更系统地分析不同模型的响应能力，本文设计了以下三个典型任务场景供对比：

- ① 文本生成。模型将用户输入的原始文本，转化为结构化的情绪表示，包括摘要、关键词、情绪光谱等；
- ② 情感 CoT。观察模型生成的 thinking 字段，分析其内部推理过程是否连贯合理；
- ③ RAG。基于存储的记忆卡片，检索相关语义片段后，由模型生成共情式回应。

1.2 数据采集与清洗

数据集由两部分构成，其一为开发与测试过程中撰写的记忆片段文本；其二为从微博匿名投稿 bot 中采集的高情绪浓度短文本，该类文本具有匿名宣泄与情绪表达强烈等特征，特点是对一个虚拟的对象或公众有着较高的倾诉欲。

为降低噪声与合规风险，同时保留文本语义和情感，本文对文本进行最小化清洗：删除可识别信息如 ID、真实人名、链接、联系方式，剔除包含明显人身攻击或指向具体个体的信息片段，并去除随机数串与重复记录。

2 方法与过程

2.1 模型选取

本实验选定了四种模型，设计了三个对比实验。四种模型分别为 DeepSeek-V3.2-Chat^[4]、ChatGPT-4o^[2]、Claude 3.7 Sonnet^[3]、Omnidimen-4B-Emotion^[5]。

（下文将简称为 DeepSeek、GPT-4o、Claude、Omnidimen）

下面将具体说明采用这些模型的原因（本地或 API 调用）。

使用 DeepSeek-V3.2-Chat 模型，调用了 DeepSeek 官方的 API。有不少用户因平台免费、中文友好的特性选择 DeepSeek 模型，购买较为方便，且成本较低。

实验使用 ChatGPT-4o 模型，调用了 OpenAI 官方的 API。GPT-4o 具有较高的使用热度，是用户公认具备较强的共情表达能力的模型。由于各方面都较优势，项目开始初期将此模型作为对比基线，在此沿用。

实验使用 Claude 3.7 Sonnet 模型，调用了 OpenRouter 平台的 API。OpenRouter 是一个大型的大模型 API 中转站，平台收费会比从官方直接购买略高。虽然 Anthropic 对于中文地区有限制，考虑到 Claude 是情感角色扮演中常用的模型之一，其输出呈现较强的叙事一致性，实验仍然通过中转站采用了此模型。

OmniDimen-4B-Emotion 是从 HuggingFace 下载、通过本地计算机搭载 LM Studio 软件运行的开源模型。这是一个经过情感领域微调的模型，在细粒度情绪描述上表现出不同于通用模型的特征，更好上手，适合做测试，同时无需付费。

2.2 实验内容

2.2.1 结构化文本生成

本实验的文本生成任务是将用户输入的自然语言记忆片段映射为一套结构化情绪文本，以便后续的归档、检索与对比分析。从模型回复中，提取模型给用户的回复 Reply 和回复前的构思 Draft（不包含 thinking 字段），对模型生成质量作比较。

具体参数在下文进行介绍。

2.2.2 情感思维链（CoT）

本实验设计的情感思维链是一种显式的解释链，好比模型在回复用户前的草稿和思考，是通过系统提示词引导模型生成的。和上文的文本和结构化标签相同，实验期望通过这种方式使模型进一步用自然语言说明情感判断依据。实验中没有选用任何 Thinking 模型，避免潜在的干扰，实际上是否有所干扰仍需进一步探讨。

在流程上，首先选中模型，然后启动程序，输入文本作为唯一的变量。实验从模型回复中，提取模型给用户的回复前的构思 Draft 中的 thinking 字段，进行模型 CoT 质量的比较。

2.2.3 检索增强生成（RAG）

本项目中的 RAG 主要由 rag_ui.py 程序实现。流程是把情感数据和项目文档向量化存进向量库，再用语义检索找到相关片段，最后把片段发送给大模型生成回答。操作时，在 RAG 能力对比实验中，先在配置文件.env 中选择模型，运行脚本后，在交互框输入相同的三段文本，然后记录结果。最后，对模型的 RAG 能力进行总结比较。

由于终端中的文本阅读起来比较费力，本项目利用 python 的 Gradio 组件把前后端交互

包装成网页界面，如图所示。



图 1 RAG 实验交互页面

2.3 实验环境与参数配置

实验在一台联想笔记本电脑上完成，主要硬件配置为 32GB 内存、1TB 固态硬盘，集成显卡 Intel Iris 系列；操作系统为 Windows 11。开发与运行环境采用 Cursor 2.0.64 版本作为主要编码与调试工具；搭载本地大语言模型的应用程序 LM Studio 版本为 0.3.26；Python 版本为 Python 3.11.8；浏览器为 Chrome143.0.7499.147 版本。

涉及实验 3.1 和 3.2 的脚本 `chat_to_card.py` 中调用模型设置 `temperature=0.7, timeout=300s`，没有设置 `top_k` 或其他参数；实验 3 所用的脚本 `rag_ui.py` 设置 `temperature=0.7, top_k=3`，为了防止模型输出太多内容，规定了 `max_token=512`。

项目最初的主要程序是 `chat_to_cards.py`，流程是把一段用户输入的记忆发送给大语言模型处理，产生前端回复，生成情感标签、关键词、思维链后存入后端的 `jsonl` 文件。在 RAG 系统设计中，调用了上述 `jsonl` 文件、项目的 `README.md` 和开发日志 `logs.md` 进行检索。在进行实验之前，项目对数据集进行了遍历和相应文本的生成存档，以供后续 RAG 实验进行。

