

서귀포보건소 AI Agent 개발 데이터셋 및 DeepSeek R1 Fine-tuning 계획

1. 서귀포보건소 AI Agent 데이터셋 분류표

1.1 공개 데이터셋 (Public Dataset)

카테고리	데이터셋명	데이터 유형	출처	용도	데이터 규모	접근 방법
감염병 관리	질병관리청 감염병 감시 데이터	JSON/XML	질병관리청 오픈 API	감염병 예방 상담	10만+ 레코드	Open API
의료기관 정보	건강보험심사평가원 의료기관 정보	JSON	심평원 오픈 API	의료기관 추천	제주 지역 500+ 기관	Open API
의약품 정보	식품의약품안전처 의약품 정보	JSON/XML	식약처 오픈 API	의약품 정보 제공	30만+ 품목	Open API
건강검진	국민건강보험공단 건강검진 통계	CSV/JSON	건보공단 오픈 데이터	건강 상담 기준 데이터	연간 통계	공공데이터포털
지역 보건	제주특별자치도 보건의료 현황	PDF/XLSX	제주도청	지역 보건 정책 정보	연간 보고서	공공데이터포털
예방접종	질병관리청 예방접종 정보	JSON	질병관리청 API	예방접종 일정 안내	전 연령 백신 정보	Open API
응급의료	응급의료포털 응급실 정보	JSON	중앙응급 의료센터	응급실 현황 안내	실시간 데이터	Open API
건강생활	국민건강영양조사	SPSS/CSV	질병관리청	건강생활 상담	연간 2만+ 대상	마이크로데이터
정신건강	정신건강 실태조사	XLSX/CSV	보건복지부	정신건강 상담	전국 통계	공공데이터포털

모자보건	모성 및 영유아 건강관리	JSON/XLSX	보건복지부	모자보건상담	연간 통계	공공데이터포털
------	---------------	-----------	-------	--------	-------	---------

1.2 비공개 데이터셋 (Private/Internal Dataset)

카테고리	데이터셋명	데이터 유형	출처	용도	데이터 규모	보안 등급
민원 처리	서귀포보건소 민원 상담 이력	텍스트/JSON	내부 시스템	대화형 AI 학습	3년간 5만+ 건	높음
건강검진	서귀포 지역 건강검진 결과	암호화 DB	내부 DB	개인 맞춤형 건강 상담	10만+ 검진 기록	최고
예방접종	서귀포 지역 예방접종 기록	암호화 DB	예방접종 시스템	접종 일정 관리	5만+ 접종 기록	최고
방문 간호	방문간호 서비스 기록	텍스트/DB	방문간호 팀	방문 서비스 최적화	2년간 3천+ 방문	높음
건강교육	건강교육 프로그램 참여 데이터	XLSX/JSON	건강증진과	교육 프로그램 추천	1천+ 프로그램	보통
정신건강	정신건강 상담 기록	암호화 텍스트	정신건강 팀	정신건강 상담	2년간 2천+ 상담	최고
모자보건	임산부 및 영유아 관리 기록	암호화 DB	모자보건 팀	모자보건 서비스	3천+ 관리 대상	최고
치매관리	치매환자 관리 기록	암호화 DB	치매안심센터	치매 관리 서비스	1천+ 관리 대상	최고
금연상담	금연 상담 및 성공 사례	텍스트/JSON	금연클리닉	금연 상담 지원	2년간 5천+ 상담	높음
직원 매뉴얼	보건소 업무 매뉴얼	PDF/DOC	각 부서	업무 처리 가이드	50+ 매뉴얼	보통

2. DeepSeek R1 모델 Fine-tuning 30일 일정표

사전 준비 (Day -7 ~ Day 0)

하드웨어 준비: RTX 4090 × 2 (서귀포보건소 부서별 AI용) 소프트웨어 환경: Ubuntu 22.04, Python 3.10, PyTorch 2.1, Transformers 4.35

Day 1 (2025.07.14): 환경 구축 및 기본 설정

작업 내용:

- DeepSeek R1-7B 모델 다운로드 및 설치
- 개발 환경 구축 (CUDA, PyTorch, Transformers)
- 기본 의존성 패키지 설치
- 모델 로딩 테스트

담당자: AI 개발팀 리더 산출물: 개발 환경 구축 완료 보고서

Day 2 (2025.07.15): 공개 데이터 수집 및 전처리

작업 내용:

- 질병관리청 API 연동 및 데이터 수집
- 건강보험심사평가원 의료기관 정보 수집
- 식약처 의약품 정보 다운로드
- 데이터 포맷 표준화 작업

담당자: 데이터 엔지니어 산출물: 공개 데이터셋 수집 완료 (50GB)

