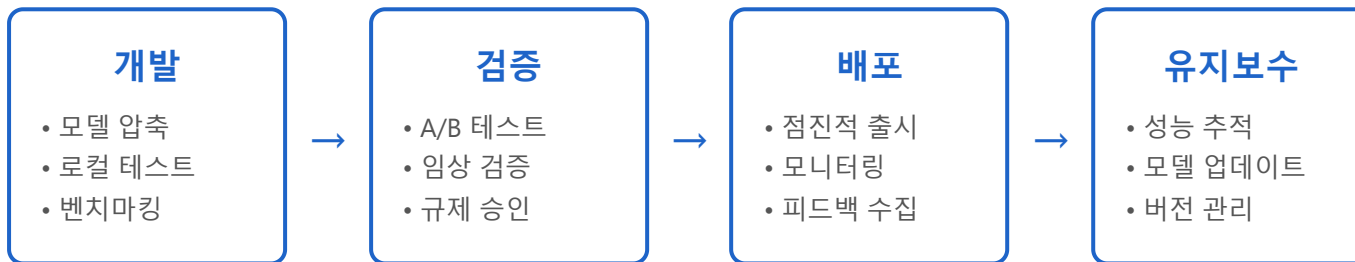


Deployment Strategies

배포 전략 및 체크리스트

압축 모델의 성공적인 프로덕션 배포를 위한 전략

배포 워크플로우



배포 전 체크리스트

성능 (Performance)

- ☐ 정확도 목표 달성 (Accuracy goal achieved)
- ☐ 지연시간 요구사항 충족 (Latency requirements met)
- ☐ 메모리 제약 내 (Within memory constraints)

안정성 (Stability)

- ☐ Edge case 테스트 (Edge case testing)
- ☐ 에러 핸들링 (Error handling)
- ☐ Fallback 메커니즘 (Fallback mechanism)

규제 (Regulatory)

- ☐ FDA/CE 인증 (필요시) (FDA/CE certification if needed)
- ☐ 개인정보 보호 (Data privacy)
- ☐ 설명 가능성 (Explainability)

운영 (Operations)

- ☐ 로깅/모니터링 (Logging/Monitoring)
- ☐ OTA 업데이트 (OTA updates)
- ☐ 사용자 교육 (User education)

플랫폼별 권장 전략

플랫폼	압축 방법	프레임워크
iOS	증류 + INT8	Core ML
Android	증류 + INT8	TF Lite
Web	경량 압축	TensorFlow.js
Embedded	INT4 + 가지치기	TF Lite Micro