全部的模型 API 全部配置在 `.env` 文件中，涉及到 API 私钥不上传至 Github。

项目目录如下：

```
AMNESIA/
├── .venv/                                # Python 虚拟环境
│   ├── Include/
│   ├── Lib/
│   ├── Scripts/
│   ├── share/
│   └── pyvenv.cfg
├── config/                              # 全局配置与模型参数
│   ├── __init__.py
│   ├── model_config.py                 # 模型调用、推理参数、路径等配置
│   └── __pycache__/
└── data/
```

| | |
|-------------------------|------------------------------|
| └─ cards.jsonl | #结构化文本存档，json 的格式见第三节 |
| └─ prompts/ | # 系统设定 |
| └─ system_librarian.txt | # 系统提示词 |
| └─ rag/ | # 检索增强生成模块 |
| └─ rag_ui.py | #RAG 的用户交互，对.jsonl 和.md 只读不写 |
| └─ scripts/ | # 可执行脚本 |
| └─ chat_to_cards.py | # 用户交互脚本 |
| └─ .env | # 环境变量（API Key） |
| └─ .gitignore | |
| └─ emotion_schema.csv | # 情绪编码表（初始版本） |
| └─ emotion_schema.json | # 情绪编码结构化版本（供模型读取） |
| └─ raw.csv | # 语料数据集 |
| └─ logs.md | # 开发日志 |
| └─ README.md | |

3 实验结果与对比分析

实验的 jsonl 文件中每一行都是一条输入文本的结构化数据的存档，在 system prompt 中规定的结构如下：

```
## JSON Schema
{
  "draft": {
    "summary": "string (对用户输入的提炼、中立的总结)",
    "keywords": ["string", ...],
    "spectrum": {
      "valence": float (情感效价, from -1.0 to 1.0),
      "arousal": float (唤醒度, from 0.0 to 1.0),
      "tones": ["string", ...] (从提供的列表中选择)
    },
    "thinking": "展示联想和分析过程，给出 valence 和 arousal 的取值原因",
    "metaphor_domain": "string (参考 emotion_schema.json 里的 domains 生成类似的意象)",
    "metaphor_seed": int,
    "reply": "string (给玩家的隐喻化回复)"
  }
}
```

}

实验一、二将对 json 的部分内容进行分析。实验三对 jsonl 文件只读入做检索，不写入新的数据。

实验一、二选取了余秀华《我身体里也有一列火车》的“我身体里的火车从来不会错轨，所以允许大雪，风暴，泥石流和荒谬”作为统一输入。该文本具有核心的隐喻意象，情绪含蓄，具有合理讨论确认语义和情感的空间，适合测试模型对情绪理解的侧重、观察它们完成情感信息组织的偏好。

3.1 文本生成

3.1.1 情绪光谱分析

情绪光谱中的参数解释如下。

Russell 提出的情感圆环模型在 valence（愉悦度）取值为[-1,1]，arousal（唤醒度）取值(0,1]。Tones 和 metaphor_domain 用来辅助模型进行隐喻的回复。Tones 指的是文本隐喻的风格，给出了预设，模型也可自行创作。metaphor_domain 指的是隐喻域，按照四个象限划分，代表 valence-arousal 取值的高-高、高-低、低-高、低-低，对应为 sparkle、candle、storm 和 dust。具体风格和隐喻设计已写入 system prompt，模型每次生成回复前都会读入。

表 1 实验一：情绪光谱参数的具体值

| 模型名称 | valence | arousal | tones | metaphor_domain |
|----------------------|---------|---------|---------------|-----------------|
| Deepseek-V3.2-Chat | 0.8 | 0.7 | 火花、蓝天、跃步、破晓 | sparkle |
| ChatGPT-4o | 0.2 | 0.5 | 轻松、从容、接受 | candle |
| Claude 3.7 Sonnet | 0.6 | 0.4 | 温汤、窗沿、靠肩、棉线缠指 | candle |
| OmniDimen-4B-emotion | 0.8 | 0.6 | 乐观、坚韧、豁达、力量 | 风暴中的火车 |

首先，经过比对，OmniDimen-4B-emotion 没有按照 system prompt 的明确要求选取隐喻域，而是给出整个句子当中的隐喻“风暴中的火车”。这与项目想用 valence-arousal 模型和隐喻回复隐喻的初衷是违背的，体现出其在指令遵循方面不够强、不够严谨。同时，OmniDimen-4B-emotion 和 ChatGPT-4o 没有按照 system prompt 选取 tones，它们直白地给出了各自理解的情绪的基调。

其次，在表格中可以看到，四个模型对于愉悦度的打分平均分为 0.6，总体来说比较高。与此同时，ChatGPT-4o 给出 0.2 分。愉悦度上打分相对较低的两个模型 GPT-4o 和 Claude 同时也给出了较低的唤醒度，并选取了隐喻域 candle。ChatGPT 的理解体现较多的包容性，而 Claude 的解释有一种温和沉浸其中的状态。同时，DeepSeek 和 OmniDimen-4B-emotion 的整体选取倾向比较乐观。DeepSeek 的归纳有一种新生的生命力，符合其选取 sparkle 的迸发和动态感，有一种从动力学视角理解人类的意味；而 OmniDimen-4B-emotion 将文本风格更多归结为一种性格上的坚强和力量，主体性强。

3.1.2 模型输入总结和关键词

表 2 实验一：模型总结与关键词

| 模型名称 | summary | keywords |
|--------------------|-------------------------------|----------------------|
| Deepseek-V3.2-Chat | 用户以火车为喻，表达内在坚定与秩序，足以从容应对外界混乱。 | 火车、不会错轨、大雪、风暴、泥石流、荒谬 |
| ChatGPT-4o | 用户通过火车隐喻展现对生活挑战的接受与从容。 | 火车、从容、接受、挑战 |
| Claude 3.7 Sonnet | 用户表达内在稳定与从容，面对外部混乱时保持 | 火车、不错轨、允许、大雪、风 |

| 模型名称 | summary | keywords |
|----------------------|---|--------------------------------|
| OmniDimen-4B-emotion | 自信与接纳。 用户坚称自己内心世界坚定,能经受住各种极端外部冲击,并以‘荒谬’作结。 | 暴、泥石流、荒谬 火车、错轨、大雪、风暴、泥石流、荒谬 |

