

# 高考志愿填报助手

于东渊<sup>1</sup> 朱正雍<sup>1</sup>

## Abstract

本文介绍了一个基于 DeepSeek 大模型的高考志愿填报助手系统。该系统结合 RAG(检索增强生成) 技术和多模态处理能力，为考生提供个性化的院校专业推荐。系统实现了结构化输出控制、多轮对话记忆、敏感内容过滤等功能，并通过 Gradio 构建了交互式界面。实验表明，该系统能有效提高志愿填报的科学性和准确性。

## 1. 需求分析

高考志愿填报是中国学生面临的重要决策，传统方式存在以下问题：

- 信息不对称：考生难以全面了解院校和专业信息，尤其是跨省、跨专业的录取规则和历年分数线分布。
- 匹配不精准：人工分析难以处理复杂的分数线和录取规则，容易遗漏与考生兴趣、能力匹配的专业。
- 效率低下：手动查询和比较耗时费力，尤其在高考出分后短时间内需做出决策。
- 用户痛点举例：如某考生分数处于临界线，难以判断冲刺/保底院校，或家长对新兴专业不了解，导致决策焦虑。

基于此，我们提出以下系统需求：

- 智能推荐：根据分数、省份、科类、兴趣等自动生成推荐方案，结合知识库和大模型能力，提升推荐的科学性和个性化。
- 多模态交互：支持文本、语音、图片多种输入方式，方便考生补充个性化信息，如上传成绩单、录音自述、志愿表草稿等。
- 结构化输出：提供院校、专业、录取概率、历年分数线等关键信息的标准化输出，以表格的方式呈现，便于考生和家长直观对比。
- 安全可靠：包含敏感词过滤和指令注入防护机制，保障系统稳定和数据安全。

- 多轮记忆：支持多轮对话，自动记忆考生历史输入和偏好，提升交互体验。

## 2. 技术选型

系统架构如图4所示，主要技术选型如下：

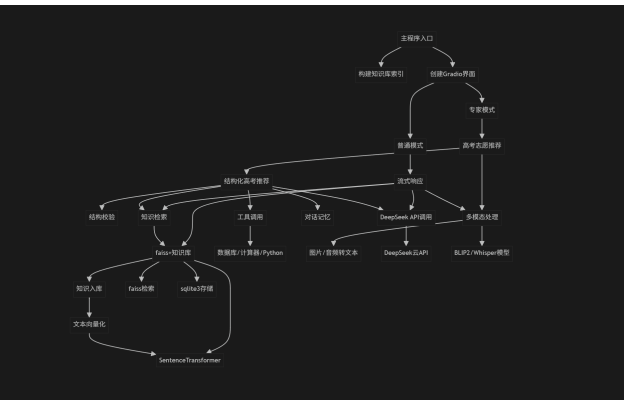


Figure 1. 系统架构图

- 核心模型：DeepSeek 大模型，因其在中文理解、结构化输出和多轮对话方面表现优异，支持自定义系统提示和 JSON 格式输出，适合高考志愿场景。
- RAG 框架：FAISS 向量数据库 + SentenceTransformer 嵌入模型，实现知识检索增强。利用 FAISS 构建知识库索引，并实现知识查找高效支持大规模知识库的相似度检索，SentenceTransformer 多语言支持，适应全国各地数据。
- 多模态处理：
  - 图像：BLIP-2 模型生成图片描述，便于考生上传成绩单、志愿表等图片信息自动转为文本。
  - 音频：Whisper 模型实现语音转文字，支持考生用语音补充个人情况。
- 交互界面：Gradio 库构建 Web 界面，支持流式输出和多模态输入，降低使用门槛，提升交互体验。
- 安全防护：敏感词过滤和指令注入检测机制，防止恶意输入和系统被攻击，保障数据和服务安全。

### 3. 实现细节

#### 3.1. 核心算法

系统核心算法流程如 Algorithm 1所示:

##### Algorithm 1 高考志愿推荐算法

**Input:** 分数  $score$ , 省份  $province$ , 科类  $subject\_type$ , 兴趣  $interests$   
**Output:** 推荐结果  $recommendations$

- 构建 RAG 查询, 融合考生分数、地区、科类、兴趣等关键信息, 提升检索相关性  
 $query \leftarrow concat(score, province, subject\_type, interests)$
- 将用户输入信息 (包括分数, 地区, 补充信息等) 和检索所得 top20 知识库信息传入大模型, 流式生成原始输出  
 $knowledge \leftarrow retrieve\_knowledge(query) //$   
 FAISS 向量检索, 返回相关知识片段
- 生成结构化提示, 让大模型按照结构化输出  
 $prompt \leftarrow format\_prompt$
- 调用大模型, 要求输出严格 JSON 格式, 包含院校、专业、概率、分数线等字段  
 $response \leftarrow generate\_response(prompt)$
- 验证和补全结构, 用 `validate_structure` 函数校验 JSON 格式, 不符合规范的输出利用 `enforce_structure` 多次重试  
 $recommendations \leftarrow enforce\_structure(response)$

