

**SK네트웍스 Family AI 과정 15기  
 모델배포 시스템 구성도**



| **산출물 단계** | 모델배포 |
| --- | --- |
| **평가 산출물** | 시스템 구성도 |
| **제출 일자** | 2025.10.02 |
| **깃허브 경로** | <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN15-FINAL-4TEAM> |
| **작성 팀원** | 권주연,임경원,정민철 |

| **개요** |  |
| --- | --- |

| **구성 요소** | * **Backend -FastAPI on AWS**   **1. EC2 (또는 ECS/Fargate) :**  FastAPI 애플리케이션 실행 환경.  **2.** **Nginx + Gunicorn :**  리버스 프록시 + WSGI 서버  **3.** **FastAPI 요청 처리 및 오케스트레이션:**  4.  **LangGraph 워크플로우 관리:**  sLLM 호출, 프롬프트 최적화, 이미지 모델 호출  **5.** **PostgreSQL (Amazon RDS) :** 사용자 계정, 요청 로그, 가구 메타데이터 저장  **6.** **S3 (Amazon Simple Storage Service) :** 학습 시 필요한 사진, 생성된 인테리어 이미지 저장, URL 반환   * **Backend**   **1.**  **sLLM :**  인테리어 추천 방안 텍스트 생성  **2.** **Image Model (Diffusion/FLUX) :**  RunPod에서 프롬프트 기반 인테리어 이미지 생성  **3.**  **OpenAI API (ChatGPT) :** 가구 키워드 추출 및 보조 질의 처리  **4.**  **Naver Shopping API:** 키워드 기반 가구 검색   * **Frontend (JavaScript Web Client)**   **1. 브라우저 기반 웹 UI (JavaScript/HTML/CSS):**  사용자 입력(스타일, 옵션) → 요청 전송  **2. Cognito(관리자 생성 계정 기반 로그인)**  사용자 인증 및 권한 관리     * **DNS 및 HTTPS 통합**   **1.**  **Route 53 도메인 / DNS 관리**  **2.** **AWS Certificate Manager**  SSL 인증서 관리 (HTTPS 지원) |
| --- | --- |
| **데이터 흐름** | * **개발자 작업**   1. GitHub에 코드 푸시  2. Backend CI/CD Pipeline → Docker 이미지 빌드 → EC2에 배포  3. Frontend CI/CD Pipeline으로 정적 파일을 AWS Amplify에 배포.     * **백엔드**   1. FastAPI 기반 백엔드 애플리케이션이 EC2 인스턴스에서 실행  (Nginx + Gunicorn)  2. ALB를 통해 HTTP/HTTPS 요청을 라우팅.  3. LangGraph 오케스트레이션:   * sLLM 호출 (인테리어 추천 텍스트 생성) * Prompt Engineering 계층 거쳐 Image Model 실행   4. RunPod: sLLM + Image Model 실행 환경  5. OpenAI API(ChatGPT): 가구 키워드 추출  6. PostgreSQL (RDS): 사용자 계정, 요청 로그, 가구 메타데이터 저장  7. Amazon S3: 학습용 데이터와 생성 이미지 저장, URL 반환   * **프론트엔드**   1. JavaScript 웹 클라이언트 (브라우저 기반 UI)  2. Cognito (관리자 생성 계정 기반 로그인) → 사용자 인증 및 권한 부여     * **사용자 요청**   1. Route 53: DNS 관리, 사용자 요청을 프론트엔드(Amplify/S3) 및 백엔드(ALB)로 전달  2. AWS Certificate Manager: HTTPS 통신 지원 |