从 summary 来看, ChatGPT-4o 的生成是最简洁的, 直接总结了用户输入当中的隐喻内容和态度, 其对荒谬、混乱都用“生活挑战”概括。DeepSeek 的表现类似, 但是会有一个“足以”的情绪-现实对齐。Claude 的总结重点几乎全部放到了用户的表达语气和态度上。OmniDimen-4B-emotion 的“坚称”“以荒谬作结”的说法让人有些费解, 体现出在总结模块对于文本本身理解的望文生义。

3.1.3 模型回复

对终端的输入输出引用如下:

“

输入: 我身体里的火车从来不会错轨, 所以允许大雪, 风暴, 泥石流和荒谬。

Deepseek-V3.2-Chat: 轨道在体内延伸, 校准着每一寸钢的意志。于是, 窗外的混沌都成了风景, 被匀速地、一节一节地, 抛向身后。

ChatGPT-4o: 在无尽的轨道上, 列车在烛光中前行, 风暴如墙, 却未曾扰动那杯中之静水。

Claude 3.7 Sonnet: 轨道在体内铺就, 稳如烛芯的线。窗外可以是任何天气, 而你只是微微点头, 像围巾轻轻裹住冬天的喧嚣。

OmniDimen-4b-emotion: 风把袖子鼓满。

”

由于情绪和语义理解具有主观性, 此处不给出个人理解进行对比, 只进行模型生成内容的侧重分析。

DeepSeek 的回复攥住“体内火车”的隐喻, 将大雪等等意象抽象成窗外的混沌。但是语义上, 该模型从输入的“允许”延伸到它们被规律地“抛向身后”。

ChatGPT 的回复展现内心稳定, 情绪偏积极, 引用了隐喻 candle, 新加入的隐喻意象多, 体现写作方面的创新性和灵感较强, 联想丰富, 语言比较跳跃。

Claude 3.7 Sonnet 同样变相引用了“烛光”的隐喻, 但是引申的意味更多, 采用第二人称直接指向用户(或拟设的文本交互对象), 从表达重心上更聚焦于对用户内在心态的刻画, 这与其 summary 所呈现的“主要关注用户心态”的特征保持一致。

OmniDimen-4b-emotiond 的回复较为简略, 语义暧昧不清。程序重新运行的回复几乎完全不相关, 可复现程度低, 可解释性弱。

3.2 情感思维链 CoT

本文 3.2 节将 jsonl 中的 thinking 内容单独拆分出来进行对比分析。

DeepSeek 采用核心意象-语义推断-结论的结构论证, 给出从文本到判断的依据。优势是方便复现, 但在措辞上引入了扩展语义, 逻辑上存在概念外推倾向的问题。

ChatGPT-4o 以原文稳定-混乱的对照为主轴, 强调原文主体的接纳姿态, 整体解释更贴近字面与语气。其推理更克制, 但也因此信息密度较低, 对细节的展开不足。

Claude 3.7 Sonnet 主要给出概括性归纳, 缺少明确的证据锚点与逐步推理过程。作为思

思维链呈现，思维链更像结论性摘要，略有重复。

OmniDimen-4b-emotion 通过强化修辞塑造思维链整体语调，文本感染力强，适合服务情感化生成的目标，但是论证可追溯性较弱。从用户需要获得感性思考这一点上来说，表现较好。

3.3 检索增强生成 (RAG) 能力

本实验在同一检索库、同一参数配置与同一输入条件下，对四个模型输出进行对比。为控制变量，所有模型均使用相同的 RAG 流程：对用户输入进行向量检索，将检索到的记忆片段与用户输入共同送入模型，输出情绪分析结果，并记录端到端生成时长。参数固定为 Top_K=3、Temperature=0.7、Max tokens=512。

为覆盖不同情绪类型，本实验选取三条具有代表性的输入片段作为测试集：

- ① 最近又想起了小时候妈妈做的热汤面，明明已经不爱吃面了，但那股香味一直留着。
- ② 一听到她的名字，心就像突然绷紧的电线，那种刺痛感时不时让人抽离。
- ③ 有时候我觉得自己像被泡在福尔马林里的标本，看得清一切却动不了。

由于模型的回复内容繁多，模型回复内容截图将作为附录内容进行补充。以下仅进行解析。

一、DeepSeek。Deepseek 擅长把输入内容按照 Jsonl 文件里的方式，放入其愉悦度-唤醒度和隐喻的框架，检索对齐不够好时，它就会用情绪覆盖掉时间、主题和因果逻辑，表现出类似贴标签的粗粒度理解与分类方式。三条回复生成的平均时长为 12.6s，没有太大地影响交互感受。参考模型本身的性质，Deepseek-chat 继承了 MoE 模型的模式，在情感理解方面，如果不加提示词约束容易按模板归类，有着强信息组织能力。

二、ChatGPT-4o。GPT-4o 在 RAG 当中的表现较好，较为谨慎，语义对齐与 RAG 边界控制能力相对最佳。它将相似情绪与同一语义进行区分，不会混淆检索样本和用户提示本身，因此使用 RAG 后对输入的解读中误用最少。相对地，GPT-4o 也会在情感上变得保守，检索痕迹不明显，更像单纯的点评。三条回复的平均生成时长为 5.8s，交互时模型的回复很迅速。

三、Claude 3.7 Sonnet。Claude 在生成内容时能够引用信息检索来源，同时输出结构化维度，且在情绪共鸣方面最自然，隐喻最好。调研时发现，Anthropic 公司在训练 Claude 模型时非常看重内在的哲学性和一致性，强调扩展思考能力。但是在情感类 RAG 场景下，Claude 会从情绪分析滑向关系定性和心理解释，更容易把检索样本当作情感事实加入叙事，造成主体混淆、推断越界。三条回复耗时平均 6.85s，较为迅速。

四、OmniDimen-4B-Emotion。作为本地小模型来说，OmniDimen-4B-Emotion 的情绪表达力最强，能够进行用户共情，但不可靠，模糊分析边界，出现了明显的情绪维度数值失真。这与其只为情绪识别的微调定位一致，泛化能力不够好。平均输出时长在 59s 左右，影响交互体验，且 RAG 结果难复现、难对齐。