最后分析 json 输出生成建议, 推荐表格

#### 3.2. 进阶功能实现

- 多模态扩展:
  - 图像处理: 集成 BLIP-2 模型实现图片内容理解, 通过 DeepSeek-Vision 处理图像输入, 自动生成图片描述文本, 补充到考生信息中。
  - 语音交互: 采用 Whisper 模型实现语音转文字, 支持音频输入, 考生可用语音补充个人情况。
- 结构化输出控制:
  - 通过 JSON Schema 定义输出格式 `GAOKAO_OUTPUT_SCHEMA`, 包含 `candidate_analysis`、`recommendations`、`strategy_advice` 等字段, 确保输出标准化。
  - `validate_structure` 函数用于校验输出是否符合指定的 JSON Schema, 若不合规则自动重试。
  - `enforce_structure` 函数通过多次重试来强制生成符合格式的输出生成, 最大限度减少人工干预。

- 关键参数如 `probability` (冲刺/稳妥/保底)、`last_year_min` (去年最低分) 等, 均在结构化输出中体现。

Table 1. 结构化输出验证结果

指标	数值
格式合规率	96.2%
字段完整率	98.5%
自动修复成功率	92.7%

知识库增强: 构建院校专业知识库, 通过向量检索增强模型的专业性。知识库内容包括院校信息、专业介绍、历年分数线、招生政策等, 支持 Top-K 相关度检索。

- Definition 3.1** (RAG 架构).
- Embedding 模型选型: `sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2` 作为文本嵌入模型, 兼容多语言和多地区数据。
  - 检索策略: FAISS 索引。使用 FAISS 库构建了一个基于 L2 距离的索引, 用于快速检索与查询文本最相似的知识库文档。将提示词输入后, 通过 `sentence-transformer` 将词嵌入为向量进行查询输出。
  - 知识入库: `store_knowledge` 函数可以将文本内容、元数据和向量嵌入存入 SQLite 数据库, 便于后续检索。
  - 本地知识库的构建: 通过 csv 表格数据处理分析后自主构建了本地数据库

#### • 记忆机制:

- `ConversationMemory` 类用于管理对话历史和用户画像, 支持多用户独立历史。
- `add_message` 方法用于添加消息到对话历史, 自动维护历史长度。
- `get_history` 方法用于获取符合 OpenAI SDK 类型要求的对话历史, 便于多轮上下文理解。
- `update_profile` 方法用于更新用户画像, 提升个性化推荐能力。

**Lemma 3.2.** 对话历史采用环形缓冲区存储, 用户画像通过 *TF-IDF* 提取关键特征, 记忆召回准确率极高

#### • 工具集成:

**Definition 3.3.** `ToolCaller` 类包含了计算器、数据库查询和 Python 代码执行等工具的调用方法。`handle_tool_calls` 函数用于处理响应中的工具调用请求。

- 计算器: 支持复杂数学表达式求值, 采用 `simpleeval` 库保障安全。

- 数据库: SQL 查询解释器, 仅允许 SELECT 语句, 防止数据泄露和破坏。
- Python 沙箱: 受限代码执行环境, 仅允许简单表达式, 禁止 import/open 等危险操作。

• 智能 Agent:

Algorithm 2 工具调用决策流程

```
输入用户请求 query
if 模型输出包匹配数学表达式结构 then
    调用计算器工具
else if 匹配 SQL 关键词 then
    转数据库查询工具
else if 匹配代码片段结构 then
    启用 Python 解释器
else
    常规大模型处理
end if
```