Day 3 (2025.07.16): 비공개 데이터 수집 및 익명화

작업 내용:

- 서귀포보건소 내부 데이터 추출
- 개인정보 익명화 처리
- 데이터 보안 검토 및 승인
- 민원 상담 이력 텍스트 전처리

담당자: 보건소 정보담당자 + 보안팀 산출물: 익명화된 내부 데이터셋 (30GB)

Day 4 (2025.07.17): 데이터 품질 검증 및 정제

작업 내용:

- 중복 데이터 제거
- 데이터 무결성 검증

- 누락 데이터 처리
- 데이터 라벨링 품질 확인

담당자: 데이터 사이언티스트 산출물: 정제된 데이터셋 품질 보고서

Day 5 (2025.07.18): 학습 데이터셋 구성

작업 내용:

- Train/Validation/Test 분할 (8:1:1)
- 대화형 데이터 형태로 변환
- 프롬프트 템플릿 설계
- 데이터 로더 구현

담당자: ML 엔지니어 산출물: 학습용 데이터셋 구성 완료

Day 6-7 (2025.07.19-20): 주말 - 시스템 안정성 검증

작업 내용:

- 메모리 사용량 최적화
- GPU 활용률 모니터링
- 백업 시스템 구축

Day 8 (2025.07.21): Fine-tuning 전략 수립

작업 내용:

- LoRA (Low-Rank Adaptation) 설정
- 하이퍼파라미터 튜닝 계획 수립
- 학습률, 배치 크기, 에폭 수 결정
- 체크포인트 저장 전략 설정

담당자: AI 연구원 산출물: Fine-tuning 전략 문서

Day 9 (2025.07.22): 보건소 도메인 특화 프롬프트 설계

작업 내용:

- 보건소 업무별 프롬프트 템플릿 개발
- 시스템 프롬프트 설계
- 응답 형식 표준화

- 전문 용어 사전 구축

담당자: 보건소 전문가 + AI 개발팀 산출물: 도메인 특화 프롬프트 라이브러리

Day 10 (2025.07.23): 1차 Fine-tuning 시작

작업 내용:

- 기본 의료 상식 데이터로 사전 학습
- 일반적인 보건 관련 Q&A 학습
- 학습 진행 상황 모니터링
- 초기 성능 평가

담당자: ML 엔지니어 산출물: 1차 모델 체크포인트

Day 11 (2025.07.24): 보건소 특화 데이터 Fine-tuning

작업 내용:

- 서귀포보건소 민원 데이터 학습
- 지역 특성 반영 데이터 학습
- 실시간 학습 모니터링
- 과적합 방지 조치

담당자: AI 개발팀 산출물: 보건소 특화 모델 v1.0

Day 12 (2025.07.25): 모델 성능 평가 및 개선

작업 내용:

- 테스트 데이터로 성능 평가
- BLEU, ROUGE 점수 측정
- 의료진 전문가 평가
- 개선점 도출

담당자: 평가팀 + 의료진 산출물: 성능 평가 보고서

Day 13-14 (2025.07.26-27): 주말 - 모델 최적화

작업 내용:

- 모델 경량화 작업

- 추론 속도 최적화
 - 메모리 사용량 최적화
-

Day 15 (2025.07.28): 2차 Fine-tuning

작업 내용:

- 1차 평가 결과 반영하여 재학습
- 부족한 영역 데이터 추가 학습
- 하이퍼파라미터 조정
- 학습 안정성 확인

담당자: AI 개발팀 산출물: 개선된 모델 v2.0

Day 16 (2025.07.29): 다중 업무 처리 능력 강화

작업 내용:

- 복합 질의 처리 학습
- 업무 연계 시나리오 학습
- 다단계 추론 능력 향상
- 컨텍스트 유지 능력 강화

담당자: AI 연구원 산출물: 다중 업무 처리 모델

Day 17 (2025.07.30): 실시간 대화 능력 Fine-tuning

작업 내용:

- 대화 흐름 자연스러움 개선
- 공감 표현 능력 향상
- 전문 용어 설명 능력 강화
- 맞춤형 응답 생성 학습

담당자: NLP 전문가 산출물: 대화형 AI 모델 v3.0

Day 18 (2025.07.31): 보안 및 개인정보보호 강화

작업 내용:

- 개인정보 유출 방지 학습
- 적절한 정보 공개 수준 학습

- 보안 정책 준수 확인
- 민감 정보 처리 규칙 적용

담당자: 보안팀 + AI 개발팀 산출물: 보안 강화 모델

Day 19 (2025.08.01): 통합 테스트 및 시스템 연동

작업 내용:

- 보건소 기존 시스템과 API 연동
- 전체 워크플로우 테스트
- 시스템 안정성 검증
- 부하 테스트 수행

담당자: 시스템 통합팀 산출물: 시스템 연동 완료

Day 20-21 (2025.08.02-03): 주말 - 종합 테스트

작업 내용:

- 전체 시나리오 테스트
- 성능 벤치마크 수행
- 버그 수정 및 안정성 확보

Day 22 (2025.08.04): 사용자 시나리오 기반 테스트

작업 내용:

- 실제 민원 사례로 테스트
- 보건소 직원 사용성 평가
- 응답 정확도 검증
- 사용자 피드백 수집

담당자: 보건소 직원 + 테스트팀 산출물: 사용자 테스트 결과 보고서

Day 23 (2025.08.05): 3차 Fine-tuning (최종 조정)

작업 내용:

- 사용자 피드백 반영
- 최종 성능 최적화
- 안정성 확보

- 배포 준비

담당자: AI 개발팀 전체 산출물: 최종 배포용 모델 v4.0

Day 24 (2025.08.06): 문서화 및 매뉴얼 작성

작업 내용:

- 사용자 매뉴얼 작성
- 관리자 가이드 작성
- API 문서화
- 트러블슈팅 가이드 작성

담당자: 문서화팀 산출물: 완전한 문서 패키지

Day 25 (2025.08.07): 직원 교육 준비

작업 내용:

- 교육 프로그램 개발
- 교육 자료 제작
- 실습 환경 구축
- 교육 일정 수립

담당자: 교육팀 산출물: 교육 프로그램 완성

Day 26 (2025.08.08): 파일럿 테스트

작업 내용:

- 제한된 범위에서 실제 서비스 시작
- 실시간 모니터링
- 즉시 피드백 수집
- 긴급 대응 체계 가동

담당자: 전체 프로젝트팀 산출물: 파일럿 테스트 결과

Day 27-28 (2025.08.09-10): 주말 - 최종 점검

작업 내용:

- 파일럿 결과 분석

- 최종 수정 사항 적용
- 전체 시스템 안정성 재확인

Day 29 (2025.08.11): 직원 교육 실시

작업 내용:

- 보건소 전 직원 대상 교육
- 실습 및 Q&A 세션
- 매뉴얼 배포
- 지원 체계 안내

담당자: 교육팀 + 프로젝트팀 산출물: 교육 완료 인증서

Day 30 (2025.08.12): 정식 서비스 시작

작업 내용:

- 서귀포보건소 AI Agent 정식 오픈
- 실시간 모니터링 체계 가동
- 시민 공지 및 홍보
- 성공 기념 행사

담당자: 전체 프로젝트팀 산출물: 서비스 오픈 완료

3. 기술 사양 및 요구사항

3.1 하드웨어 요구사항

- **GPU:** RTX 4090 × 2 (서귀포보건소용)
- **RAM:** 64GB DDR4
- **Storage:** NVMe SSD 2TB
- **Network:** 1Gbps 전용선

3.2 소프트웨어 스택

- **OS:** Ubuntu 22.04 LTS
- **Python:** 3.10+
- **ML Framework:** PyTorch 2.1, Transformers 4.35
- **API:** FastAPI, Nginx
- **Database:** PostgreSQL (암호화)
- **Monitoring:** Prometheus, Grafana

3.3 보안 요구사항

- 데이터 암호화: AES-256
- 접근 제어: RBAC (Role-Based Access Control)
- 감사 로그: 모든 접근 기록 저장
- 개인정보보호: K-익명성 보장

4. 성공 지표 (KPI)

4.1 기술적 지표

- 응답 정확도: 95% 이상
- 응답 시간: 3초 이내
- 시스템 가용성: 99.9% 이상
- 동시 접속자: 100명 이상

4.2 업무 효율성 지표

- 민원 처리 시간: 50% 단축
- 1차 해결률: 80% 이상
- 직원 만족도: 4.5/5.0 이상
- 시민 만족도: 4.5/5.0 이상

5. 위험 관리 계획

5.1 기술적 위험

- 모델 성능 부족: 추가 데이터 수집 및 재학습
- 시스템 장애: 이중화 시스템 구축
- 보안 위협: 24시간 보안 모니터링

5.2 운영적 위험

- 직원 저항: 충분한 교육 및 인센티브 제공
- 시민 불신: 투명한 정보 공개 및 소통
- 예산 초과: 단계별 예산 관리

※ 본 계획은 서귀포보건소의 실제 상황에 맞춰 조정될 수 있으며, 각 단계별 완료 후 다음 단계 진행 여부를 검토합니다.