

# 모델 배포

## - 배포 개요

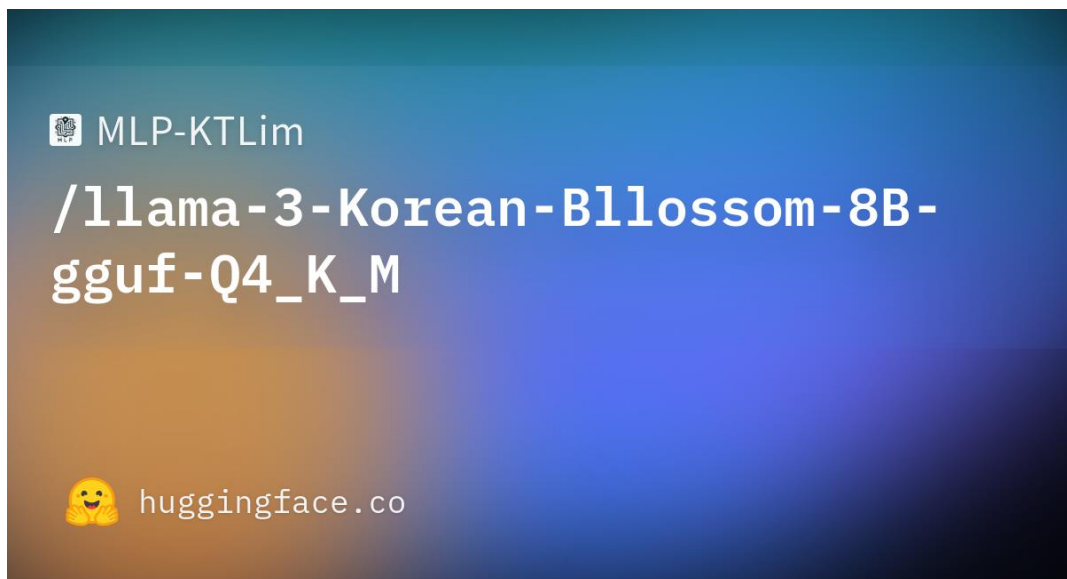
- 총 3개의 로컬 모델과 1개의 API 모델을 활용하여 서비스 배포를 수행하였음.

## - 배포 모델

- Llama-3-Korean-Blossom-8B-gguf-Q4\_K\_M
- Gemma-2-2b-it-4bit
- Lightning-whisper-mlx
- OpenAI API GPT4o-mini

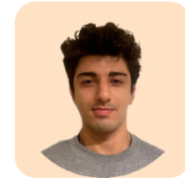
## - 배포 상세

- OpenAI API: 기능 단위로 코드를 분석 후 한글로 백로그를 생성하도록 함.



- [Llama3](#): 생성된 백로그를 바탕으로 결과 보고서를 생성함. 제목, 개요, 스택, 주요 기능, 활용 방안, 보완할 점, 완성도(원형 그래프 및 피드백)를 생성함.

# mustafaaljadery/lightning-whisper-mlx



An extremely fast implementation of whisper optimized for Apple Silicon using MLX.

3

Contributors

25

Used by

574

Stars

28

Forks



- [Lightning-whisper-mlx](#): 회의록 녹음 시 STT를 위한 모델로 활용함. 녹음된 파일에서 음성을 텍스트로 변환함.



- [Gemma-2-2b-it-4bit](#): 변환된 텍스트를 바탕으로 회의록을 요약함.