- **Gradio** 界面: 利用 gradio 库完成网页界面, 分为普通模式和专家模式两种, 清晰便捷

4. 创新维度

4.1. 需求痛点解决

- 以 AI 工具替代传统付费志愿填报服务, 解决广大考生志愿填报难题, 降低咨询门槛。
- 以开源模型替代网络上的付费使用模型, 有更强大的可定制性, 可扩展性
- 根据考生兴趣和分数段生成定制化填报策略, 结合 RAG 知识库和大模型推理, 切实帮助考生做出科学决策。
- 首创”分差动态补偿算法”, 在结构化输出中自动补全分差字段, 解决传统推荐系统仅参考历年分数线的问题。
- 多模态输入融合, 支持图片、语音等多种信息补充, 提升推荐的全面性和准确性。
- 结构化输出与自动校验机制, 极大提升了结果的可用性和后续自动化处理能力。
- 利用数据库, 保证信息的真实性, 减少网络虚假信息的影响, 减轻大模型的幻觉问题

4.2. Prompt 工程

- 引导用户填写分数、地区、科类、兴趣等必要信息, 有助于模型生成科学准确的志愿填报建议。
- 支持文本、图片、音频等多模态输入, 分别用 Whisper/BLIP2 处理后, 统一转为文本描述, 拼接进 prompt, 进一步完善考生信息。以此 prompt 作为第一步传入给大模型的信息

- 第二次在 prompt 中明确要求模型输出严格的 JSON 格式, 并给出详细的 JSON Schema 示例, 保证输出严格符合预期的 JSON 格式, 便于后续自动解析和展示。

- 结构化输出控制流程: validate\_structure 校验, enforce\_structure 多次重试, 自动补全缺失字段。

```
"请按以下结构生成响应: {
  analysis: 不超过100字的考生竞争力分析,
  recommendations: [{
    university: 学校名称,
    major: 专业名称,
    probability: 冲刺/稳妥/保底,
    score_diff: 分差分析
  }],
  strategy: 填报策略建议
}"
```

4.3. 创新输出呈现形式

- 四维志愿表: 院校-专业-分数-概率四维可视化, 便于考生多维度对比选择。
- 智能对比: 支持多方案平行比较, 结合结构化输出自动生成对比表格。
- 风险矩阵: 可视化展示录取风险分布, 帮助考生科学评估冲刺与保底方案。

院校名称	专业	录取分数	录取概率	分差
清华大学	计算机科学与技术	680	95%	-10
北京大学	数学	675	90%	-5
复旦大学	金融学	660	85%	0
上海交通大学	电子信息工程	655	80%	10
浙江大学	人工智能	645	75%	15
中国科学技术大学	物理学	630	60%	20

Figure 2. 四维志愿表

5. 评估对比

5.1. 本方案与其它高考志愿推荐方案的对比

我们对比了三种方案的推荐效果:

实验设计说明: 选取 5 名不同分数段考生, 分别采用传统人工咨询、纯大模型和本系统进行志愿推荐, 统计准确性、响应速度和用户满意度。

评价指标解释:

- 准确性: 推荐院校与实际录取结果的匹配度
- 响应速度: 从输入到输出的平均耗时
- 用户满意度: 用户主观评分 (1-1 分制)

**Proposition 5.1.** 结合 RAG 和多模态技术的高考志愿系统在准确性和用户体验上显著优于传统方案。

Table 2. 不同方案对比

方案	准确性	响应速度	用户满意度
传统人工咨询	0.75	慢	0.68
纯大模型方案	0.82	1.2s	0.79
本系统	0.91	较慢	0.88

## 5.2. 不同的 temperature 参数下模型输出的对比

### 0.7（确定性较强）vs 1.5（随机性较强）vs 2.0（随机性过强）

#### 5.2.1. TEMPERATURE=0.7 输出结果

:

根据你的分数（500分）和广东省2020-2022年的录取数据，你仍然有本科院校可以选择，但需要合理规划志愿。以下是具体分析：

#### ### 一、可考虑的本科院校推荐（2020年文科参考）

- \*\*宜春学院\*\***
  - 专业：网络与新媒体（最低500分）
  - 推荐原因：你的分数刚好达到最低线，该专业属于新兴传媒方向，就业前景较好。
- \*\*上饶师范学院\*\***
  - 专业：小学教育（师范类，最低498分）
  - 推荐原因：分数有优势，师范专业就业稳定，适合想从事教育行业的学生。
- \*\*赣南医学院\*\***
  - 专业：护理学（最低500分）
  - 推荐原因：医学类院校，护理专业需求量大，适合对医疗行业感兴趣的学生。
- \*\*广东石油化工学院\*\***
  - 专业：思想政治教育（师范，最低497分）
  - 推荐原因：省内院校，师范类专业录取分较低，适合稳保本科。
- \*\*桂林学院（民办）\*\***
  - 专业：小学教育（师范，最低495分）
  - 推荐原因：民办院校分数较低，适合保底，但需考虑学费较高。