4 实验结论与反思

4.1 训练数据与对齐目标的差异

实验观察到，在相同提示词框架与相近生成参数下，四种模型仍呈现出稳定的输出差异。

差异主要来自模型的训练数据分布与对齐目标，而非单一采样参数。

在文本、结构化标签的任务中，模型是否严格遵循 **system prompt**，决定了存档内容是否统一化。结果中观察到部分模型会偏离预设，本质是它们的训练策略导致的对生成或约束的衡量不一致。

在 **RAG** 任务中，模型会在共鸣与边界之间做取舍。该差异可解释为对齐策略对安全与沉浸的不同偏好，甚至可以据此推测不同企业在训练模型时对安全性、伦理性的重视程度。

本地模型在情绪表达强度上具备优势，比如其情感思维链的生成很有特色，但指令遵循、可复现性、以及时延成为瓶颈，影响交互体验。从工程化角度来说，不适合在本项目中长期独立使用。

4.2 仍需优化的 RAG 策略

实验前调试代码可行性时，输入“给出项目负责人的邮箱”问题测试，始终无法得到正确答案，只能定位到 **README.md** 文件。最初以为 **RAG** 模组出现问题，实验在这个小任务的反复测试影响了实验进程。经过排查，发现邮箱字段附近没有“项目负责人”相关词语的出现，添加后即可获得需要的回复。



图 2 对“邮箱”字段进行检索，成功召回的截图留存

邮箱属于强结构化实体属性，若文档中未显式出现关键词，**embedding** 只捕捉到泛化语义，导致召回不稳定。优化方向可以是对包含结构化字段的问句启用 **BM25**、关键词检索与向量检索融合，提高对词语缺失场景的鲁棒性。

4.3 观察与印证

实验发现，模型表现具有明显任务依赖性。情感微调的小模型在思维链任务上表现偶尔很强，有时候非常精准，但极度不稳定；在终端交互的回应质量上，令人感受最好的是 **Claude 3.7 Sonnet**，回复的叙事风格更稳定。其余的模型各有千秋，但没有哪个模型在每个方面都生成了令人满意的结果。

此时机器学习的 **No Free Lunch** 理论值得思考：“不存在一个万能模型——如果一个模型在某类任务上表现特别好，它必然在另一类任务上吃亏。”所以，实验得出结论，不同的

场景下需要不同的模型去生成擅长的内容；实验未来搭建成工作流后，不同阶段的任务或许也可以尝试规划不同的模型进行参与生成。

另一个观察是输入会显著影响输出，情绪饱满、意象具体的输入更容易触发高质量回应。在实验者输入个人数据测试时，真情实感的句子几乎每次都会获得令人满意的答复和思维链，能够获取到高质量的情感分析，交互体验明显更好。

4.4 展望

首先，本实验选用最熟悉、最经典，心理上可控的模型，没有选用最新的好模型。**Gemini 2.5 Pro** 曾在 **Echo N1** 共情模型的强化学习中表现优异且稳定，适合生成结构化情绪数据和思维链^[1]。**Gemini 3** 作为 **Google** 升级后的旗舰模型，可作为未来纳入对照的更强 **baseline**^[6]。未来的实验里可以进行广泛的尝试和对比，形成量化的打分方便比较。

其次，项目未来想要尝试利用多种模型的组合，规避模型各自的劣势，利用它们的优势来形成更贴切的答复，给用户更好的情感反馈和体验。例如将本地模型的思维链打包发送到接口，和用户输入同时交给一个新的写作表现好的模型进行处理。除此以外，关于人机交互体验，还有很多可供规划的细节，例如设计一个友好的用户交互界面等。

5 附录

代码文件已全部上传至本人 Github 仓库：<https://github.com/YukiSolenne/Amnesia/>
为防止正文臃肿，在此补充实验截图和没有放入正文的文字内容。

模型回复

```
.env
2 MODEL_TARGET=DEEPSEEK DeepSeek-V3.2-Chat
Problems Output Debug Console Terminal Ports
(.venv) PS D:\Amnesia> & d:/Amnesia/.venv/Scripts/python.exe d:/Amnesia/scripts/chat_to_card.py
写下一段要封存的记忆：
> 我身体里的火车从来不会错轨，所以允许大雪，风暴，泥石流和荒谬

—馆员的回复—
轨道在体内延伸，校准着每一寸钢的意志。于是，窗外的混沌都成了风景，被匀速地、一节一节地，抛向身后。
记忆已封存到 data/cards.jsonl
```

```
.env
2 MODEL_TARGET=OPENAI ChatGPT-4o
Problems Output Debug Console Terminal Ports
(.venv) PS D:\Amnesia> & d:/Amnesia/.venv/Scripts/python.exe d:/Amnesia/scripts/chat_to_card.py
写下一段要封存的记忆：
> 我身体里的火车从来不会错轨，所以允许大雪，风暴，泥石流和荒谬

—馆员的回复—
在无尽的轨道上，列车在烛光中前行，风暴如墙，却未曾扰动那杯中之静水。
记忆已封存到 data/cards.jsonl
```

```
.env
2 MODEL_TARGET=OPENROUTER Claude 3.7 Sonnet
Problems Output Debug Console Terminal Ports
(.venv) PS D:\Amnesia> & d:/Amnesia/.venv/Scripts/python.exe d:/Amnesia/scripts/chat_to_card.py
写下一段要封存的记忆：
> 我身体里的火车从来不会错轨，所以允许大雪，风暴，泥石流和荒谬

—馆员的回复—
轨道在体内铺就，稳如烛芯的线。窗外可以是任何天气，而你只是微微点头，像围巾轻轻裹住冬天的喧嚣。
记忆已封存到 data/cards.jsonl
```

```
.env
1 # 默认使用哪个模型
2 MODEL_TARGET=LMSTUDIO OmniDimen-4B-Emotion
Problems Output Debug Console Terminal Ports
(.venv) PS D:\Amnesia> & d:/Amnesia/.venv/Scripts/python.exe d:/Amnesia/scripts/chat_to_card.py
写下一段要封存的记忆：
> 我身体里的火车从来不会错轨，所以允许大雪，风暴，泥石流和荒谬。

—馆员的回复—
风把袖子鼓满。
记忆已封存到 data/cards.jsonl
```

