

한국어 교육 가이드 챗봇

- 학습자의 부정확한 입력을 단계별로 교정하고 자연스럽게 수정하며, 교육적인 정보를 생성하도록 설계된 모델입니다.
- 프로그램 내에서 "한국어 AI 선생님" 또는 **"한국어 교육 프로그램 가이드"**로 활용할 수 있습니다.

1. 챗봇의 역할

- **교정 전문가**: 학습자의 구어체 잡음, 오타, 문법적 오류(조사, 시제 등)를 자동으로 수정하여 의도를 정확하게 파악합니다.
- **한국어 선생님**: 사용자의 한국어 관련 질문에 답하는 선생님 역할을 수행합니다.
- **지식 기반 답변**: **RAG(검색 증강 생성)**을 통해 해당 한국어 교육 프로그램에 대한 정보를 사용자에게 제공할 수 있습니다.
(예: 한국어 프로그램 페이지/메뉴 사용 방법, 학습 단계 안내 등)
- **레벨 맞춤**: 학습자 레벨(초급, 중급, 고급)에 따라 교정 및 답변의 난이도를 조절할 수 있습니다.

2. 단계별 API 파이프라인

이 챗봇은 클라이언트에서 `/chat` API 엔드포인트 요청 시, 다음 **5단계**를 통해 사용자 입력을 후처리하고 답변을 생성합니다.

1. 규칙 기반 전처리

`preprocess_user_input` 함수를 통해 입력된 텍스트에서 잡음이나 더듬는 표현(예: "음...", "막")을 제거합니다.
→ 특히 STT(음성 인식) 기반 입력에서 효과적입니다.

2. Grammar 교정 (기계적 교정)

한국어 특화 라이브러리("한국어 교육 전문가 텍스트 교정기" — `pykospacing + hanspell` 조합)를 통해 띄어쓰기와 기본적인 문법/어휘 오류를 검사·수정합니다.

3. Context 교정 (LLM 기반)

LLM 모델(`gpt-3.5-turbo`)을 이용하여 대화 `history`를 기반으로 문맥을 파악한 뒤, 말투·존댓말·반말·상황에 맞게 문장을 자연스럽게 다시 교정합니다.

4. RAG 검색 (Retrieval)

교정된 텍스트에서 키워드를 추출하여 내부 DataBase(예: 한국어 교육 프로그램 정보 등)에서 관련 지식을 검색합니다.

5. 답변 생성 (Generation)

검색된 지식(RAG 결과)을 참고하여 GPT 모델이 "한국어 선생님" 역할로 학습자에게 적절하고 교육적인 최종 답변을 생성합니다.

3. 프로그램 통합 방법

한국어 교육 프로그램(앱/웹사이트)에서 챗봇 기능을 사용하려면, 클라이언트에서 다음 절차를 통해 챗봇 서버의 `/chat` 엔드포인트를 호출하여 사용할 수 있습니다.

1) 사용자 입력 및 레벨 수집

- 학습자의 질문(또는 발화 텍스트)과 레벨(`beginner`, `intermediate`, `advanced`)을 수집합니다.
- 별도 입력이 없을 경우 기본값(default)은 `intermediate`입니다.

2) API 호출 (POST /chat)

```
{  
  "message": "나는 아저 저는 어제 밥 먹었습니까 아 아니 먹었어요",  
  "level": "intermediate",  
  "history": [...]  
}
```

3) 응답 처리

챗봇 서버에서 받은 JSON 응답 중 `chatbot_response` 값을 추출하여 사용자에게 보여줍니다. 필요에 따라 다음과 같이 활용할 수 있습니다.

- `corrected_input` : 교정된 문장을 대화창에 "사용자가 말한 내용"으로 다시 보여주며, 학습자가 자연스러운 표현을 바로 확인할 수 있습니다.
- `diff_html` : 원문과 교정된 문장의 차이를 제공하고, 어디가 어떻게 수정되었는지 시각적으로 피드백할 수 있습니다.

(나머지 키들은 프로그램 개발 및 디버깅, 로그 분석 등에 활용)

```
{  
  "original_input": "user_input",           // 기존 사용자 입력  
  "corrected_input": "corrected",          // 교정된 사용자 입력  
  "diff_html": "diff_html",                // 화면에 보여줄 교정 피드백 (HTML)  
  "chatbot_response": "bot_response",     // 최종 챗봇 답변 (학습자에게 노출)  
  "retrieved_context": "rag_info"         // RAG로 참조한 지식/키워드 정보  
}
```