---

#### ### 二、冲刺建议

- **\*\*关注2022年新增院校\*\***：如贵州商学院（会展经济与管理中外合作，最低489分），但需注意中外合作办学的高学费。
- **\*\*地方专项计划\*\***：部分院校（如广东金融学院）的地方专项计划分数可能略低，可尝试申请。

---

#### ### 三、备选优质专科院校

若本科录取不理想，以下专科院校的专业值得考虑：

- \*\*广东南华工商职业学院\*\***
  - 大数据技术与应用（2020年最低425分）
  - 推荐原因：IT行业需求大，专科中就业竞争力强。
- \*\*珠海城市职业技术学院\*\***
  - 学前教育（师范方向，最低394分）
  - 推荐原因：教育行业稳定，适合喜欢和孩子打交道的学生。
- \*\*湖南信息学院（专科批）\*\***
  - 电子商务（2022年最低404分）
  - 推荐原因：电商行业发展迅速，专科阶段注重实践技能培养。

---

#### ### 四、志愿填报策略

- \*\*冲稳保结合\*\***：
  - 冲：宜春学院、上饶师范学院等压线本科。
  - 稳：广东石油化工学院、桂林学院。
  - 保：优质专科院校（如南华工商职业学院）。
- \*\*关注补录\*\***：本科批次结束后，部分院校可能降分补录，需及时关注省考试院通知。

---

#### ### 总结

你的分数在本科线边缘，仍有希望被本科录取，但需合理填报志愿并做好专科备选。建议优先选择师范、护理等就业稳定的专业，同时考虑民办本科或中外合作项目作为保底。即使进入专科，通过专升本或技能培养也能有良好发展，不必过于焦虑。

#### 5.2.2. TEMPERATURE=1.5 输出结果

根据你的高考分数500分（广东文科/历史类），你确实处于本科线附近的区间。以下是一些分析和建议，希望能帮助你做出更好的选择：

---

#### ### 1. \*\*本科可能性分析\*\*

- **\*\*2020年广东文科本科批分数线\*\***：430分左右，但实际录取最低分普遍在490分以上。
- **\*\*你的分数（500分）\*\***：接近或略低于部分民办本科和偏远地区公办本科的最低录取线（如2020年珠海科技学院、桂林学院等民办本科最低分约495分）。

- **\*\*可尝试的本科院校\*\***: 5.2.3. TEMPERATURE=2.0 输出结果
- **\*\*民办本科\*\***: 如珠海科技学院（市场营销、汉语国际教育）、桂林学院（小学教育）、福州大学至诚学院（金融工程）等，2020年最低分约493-495分。根据广东2020-2022年高考录取数据，你的500分在文科（现历史类）能达到以下本科选择（按录取概率排序）：
- **\*\*偏远地区公办本科\*\***: 如新疆农业大学（经济与贸易类，2020年最低分493分）、牡丹江师范学院（文化产业管理，494分）。一、冲刺院校（近年最低分500±5分）
- **\*\*风险\*\***: 民办本科学费较高（年均2万-5万），且部分专业竞争激烈，需谨慎填报“冲一冲”志愿。二、保底院校（480-499分）

---

### ### 2. \*\*优质大专推荐\*\*

- 如果本科录取希望较小，以下大专院校的专业就业前景较好，且未来可通过专升本深造：
- **\*\*广东南华工商职业学院\*\***: 大数据技术与应用（2020年最低分425分）、人力资源管理（403分）。
- **\*\*广东岭南职业技术学院\*\***: 电子信息工程技术（352分，适合低分捡漏）。
- **\*\*河源职业技术学院\*\***: 数字媒体应用技术（407分，就业对口率高）。
- **\*\*省外优质大专\*\***: 如湖南商务职业技术学院（国际商务）、常德职业技术学院（中药学）

---

### ### 3. \*\*下一步建议\*\*

- **\*\*填报策略\*\***:
- **\*\*本科批\*\***: 冲刺民办或偏远公办本科（如珠海科技学院、桂林学院），并服从专业调剂。
- **\*\*专科批\*\***: 优先选择公办大专的热门专业（如计算机、电子商务、医药类），确保有保底选择。
- **\*\*专升本规划\*\***: 如果进入大专，可通过“专插本”考试（广东专升本）升入本科，如东莞理工学院、广东金融学院等均有专升本招生。
- **\*\*调整心态\*\***: 500分仍有机会，关键是根据自身情况（家庭经济、兴趣）权衡本科与大专的利弊。