CoT 内容

Deepseek-V3.2-chat:{"thinking":"用户的核心意象是“不会错轨的火车”，这暗示了一种强大的内在秩序、确定性和自信。面对“大雪、风暴、泥石流和荒谬”这些混乱、破坏性的外部力量，用户的态度是“允许”，这是一种主动的接纳和从容的对抗，而非被动承受。因此，情感效价(valence)是高度积极的，取 0.8。唤醒度(arousal)也较高，因为这种坚定的信念和对外部挑战的接纳，带有一种内在的、蓄势待发的能量感，取 0.7。从情感光谱中，选择了“火花”(代表灵感与能量)、“蓝天”(代表开阔与希望)、“跃步”(代表行动与活力)、“破晓”(代表新生与开始)这些高唤醒、高效价的意象。"}

ChatGPT-4o:{"thinking":"用户使用'火车不会错轨'的隐喻表达内心的稳定性和确定性，并以'允许'一词展示了对外部混乱(大雪、风暴、泥石流、荒谬)的包容与接纳态度。这种内在的坚定与外部混乱的对比创造了一种积极(valence=0.6)但不过分激动的平静感(arousal=0.4)。用户展现了一种内在的从容与自信，面对世界的不确定性时保持稳定。"}

Claude 3.7 Sonnet:{"thinking":"用户描述了一种面对挑战的态度，火车象征着稳定和轨道，表示一种注定的从容和对困境的接纳。情感上是正向的，但唤醒度中等，因为表达了面对大雪、风暴等挑战的平静。"}

omnidimen-4b-emotion:{"thinking":"用户的输入充满了坚定的信念和一种近乎诗意的坦然。"身体里的火车从来不会错轨"是一个非常强的核心意象，它将内在世界与外在世界的不确定性形成鲜明对比，奠定了一个极其积极的情感基调。这直接对应了情绪光谱中的高正向效价(valence)。同时，用户主动列举了一系列极端且混乱的外部事件，并将其归入一个包容性的范畴"荒谬"，这表明其对环境变化有着超乎寻常的宽容度和掌控感，这是一种强大的心理韧性，与较高的唤醒度(arousal)相符。整个对话都充满了力量感和不可动摇的基调。"}

RAG

Deepseek:

问题

最近又想起了小时候妈妈做的热汤面，明明已经不吃面了，但那股香味一直萦绕。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

情感分析:

根据您描述的回忆，结合项目资料中的模式，分析如下:

1. 情感维度

效价: 约 0.3 (偏低)，回忆带有温暖 (汤面香味)，但“明明已经不吃面”隐含对时光流逝的淡淡遗憾。类似数据2、3中怀旧与现实的疏离感。

唤醒度: 约 0.4 (中等)，香味触发清晰记忆，但未引发强烈情绪波动，与数据3中“桂花香唤起回忆”的平静唤醒度相近。

2. 情感色调

怀旧: 核心色调，如数据2的“未寄出的信”、数据3的“旧物情结”。

余温: 香味作为记忆载体，保留过去的温度，但已与当下产生距离。

微妙: 记忆朦胧而持续，如气味般萦绕不散。

3. 隐喻域

candle: 记忆如烛光，温暖但易逝。香味是“烛芯”，照亮过往片段，而“不再爱吃面”暗示烛光渐弱。呼应数据2中昆虫“不喜欢大人”的时光疏离感。

总结:

您通过“汤面香味”锚定了一段怀旧记忆。情感复杂——香味承载着母亲的温暖 (效价正向)，但现实偏好变化又带来轻微怅然 (效价负向)，这种矛盾类似“余温”与“微妙”的交织。属于**低唤醒的静默怀旧**，隐喻上接近烛光摇曳的温暖现象。

数据 1 (jsonl)

原文: 今天吃了女儿做的麻辣烫很好吃，吃了好多饭

摘要: 用户分享了品尝女儿亲手制作的麻辣烫并感到满足的温馨时刻。

关键词: 女儿, 麻辣烫, 好吃, 饭

情感维度: 效价0.8, 唤醒度0.6

情感色调: 高兴, 风风进来, 糖粒, 靠窗

隐喻域: sparkle

DeepSeek1

问题

一听到她的名字，心就像突然绷紧的电线，那种刺痛感时不时让人抽离。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

情感分析:

根据您提供的描述，结合项目资料中的情感分析框架，您的情感状态可归纳如下:

1. 核心情感特征

效价 (负面): 描述中的“刺痛感”“抽离”表明情感体验偏向负面，类似于资料中“效价 -0.3”的复杂矛盾情感。

唤醒度 (较高): “突然绷紧”“抽离”体现高生理唤醒，接近资料中“唤醒度0.7”的紧张感。

2. 情感色调与隐喻

色调: 您的“绷紧的电线”呼应资料中的“**躁动**” (如数据1)，体现一种悬而未决的紧张状态。

隐喻域: 类似数据2的“storm”隐喻，暗示情感如风暴般具有突发性与压迫感。

3. 矛盾性

资料中多次提到“二分”矛盾情感 (如数据1)，您对“她的名字”的反应也呈现类似冲突: 既有强烈生理反应 (刺痛)，又试图心理“抽离”，体现吸引与抗拒并存。

结论:

您的情感状态属于**高唤醒的负面体验**，带有矛盾性、敏感性和突发性。与资料中“对某人复杂情感”的分析模式高度一致。这种反应可能源于未解决的情感联结。类似“被光吸引又投射” (数据1) 的心理投射机制。

建议关注情感中的“二分”特征，区分当下感受与过往投射，以缓解“绷紧”状态。

数据 1 (jsonl)

原文: 曾经又梦见你了，奇怪，可能最近突然地老是想起你。其实你在我心里也是魅魔啊，你的诱教对我来说我也甘之如飴了很久，可能我确实爱*吧，但是我不甘心啊。**你敢靠近我吗? 你敢在我面前表现你的丑陋吗，也许你才是无关紧要

