

LLM 마스터클래스 #7

LLM 서비스 운영: 실전 서비스 구현



Part 7-1.

앱 만들기, 배포, 운영 소개



실제로 만들어서 배포한다면

노 코드 - MS Copilot 만들기

가장 간단한 옵션 - local notebook, Ollama, Google Colab

Flask + # Google Cloud Run ◀ 우리의 샘플!

Full application



Part 7-2.

No-code 옵션



MS Copilot 만들기

- 데모



Part 7-3. **PromptFlow /** **Platform option**



PromptFlow 란?

- PromptFlow 소개
- 데모



Part 7-4.

가장 간단한 파이썬 프로그램 - Flask & Google Cloud Run



Flask app

- 가장 간단한 Flask app - on localhost
- Demo
- Run on Google Cloud Run



Part 7-5. **Full application**



Full application 이란?

- 실제 프로젝트 예
- 당신이 맡을 일



Part 7-6.
가장 현실적인 시나리오



당신이 맡은 역할과 당신의 목적

노트북으로 OK

- 아주 빨리 POC 테스트
- 새로 나온 LLM 모델 테스트

No/low 코드 옵션

- 이미 MS 제품을 쓰고 있음
- 이미 LLM Platform 이 있음
- 데이터 베이스와 auth 통합이 복잡하고, 이미 구현되어 있음

실제 서비스를 한 번 만들어 보고 싶음

- Flask + Google Cloud Run
- Amazon

실무자인

- 회사 시스템에 integration