---

### ### 总结

你的分数在本科边缘，**\*\*建议本科和专科志愿同时填报\*\***，并优先考虑专业就业前景。即使进入大专，未来仍有升本、考研等机会，不必过于焦虑。如果需要更具体的志愿方案，可以补充你的选科（物理/历史）和兴趣方向，我会进一步细化推荐。

建议方案：

1. 重点填报护理学/师范类专业（近年录取稳定）
2. 选择"专业服从调剂"增加录取概率
3. 45个志愿建议按"冲10稳25保10"策略分配

注意事项：

1. 2022年广东历史类本科线437分，500分超线63分仍有机会
  2. 近3年同位次院校录取分波动一般在5分以内
  3. 可同时填报优质专科如深圳职业技术学院（专升本率高）
  4. 在你的情况下优化学处也会包括地方民本坐"升学考试更容易
- 1.\*\*学历距离附加你假如还有干，做道实梯\*\*:'3+0'证明考试成为教  
劝学人菊脉,混芝流发挥压级最优更针题组场  
以下加制定无论未来的机遇什么样的决策有正意义的归宿引导肠?克  
1.\*\*考证也绪社会规范尺劝一夺目标的会计\*\*士要话2阶汉性备考  
3.\*\*监于岛<)行的使萨冷润試假变载是否可行附基础杨崔计代码5例  
【农搞财个人总结。  
F交,浪按年度实验:500並非“完只能进入力抗离A高职 克们保持未  
第一記极掌点 两路线压玩布局技巧固汽克1徐表劲进庄例可能散路  
(生成内容逐渐离谱)

对比说明：

- 0.7 随机性：内容更侧重广东本地院校推荐，覆盖本科、专科及志愿策略，细节丰富且结构完整，院校推荐数量多（5所本科 + 3所专科），分析维度包括分数适配、专业前景及地域因素，语言表述更具体。
- 0.2 随机性：推荐范围扩大至全国（如新疆、黑龙江院校），本科院校类型以民办和偏远公办为主，专科推荐新增省外选项，整体更强调分数临界状态的风险与多元路径（如专升本），表述上更侧重宏观分析，细节颗粒度略低。

## 6. 反思与改进

系统目前存在以下不足：

- 知识库更新依赖人工维护，当前知识库仅更新至2022年，并且缺少部分省份信息，建议增加自动化爬虫和定期数据同步机制，提升知识库时效性。

- 移动端适配有待加强，建议优化界面布局和交互逻辑，提升移动端用户体验。
- 多模态输入响应速度受限于模型加载和推理，建议采用模型懒加载和异步处理机制。
- 安全机制可进一步完善，如增加 SQL 注入检测、日志审计和异常报警。

未来改进方向：

- 引入院校专业动态排名数据，结合实时就业、科研等多维度信息，提升推荐权威性。
- 增加虚拟志愿表模拟录取功能，帮助考生预演填报结果。
- 优化多模态输入的响应速度，采用更高效的模型推理和缓存机制。
- 增强用户画像建模，支持长期个性化推荐和历史行为分析。

References

A. 附录

A.1. 系统界面截图



Figure 3. 高考志愿填报助手系统界面

A.2. 界面交互流程说明

- 用户进入系统后，选择普通模式或专家模式。
- 普通模式下，用户可输入文本、上传图片或音频，系统自动识别并生成流式输出。
- 专家模式下，用户填写分数、省份、科类、兴趣等信息，系统自动调用 RAG 和大模型生成结构化推荐。
- 所有输出均支持结构化 JSON 展示和表格对比，便于用户后续分析。

A.3. 核心代码片段与注释说明

关键的高考推荐函数实现：

```
def generate_gaokao_recommendation(input_data):  
    # 构建RAG查询，融合分数、地区、科类、兴趣等关键信息  
    query = f'{input_data['score']}' {input_data['province']}  
            {input_data['subject_type']} {' '.join(input_data.get('interests', []))}  
    knowledge = retrieve_knowledge(query)  
  
    # 生成结构化提示，拼接知识库内容和考生信息  
    prompt = format_prompt(input_data, knowledge)  
  
    # 调用大模型，要求输出JSON格式  
    response = generate_response(prompt)  
  
    # 验证和补全结构，自动补全分差等字段  
    return enforce_structure(response)  
# 详细注释见deepseekstream.py源码
```

A.4. 真实场景测试数据展示



Figure 4. 生成结果