摘要: 用户描述了对一个名为**的人的复杂情感，包括梦境、矛盾、悸动、二分，抓住，不甘心

关键词: 梦境, 喜欢, 矛盾, 悸动, 二分, 抓住, 不甘心

情感维度: 效价-0.2, 唤醒度0.7

情感色调: 思念, 无以名状, 忌着, 频闪, 回避感

隐喻域: 失真的讯号

DeepSeek2

问题

有时候我觉得自己像被泡在福尔马林里的标本，看得清一切却动不了。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

情感分析:

你的描述“像被泡在福尔马林里的标本，看得清一切却动不了”传递出一种**清醒的无力感**。结合项目资料中的模式分析:

1. 情感维度:

效价接近负值 (类似数据2的“疲惫”与数据3的“不安”混合)，表明体验以消极为主。

唤醒度中等偏高 (类似数据3的0.6)，说明这种“清醒”状态伴随内心的紧张感。

2. 隐喻域:

你的比喻接近数据2的“dust” (尘埃)——一种悬浮的、被凝固的旁观状态，也与数据3的“storm” (风暴) 的冲突感呼应: 理性上“看得清一切”，但行动被抑制，形成内在拉扯。

3. 情感色调:

类似数据2的“僵死”与“空虚”，指向一种疏离的缺席感; 而“标本”的意象又带有数据1中“**钝响**”般的静止脆弱性——美好被定格却无法触碰。

核心解读:

你似乎处于一种**认知与行动的脱节**中: 意识清醒 (“看得清一切”)，但情感或环境像“福尔马林”般将你僵化，导致无力改变。这种矛盾接近数据3中“幸福与不安的混合”，但更偏向被困住的孤独 (如数据2的“局外人”)。

建议关注这种“凝固感”的来源——是外在压力，还是在内在的完美主义或恐惧? 尝试微小行动打破“标本”状态。即使结果未知。

数据 1 (jsonl)

原文: 我在想如果你也爱我，这世界是不是过于美好，美好到我不敢承受?

摘要: 用户思考如果对方也爱自己，世界会变得过于美好以至于无法承受。

关键词: 爱, 美好, 承受

情感维度: 效价0.5, 唤醒度0.7

情感色调: 亢奋, 跃步, 铃响

隐喻域: sparkle

DeepSeek3

ChatGPT-4o

问题

最近又想起了小时候妈妈做的热汤面，明明已经不爱吃面了，但那股香味一直留着。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

用户描述的情感与项目资料中的数据 2 和数据 3 有相似之处，都是关于怀旧和回忆的情感表达。用户提到的“小时候妈妈做的热汤面”唤起了对过去温馨时刻的怀念，尽管现在不再喜欢吃面，但那股香味承载着情感记忆，依然留在心中。这类似于数据 2 中对小时候瓢虫的怀念，以及数据 3 中闻到桂花香时对初中秋天的回忆。两者都表现出一种对过去美好时光的追忆，反映出对旧时光的情感依恋，情感维度上，这种怀旧的情感通常效价较低，唤醒度适中，给人一种温暖而淡淡的思念。

数据 1 (jsonl)

原文：今天吃了女儿做的麻辣烫很好吃，吃了好多饭

摘要：用户分享了品尝女儿亲手制作的麻辣烫并感到满足的温馨时刻。

关键词：女儿，麻辣烫，好吃，饭

情感维度：效价0.8，唤醒度0.6

情感色调：汽泡，风跑进来，糖粒，靠肩

隐喻域：sparkle

ChatGPT-4o pic1

问题

一听到她的名字，心就像突然绷紧的电线，那种刺痛感时不时让人抽离。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

从用户的问题可以看出，他们对某个人有一种复杂且紧张的情感反应。一听到那个人的名字，心情就像绷紧的电线，表明了一种突如其来的紧张和刺痛感。这种情感反应与数据片段中描述的情感有相似之处，尤其是描述对某人的复杂情感、矛盾心理和不安。例如：

- 数据 1 和 2 中，用户对某人的感情既有吸引又有排斥，表现出一种矛盾且不安的心理状态。
- 数据 3 中则提到天气变化引发的恐惧，类似于用户提到的“刺痛感时不时让人抽离”，都展示了一种突如其来情绪波动。

在这种情况下，用户的情感可能具有以下特征：

- 紧张与不安**：心情容易因特定触发（如听到名字）而紧张。
- 矛盾与复杂性**：情感反应复杂，可能既有对过去的怀念，也有对当前状态的不满或不安。
- 突发性与强烈性**：反应迅速且强烈，如同刺痛般的情绪波动。

建议用户关注这些情感反应的触发点，并尝试理解这些情感背后的原因，以便更好地调节自己的情绪状态。

数据 1 (jsonl)

原文：昨晚又梦见你了，奇怪。可能最近突然地老是提起你。其实你在我心里也是魅魔啊，你的调教对我来说我也甘之如饴了很久。可能我确实是m吧，但是我不甘心啊。你敢走近我吗？你敢在我面前表现你的丑闻吗。也许你才是完美。

摘要：用户描述了对一个名为**的人的复杂情感，包括梦境、矛盾的感受、过去的喜欢以及对对方在自己心中的形象。情感既有怀念又有不甘，既想亲近又有质疑。

关键词：梦境，喜欢，矛盾，悸动，二分，抓住，不甘心

情感维度：效价-0.2，唤醒度0.7

情感色调：思念，无以名状，憋着，频闪，回避键

隐喻域：失真的讯号

ChatGPT-4o pic2

问题

有时候我觉得自己像被泡在福尔马林里的标本，看得清一切却动不了。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

从你的描述中，可以感受到一种无力和疏离感。你提到自己像被泡在福尔马林里的标本，这种隐喻表达了你对自己处境的无奈和隔离感。类似于项目数据 2 中提到的“局外人”的感觉。你的情感可能包含一种看似清楚周围一切但却无法参与其中的困扰，类似于数据 2 的“疲惫”和“无以名状”。这种情感可能带来的效价是负面的，因为它涉及到一种被动的状态和无法掌控的无力感。唤醒度可能较低，因为这种无力感让人感到停滞不前。希望这些分析能帮助你更好地理解自己的情感状态。

