

제주국세청 AI Agent 개발 데이터셋 및 Fine-tuning 일정

데이터셋 분류표

공개 데이터셋

분류	데이터명	출처	규모	용도
세법·규정	국세기본법 및 시행령	국가법령정보센터	~800MB	세법 기본 원리 학습
	소득세법 및 시행령·규칙	국가법령정보센터	~1.2GB	소득세 업무 처리
	법인세법 및 시행령·규칙	국가법령정보센터	~900MB	법인세 업무 처리
	부가가치세법 및 시행령·규칙	국가법령정보센터	~600MB	부가가치세 업무
	상속세및증여세법	국가법령정보센터	~400MB	상속·증여세 업무
	종합부동산세법	국가법령정보센터	~200MB	종부세 업무
조세정책	국세청 조세정책 자료	국세청 홈페이지	~300MB	최신 세정 동향
	지방국세청 공지사항	각 지방국세청	~150MB	지역별 세정 현황
	세정개혁 방안	국세청	~100MB	정책 변화 대응
행정절차	국세 민원처리 안내	국세청	~250MB	민원 처리 절차
	홈택스 이용 가이드	국세청	~200MB	전자세정 서비스
	세무조사 절차 매뉴얼	국세청	~300MB	조사 업무 프로세스
	체납처분 절차 안내	국세청	~180MB	징세 업무
조직·업무	국세청 조직도 및 업무분장	국세청	~50MB	조직 구조 이해
	세무서 표준 업무분장	국세청	~80MB	부서별 역할

	지방국세청 현황	국세청	~40MB	지역별 특성
양식·서식	각종 신고서 양식	국세청, 홈택스	~1GB	신고서 작성 지원
	민원증명 발급 양식	국세청	~300MB	증명서 안내
	세무조사 관련 서식	국세청	~200MB	조사 업무 서식
FAQ·사례	국세청 자주묻는질문	국세청 홈페이지	~400MB	일반 세무 상담
	세목별 해석사례	국세법령정보시스템	~600MB	법령 해석
	국세심판소 판례	국세심판소	~800MB	분쟁 사례 학습
통계·현황	국세통계연보	국세청	~500MB	세수 현황 분석
	지역별 세수 현황	국세청	~200MB	제주 지역 특성
	업종별 과세 현황	국세청	~300MB	업종별 특징

공개 데이터 총합: 약 **9.0GB**

비공개 데이터셋 (내부 구축 필요)

분류	데이터명	출처	예상 규모	용도
실제 업무사례	최근 5년간 세무상담 사례	제주세무서·서귀포지서	~3GB	실제 상담 상황 학습
	세목별 처리 사례	각 세무과	~2.5GB	세목별 전문 지식
	복잡 사안 해결 과정	조사과, 징세과	~1.5GB	복합 업무 처리
	이의신청·심판청구 사례	납세자보호실	~800MB	분쟁 해결 노하우
내부 매뉴얼	부서별 업무 처리 지침	각 부서	~600MB	세무 업무 절차
	세무조사 실무 매뉴얼	조사과	~400MB	조사 기법 및 절차

제주 특화정보	체납징수 실무 가이드	징세과	~300MB	징수 업무 노하우
	고객응대 스크립트	민원봉사실	~200MB	상담 화법 학습
	제주 지역 주요 납세자 유형	제주세무서	~400MB	지역 특성 반영
	관광업 과세 특례 사례	제주세무서	~300MB	관광산업 특화
	농업·어업 세무 처리	제주세무서	~350MB	1차 산업 특성
직원 노하우	중국인 투자 관련 세무	제주세무서	~250MB	외국인 투자 대응
	숙련 세무공무원 인터뷰	현장 조사	~500MB	실무 경험 학습
	부서장급 업무 노하우	관리자 인터뷰	~400MB	관리 업무 기법
	어려운 사안 해결 사례	전문가 인터뷰	~300MB	고급 업무 처리
납세자 피드백	민원 만족도 조사 결과	민원봉사실	~150MB	서비스 개선점
	개선 요청 사항	납세자 설문	~200MB	사용자 니즈 파악
	불만 사례 및 해결책	민원 분석팀	~180MB	문제 상황 대응
시스템 연계정보	홈택스 연계 프로토콜	전산담당	~100MB	시스템 연동
	타 기관 정보 공유	정보통신담당	~150MB	기관간 협업
	데이터베이스 구조	IT팀	~120MB	데이터 처리 최적화

