



# 외부 서비스 정리 문서

## 1. LiteLLM

GPT / Claude / Gemini 등 외부 AI 모델을 LiteLLM을 통해 호출합니다.  
사용자 개인 API Key를 입력받아 요청을 위임하는 구조입니다.

LiteLLM은 내부 서비스이지만, LiteLLM이 호출하는 GPT/Claude/Gemini API는  
모두 외부 모델 제공사 API이므로 외부 서비스로 분류합니다.

### LiteLLM 기본 설정

```
LITELLM_BASE_URL=http://litellm:4000  
LITELLM_MASTER_KEY=sk-1234
```

### LiteLLM 사용 목적

- 사용자 API Key 기반으로 LLM 요청을 통합 처리
- OpenAI, Anthropic, Google Gemini 모델을 단일 인터페이스에서 호출
- 모델 라우팅 / 모델별 파라미터 통일

### 지원 LLM 모델 목록

#### OpenAI

- GPT-4o
- GPT-4o-Mini
- GPT-5
- GPT-5-Mini
- GPT-5-Nano

#### Anthropic (Claude)

- Claude-Sonnet-4
- Claude-Opus-4-1
- Claude-3-7-Sonnet
- Claude-3-5-Haiku

#### Google Gemini

- Gemini-2.5-Pro
- Gemini-2.5-Flash

- Gemini-2.5-Flash-Lite

## 사용자 Key 입력 구조

모델 제공사	필요한 사용자 키
OpenAI	OPENAI_API_KEY
Google Gemini	GEMINI_API_KEY
Anthropic	ANTHROPIC_API_KEY

| 모든 사용자 요청은 **사용자 자체 키** 기반으로 처리됩니다.

## 2. GMS API (SSAFY 공식 외부 서비스)

| SSAFY에서 제공하는 외부 API 서비스

모델 실행 및 내부 기능 보조용으로 활용됩니다.

### 기본 설정

GMS\_BASE\_URL=https://gms.ssafy.io/gmsapi/

### 사용 목적

- SSAFY GMS 모델 연동
- 자체 서비스가 아닌 외부 API에 의존하는 구조