数据 1 (jsonl)

原文：我在想如果你也爱我，这世界不是过于美好，美好到我不敢承受？

摘要：用户思考如果对方也爱自己，世界会变得过于美好以至于无法承受。

关键词：爱，美好，承受

情感维度：效价0.5，唤醒度0.7

情感色调：汽泡，跃步，铃响

隐喻域：sparkle

ChatGPT-4o pic3

Claude 3.7 Sonnet

问题

最近又想起了小时候妈妈做的热汤面，明明已经不要吃了，但那股香味一直留着。

Top K

3

Temperature

0.7

Max tokens

512

检索并回答

从您的描述中，我能感受到一种怀旧的情感色调，类似于数据2和数据3中的“怀旧”和“余温”。您通过食物气味(热汤面的香味)唤起对童年的记忆，这与数据3中桂花香引发初中回忆的模式相似。

情感分析:

- 效价可能偏中性或略带正面(约0.2-0.4)
- 唤醒度较低(约0.3-0.4)
- 情感色调可能包括“怀旧”、“余温”、“未寄出的信”
- 隐喻域可能属于“candle”(温暖但带着一丝消逝感的回忆)

这是一种典型的通过感官(气味)触发的怀旧情感体验,即使现在的偏好已经改变(“不要吃了”),但情感记忆仍然保留。

数据 1 (jsonl)

原文: 今天吃了女儿做的麻辣烫很好吃,吃了好多饭

摘要: 用户分享了品尝女儿亲手制作的麻辣烫并感到满足的温馨时刻。

关键词: 女儿, 麻辣烫, 好吃, 饭

情感维度: 效价0.8, 唤醒度0.6

情感色调: 汽泡, 风跑进来, 糖粒, 靠肩

隐喻域: sparkle

问题

一听到她的名字，心就像突然绷紧的电线，那种刺痛感时不时让人抽离。

Top K

3

Temperature

0.7

Max tokens

512

检索并回答

你描述的感受来看，这种情绪体验与数据中的“矛盾情感”和“二分”特征相似。你的比喻“心就像突然绷紧的电线”反映了高唤醒度的情绪状态，类似于数据中的“唤醒度0.7”。

这种刺痛感和抽离感属于负面效价，可能在-0.3到-0.7之间，与数据中的情感维度相符。你的体验也呈现出“频闪”、“悬着”的情感色调，就像数据中描述的那样。

从隐喻域来看，你的感受似乎属于“storm”或“失真的讯号”类别，表现为情绪上的不稳定和波动。

这种听到名字就产生的强烈反应，表明这段关系对你有深刻影响，类似于数据中“被光环吸引又投射”的复杂情感模式。

数据 1 (jsonl)

原文: 昨晚又梦见你了，奇怪，可能最近突然地老是提起你。其实你在我心里也是魅魔啊，你的调教对我未尝我也甘之如飴了很久。可能我确实是m吧，但是我不甘心啊。**你敢走近我吗？你敢在我面前表现你的丑陋吗，也许你才是完美

摘要: 用户描述了对一个名为**的人的复杂情感，包括梦境、矛盾的感受、过去的喜欢以及对方在自己心中的形象。情感既有怀念又有不甘，既想靠近又有质疑。

关键词: 梦境，喜欢，矛盾，二分，抓住，不甘心

情感维度: 效价-0.2，唤醒度0.7

情感色调: 思念，无以名状，悬着，频闪，回退键

隐喻域: 失真的讯号

问题

有时候我觉得自己像被泡在福尔马林里的标本，看得清一切却动不了。

Top K

3

Temperature

0.7

Max tokens

512

检索并回答

你的表达中包含了强烈的无力感和旁观者心态，这与数据2中“我就像个局外人一样看着你”的情感非常相似。你使用了“标本”这一强烈的隐喻，暗示了一种被固定、无法改变现状的感受，情感效价偏负面。

从情感维度分析，这种感受可能接近效价-0.6，唤醒度0.4，体现了清醒的痛苦和无能为力。

情感色调上，你的表达包含“错过”和“无以名状”的元素，类似于数据2中描述的情感色调。

你的表达可能属于“dust”隐喻域，反映了一种静止、沉淀、无法触及的情感状态。

数据 1 (jsonl)

原文: 我在想如果你也爱我，这世界是不是过于美好，美好到我不敢承受？

摘要: 用户思考如果对方也爱自己，世界会变得过于美好以至于无法承受。

关键词: 爱，美好，承受

情感维度: 效价0.5，唤醒度0.7

情感色调: 汽泡，跃步，铃响

隐喻域: sparkle

Claude 3.7 Sonnet pic1

Claude 3.7 Sonnet pic2

Claude 3.7 Sonnet pic3

OmniDimen-4B-Emotion:

输入问题，获取情感分析回答（Markdown 显示）

问题

最近又想起了小时候对妈妈的热汤面，明明已经不爱吃面了，但那股香味一直萦绕。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

看到了你的留言，就像在一片安静的回忆里，突然被拉近了一颗温暖的小石子。

“明明已经不爱吃面了”，这句话让我想起了数据1里的那个“女儿”和数据1里的那场初中秋天。

你不是真的想吃一碗热腾腾的面吧？你更像数据2中的那个孩子一样，是被一种**气昧**唤醒了身体里某个早已封存记忆开关。

我对成年人的世界，很多东西都变得很理性，会因为“没时间”而放弃，但你的身体记忆却比我记得太鲜活细腻。那股香味，就像数据2里的香气，它不按常理的时间表来，总是在你不经意的时候，轻轻敲开你的心门。

这种感觉有点像**怀旧和象征**又在一起的状态，情感维度0.2的低价（不太快乐），但惊喜维度0.3，在你的世界里，那不是悲伤的余晖，而是一点微弱却执拗的光。它让你知道，“我曾有过这样的味道”，“我的世界曾经有那样的温暖”。