비공개 데이터 총합: 약 **12.8GB**



DeepSeek R1-32B Fine-tuning 일일 작업 일정

Phase 1: 환경 구축 및 데이터 준비 (1-10일차)



Day 1: 하드웨어 환경 구축

- GPU 서버 설정 (A100 40GB × 4 구성)
- CUDA 12.1, PyTorch 2.1+ 설치
- DeepSeek R1-32B 모델 다운로드 및 테스트
- 개발 환경 최적화 및 벤치마크

Day 2: 공개 데이터 수집 시작

- 국가법령정보센터 세법 데이터 수집
- 국세청 홈페이지 자료 크롤링
- 홈택스 가이드 및 FAQ 수집
- 데이터 저장소 구축 및 1차 분류

Day 3: 세법 데이터 전처리

- 법령 조문 구조화 및 계층 분석
- 시행령·시행규칙 연계 구조 파악
- 개정 연혁 및 적용 시점 정리
- 법령 인용 관계 매핑

Day 4: 행정절차 데이터 정리

- 민원처리 절차 플로우차트 작성
- 부서별 업무 프로세스 매핑
- 예외상황 처리 방법 정리
- 권한별 업무 범위 정의

Day 5: 비공개 데이터 수집 계획

- 제주세무서·서귀포지서 협조 요청
- 개인정보 비식별화 프로토콜 수립
- 데이터 수집 일정 및 담당자 배정
- 보안 가이드라인 설정

Day 6: 제주 특화 데이터 수집

- 제주 관광산업 세무 특례 수집
- 농축수산업 과세 특성 자료 수집
- 외국인 투자 관련 세무 사례 수집
- 지역 경제 특성 반영 자료 정리

Day 7: 데이터 전처리 도구 개발

- 세무 전문용어 사전 구축
- 자동 식명화 도구 개발
- 데이터 품질 검증 스크립트 작성
- 세법 조문 파싱 도구 개발

Day 8: 학습 데이터 구조 설계

- 세무상담 Q&A 형태로 변환

- 멀티턴 대화 시나리오 설계
- 복잡한 세무상담 시뮬레이션 구조
- 평가 데이터셋 분리 (80:10:10)

Day 9: Fine-tuning 설정 최적화

- LoRA 설정 (rank=16, alpha=32)
- 학습률 스케줄링 ($2e-5 \rightarrow 1e-6$)
- 배치 크기 및 그래디언트 누적 설정
- 조기 종료 및 체크포인트 전략

Day 10: 1차 데이터 통합 테스트

- 수집된 공개 데이터 통합
- 데이터 일관성 및 품질 검증
- 샘플 학습 테스트 실행
- 초기 모델 성능 벤치마크

Phase 2: 기본 세무 도메인 학습 (11-25일차)

Day 11-13: 세법 기초 학습

- 국세기본법 중심 기본 개념 학습
- 세법 체계 및 용어 정의 학습
- 과세요건 및 절차법 학습
- 세법 해석 원칙 학습

Day 14-17: 주요 세목 전문 학습

- 소득세법 상세 학습 (개인소득, 사업소득)
- 법인세법 중점 학습 (법인 과세체계)
- 부가가치세법 실무 학습 (매입·매출세액)
- 상속·증여세 특례 학습

Day 18-21: 행정절차 프로세스 학습

- 신고·납부 절차 상세 학습
- 세무조사 절차 및 권리구제 학습
- 체납처분 및 징수절차 학습
- 민원처리 및 상담 프로세스 학습

Day 22-25: 1차 평가 및 보완

- 세법 해석 정확도 테스트
- 절차 안내 적절성 검증
- 부족한 영역 식별 및 추가 학습
- 모델 성능 지표 측정

Phase 3: 실무 사례 통합 학습 (26-40일차)

