### **人工智能基础期末大作业细则**

**主题：（多模态）大语言模型应用开发**  
**目标**：开发一个解决实际需求的大语言模型应用，需体现对LLM或者VLM核心能力的理解与扩展

#### **一、项目要求**

1. **基础功能层**（必须完成）
   * 集成至少1个大语言模型API（如GPT-3.5/4、qwen、deepseek等），展示不同参数（如temperature=0.7 vs 1.2）对输出的影响对比
   * 实现核心功能闭环，需定义明确的输入输出规范（如JSON Schema）并实现校验机制
   * 包含基础交互界面（命令行/CUI/GUI任选）
   * 处理一些真实场景测试数据作为展示
   * 包含下方“技术实现规范”中的要点。
2. **进阶功能层**（选择性完成）
   * 实现多模态扩展（图像/语音输入输出）
   * 实现高质量结构化输出控制，能95%的概率生成符合要求的格式
   * 构建领域知识库增强，如RAG架构需说明embedding模型选型（如text2vec）和检索策略（如FAISS索引）
   * 开发记忆机制（对话历史/用户画像）
   * 集成外部工具（包括但不限于计算器、数据库、API、Python代码解释器）
   * 实现一个自主“Agent”，需实现至少3种工具调用决策逻辑
3. **创新维度**（选择性完成）
   * 解决未被主流产品覆盖的需求痛点
   * 提出新颖的prompt engineering方案
   * 设计独特的输出呈现形式

* 注：以上括号内的，均是举例，而非必须和仅仅。比如集成的外部工具可以从括号中选几个，也可以自己再额外设计。

#### **二、技术实现规范**

1. **模型调用**
   * 必须展示API调用参数调优过程（temperature/top\_p等）
   * 需处理流式响应，实现逐字/分块输出效果（如ChatGPT式打字机效果），禁用单次完整响应
   * 实现输入预处理（敏感词过滤/指令注入防护）
2. **系统架构**
   * 需提供架构设计图（数据流图/模块关系图）
   * 要求模块化开发（至少3个独立功能模块）
   * 必须包含异常处理机制（API失败重试、结构化输出失败等）

#### **三、评分标准矩阵**

| 评估维度 | 权重 | 评估要点 |
| --- | --- | --- |
| **技术深度** | 20% | API高级参数应用、多模型协同、系统扩展性 |
| **功能完整度** | 30% | 核心功能实现度、异常处理覆盖率、单元测试通过率 |
| **创新价值** | 20% | 需求新颖性（需附竞品分析）、解决方案独创性、技术组合创新性 |
| **工程规范** | 10% | git管理、API密钥安全管理、依赖管理（requirements.txt） |
| **展示质量** | 20% | 演示视频流畅度（3-5分钟）、技术文档完整性（API文档+用户手册）、答辩表现 |

**加分项**（上限10分）：

* 获得真实用户反馈（需提供证明）
* 开源项目并获得star

**扣分项**：

* 大规模代码抄袭：零分处理

#### **四、交付清单**

1. 源代码（以git仓库的形式提交）
2. 技术报告（10-15页，需包含：需求分析→技术选型→实现细节→评估对比→反思）
3. 演示视频（MP4格式，含功能演示与技术解说）
4. 可执行文件或部署方案README或部署链接（需保证期末周可访问）
5. 多人组队时，需要提供小组分工表

#### **五、时间节点**

* 选题申报：DDL+7天（需提交1页提案）
* 中期检查：DDL-14天（代码完成度≥40%）
* 最终提交：YYYY-MM-DD 23:59

#### **六、补充说明**

* 推荐使用dify搭建应用。
* 鼓励使用LLM生成代码；如遇查重，可提供原始代码生成网址、截图等信息免于查重。