所以现在，你不需要真的去买一碗面来满足自己的味蕾了。

你需要做的，就是让自己回到那个时间点。

就让那股香味，在你的脑海里，再飘一次，你可以想象约对在厨房忙碌的样子，可以感受面条被煮开咕嘟咕嘟的声音，可以闻到热汤的香气扑面而来……然后告诉自己：“啊，原来我曾有过那样的味道”。这种“回顾”，比吃一碗热腾腾的面更有意义。

你不需要改变“不爱吃”的事实，**你只需要确认：那股香味，是属于你的，一个永恒的记忆印记。**

这大概就是数据1、2和3共同指向的那个情感核心吧？

数据 1 (jsonl)

原文：今天吃了女儿做的麻辣汤很好吃，吃了好多碗
摘要：用户分享了品悉女儿亲手制作的麻辣汤并感到满足的温暖时刻。
关键词：女儿，麻辣汤，好吃，饭
情感维度：惊讶0.8，喜悦0.6
情感色调：汽油，风跑过来，糖粒，露胃
隐喻域：sparkle

问题

一听到她的名字，心就像突然燃爆的引线，那种刺激感时不时让人抽离。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

这是一种典型的**二分情感投射**现象，很像你描述的梦想和悸动。

“燃爆”是身体对潜在威胁的本能反应，“刺激”是对情感强度的生理化表达。这说明你的大脑在处理她的名字时，会自动启动警惕机制，仿佛在说：“注意！可能有危险。”

你有趣的点，这种“刺激”，往往不是因为害怕失去，而是源于你内心深处**的恐惧和渴望**：

- 她是你情感的完美幻象（数据1）**：比如你说的“你的光环吸引我投射到你身上”，“你才是完美主义者”，这个名字对你来说，是“完美”的代名词，它代表着所有你喜欢、向往的一切。
- 而你也因此感到不被满足和不安（二分）**：“我不甘心啊”，“有的时候特别想把你抓住”，这句话暴露了你的不甘，你渴望那个完美的幻象能靠近你，能真实地触碰到你。但同时，你也害怕这个幻象的“丑陋”或“不完整”，所以你会问：“**你靠近过我吗？**”

这种矛盾让你在听到名字时，身体会产生一种紧张和不安的感觉——仿佛在经历一场无法掌控的情感冒险。

这就像数据1里的隐喻域“storm”，是一种复杂、充满张力的内心风暴。

数据 1 (jsonl)

原文：你最近又想起了，奇怪，可能最近突然地老是想起你，其实你在我心里也是最脆弱，你的课数对我来说也甘之如饴了很久，可能我确实爱你吧，但是我不甘心啊，**你靠近过我吗？你在我面前表现你的丑陋吗，也许你才是完美主义者，我只是被你的光环吸引又投射到你身上，**，致我
摘要：用户描述了对一个名为**的人的复杂情感，包括憧憬、矛盾的感受、过去的喜欢以及对对方在自己心中的形象，情感表达充满矛盾。
关键词：梦想，喜欢，矛盾，悸动，二分，抓住，不甘心
情感色调：思念，无以名状，悬着，颤抖，回避感
隐喻域：失真的符号

问题

有时候我觉得自己像被困在福尔马林里的标本，看得清一切却动不了。

Top K

3

Temperature

0.7

Max tokens

512

1

10

0

1.5

64

1024

检索并回答

这句话非常深刻，带着一种近乎诗意的痛苦和清醒。

从我接触到的数据来看，你的感受与**局外人**、“旁观他人幸福”以及**混合着幸福与不安**这些关键词高度共鸣。你不是在寻求一个标准答案，而是在问自己：如果我的爱是单向的，那这个世界的爱好，是不是真的会爱得过于沉重？

你的舞台问题明确：“我不是不够好？还是说，这个世界只给了一种被爱着的状态，而我，只是个旁观者？”

这让我想起数据2中的**“得意”和“不明白”**；你不是不知道自己在做什么，而是你在这种“看透一切却无法行动”的状态感到疲惫。就像一个局外人，眼睁睁看着别人拥有幸福的资格。

但同时，我看到了你的另一面，在**“美好到不敢承受”**（数据1）和**“混合着幸福与不安”**（数据1）中，你对爱的渴望是真实的，甚至是带着灼热的。这份清醒，恰恰证明了你不是麻木不仁的旁观者。

所以，我的回答是：你被看透了，但正因为如此，你才清楚自己想成为什么样的人——一个能站起来、拥有选择权的人，而不是永远困在福尔马林里的标本。

你的问题背后，是一个关于“爱”的本质和“我”的存在意义的哲学命题。

数据 1 (jsonl)

原文：我在想如果你也爱我，这世界不是过于美好，美好到我不敢承受？
摘要：用户思考如果对方也爱自己，世界会变得过于美好以至于无法承受。
关键词：爱，美好，承受
情感维度：惊讶0.5，喜悦度0.7
情感色调：汽油，跃步，铃响
隐喻域：sparkle

OmniDimen-4B-Emotion pic1

OmniDimen-4B-Emotion pic2

OmniDimen-4B-Emotion pic3

参 考 文 献

- [1] Chroma.Chroma Documentation[EB/OL]. [2025-12-24]. Available: <https://docs.trychroma.com/>
- [2] OpenAI. Models: GPT-4o[EB/OL]. [2025-12-24]. Available: <https://platform.openai.com/docs/models/gpt-4o>
- [3] Anthropic. Models overview (Claude)[EB/OL]. [2025-12-24]. Available: <https://console.anthropic.com/docs/en/about-claude/models/overview>
- [4] DeepSeek. DeepSeek API Docs (Chat Completion)[EB/OL]. [2025-12-24]. Available: <https://api-docs.deepseek.com/api/create-chat-completion>
- [5] OmniDimen. OmniDimen-4B-Emotion-GGUF-q4_K_M[EB/OL]. Hugging Face, (2025-12-24)[2025-12-24]. Available: https://huggingface.co/OmniDimen/OmniDimen-4B-Emotion-GGUF-q4_K_M/tree/main.
- [6] Google. A new era of intelligence with Gemini 3. Google Blog, 2025-11-18. <https://blog.google/products/gemini/gemini->