17 Day 26-29: 실제 상담 사례 학습

- 5년간 세무상담 사례 통합 학습
- 개인정보 비식별화 완료된 데이터 활용
- 상담 유형별 답변 패턴 학습
- 복잡한 질의에 대한 단계별 답변 학습

17 Day 30-33: 제주 특화 업무 학습

- 관광산업 세무 특례 심화 학습
- 농축수산업 과세 특성 반영
- 외국인 투자자 세무 상담 학습
- 지역 경제 특성 기반 상담 능력 개발

17 Day 34-37: 복합 업무 처리 학습

- 다부서 연계 업무 처리 학습
- 분쟁 및 이의신청 대응 학습
- 세무조사 상황별 대응 학습
- 체납처분 관련 상담 학습

17 Day 38-40: 고급 상담 스킬 학습

- 감정적 납세자 대응 기법 학습
- 복잡한 세법 해석 설명 기법
- 상황별 적절한 톤앤매너 학습
- 전문가 수준 상담 능력 개발

Phase 4: 최종 최적화 및 배포 준비 (41-50일차)

17 Day 41-43: 성능 최적화

- 모델 경량화 (INT8 Quantization)
- 추론 속도 최적화
- 메모리 사용량 최소화
- 동시 처리 능력 향상

17 Day 44-46: 통합 시스템 테스트

- 토크스 연계 테스트
- 실시간 상담 시뮬레이션
- 대용량 동시 접속 테스트
- 장애 상황 복구 테스트

17 Day 47-48: 실무진 검증

- 현직 세무공무원 블라인드 테스트
- 복잡 사안 처리 능력 검증
- 답변 정확도 및 적절성 평가
- 개선사항 도출 및 반영

Day 49: 최종 시스템 점검

- 보안 시스템 점검
- 개인정보 보호 체계 확인
- 모니터링 시스템 구축
- 백업 및 복구 체계 점검

Day 50: 배포 및 런칭

- 프로덕션 환경 배포
- 초기 운영 모니터링
- 사용자 가이드 배포
- 운영진 교육 완료

최종 성과 지표

평가 항목	목표 수치	측정 방법
세법 해석 정확도	98% 이상	전문가 검증
절차 안내 정확도	95% 이상	실무진 평가
응답 속도	2초 이내	시스템 측정
사용자 만족도	4.7/5.0 이상	납세자 설문
상담 해결률	85% 이상	추가 문의율 역산
처리 시간 단축	75% 이상	before/after 비교

제주 특화 기능

관광산업 특화 서비스

- 숙박업 부가가치세 특례 안내
- 외국인 관광객 대상 사업 과세
- 관광단지 투자 세제혜택 상담
- 면세점 운영 관련 세무 처리

1차 산업 특화 서비스

- 농업소득 필요경비 산정
- 어업 관련 소득세 신고
- 축산업 부가가치세 면세 범위
- 농어촌 특별세 감면 안내

외국인 투자 특화 서비스

- 외국법인 한국 지정 세무
- 부동산 취득·양도 시 세무처리
- 이중과세방지협정 적용
- 영주권자 세무 의무

지속적 개선 계획

- 주간: 새로운 세법 해석사례 업데이트
- 월간: 제주 지역 특성 반영 사례 추가
- 분기: 세법 개정사항 반영 및 재학습
- 반기: 대규모 성능 개선 및 기능 확장
- 연간: 차세대 모델 도입 및 전면 업그레이드

핵심 성공 요인

1. 제주 특화: 관광산업과 1차 산업 중심의 지역 특성 반영
2. 실무 중심: 실제 세무상담 사례와 현장 노하우 활용
3. 법령 정확성: 최신 세법 및 해석사례 정확한 반영
4. 사용자 편의: 복잡한 세법을 쉽게 설명하는 능력
5. 지속 업데이트: 세법 변경사항 신속 반영 체계