

## SK네트웍스 Family AI과정 6기

# 모델배포 시스템 구성도

### □ 개요

- 산출물 단계 : 모델배포
- 평가 산출물 : 시스템 구성도
- 제출 일자 : 2025.03.14
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN06-FINAL-2Team>
- 작성 팀원 : 정유진

개요	<p>본 시스템은 Django 웹 애플리케이션으로, LLM 기반 웹툰, 웹소설 추천 챗봇을 배포하는 것을 목표로 합니다.</p> <p>주요 기능:</p> <ul style="list-style-type: none"><li>• LLM 기반 추천 시스템: 사용자의 취향 및 추천 기록을 반영한 맞춤형 웹소설·웹툰 추천</li><li>• 벡터 검색 기술: ChromaDB를 활용하여 최적의 작품 필터링</li><li>• AWS 클라우드 인프라 배포: EC2, RDS, Nginx, Unicorn을 이용한 안정적인 서비스 운영</li></ul> <p>본 문서에서는 시스템의 구성 요소, 데이터 흐름에 대한 내용을 다룹니다.</p>
구성 요소	<ol style="list-style-type: none"><li>1. 프론트엔드<ul style="list-style-type: none"><li>◦ HTML &amp; CSS</li><li>◦ JavaScript</li></ul></li><li>2. 백엔드<ul style="list-style-type: none"><li>◦ Django Framework</li><li>◦ LLM 기반 챗봇 엔진<ul style="list-style-type: none"><li>■ LangChain + OpenAI GPT-4o 모델 챗봇</li><li>■ 벡터 검색</li><li>■ MySQL 활용</li></ul></li></ul></li><li>3. 데이터베이스<ul style="list-style-type: none"><li>◦ MySQL (AWS RDS)</li></ul></li><li>4. 배포 환경<ul style="list-style-type: none"><li>◦ 웹 서버: Unicorn + Nginx</li><li>◦ 클라우드 인프라: AWS EC2 인스턴스</li></ul></li></ol>
데이터 흐름	<ol style="list-style-type: none"><li>1. 사용자가 질문 입력<ul style="list-style-type: none"><li>• 프론트엔드에서 입력값을 받아 서버에 전달</li><li>• views.py에서 해당 모델 챗봇 호출</li></ul></li><li>2. LLM 기반 응답 생성<ul style="list-style-type: none"><li>• Django 서버에서 LangChain을 통해 OpenAI GPT-4o 호출</li><li>• 사용자 취향(Preset)과 추천 내역(RecommendedWork)을 반영해 최적의 웹소설/웹툰 추천</li></ul></li><li>3. 응답 전송 및 UI 출력<ul style="list-style-type: none"><li>• 생성된 응답을 프론트엔드에서 받아 UI에 출력</li><li>• 사용자 메시지와 AI 응답을 채팅창(chat-box)에 추가</li></ul></li></ol>

