

# 웰홈(WELHOME)

사람과 정책 사이, 더 쉽게, 더 가깝게

---

이창민, 이유진, 전하연

## 목차

- 1. 프로젝트 개요**
- 2. 프로젝트 팀 구성 및 역할**
- 3. 프로젝트 수행 절차 및 방법**
- 4. 프로젝트 수행 경과**
- 5. 자체 평가 의견**

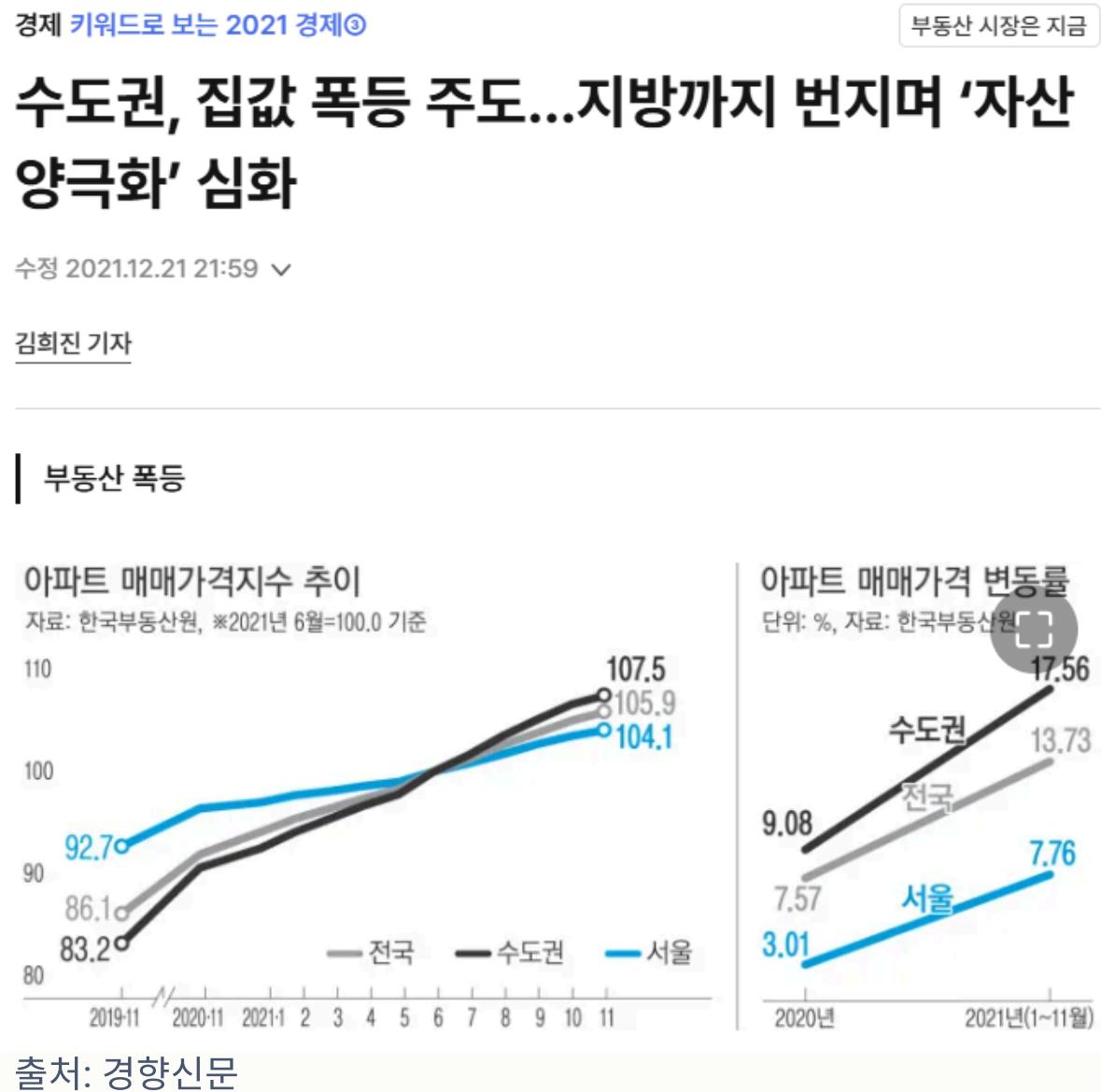


# 1. 프로젝트 개요



# 프로젝트 개요

시대가 지날 수록 집값은 **폭등** → 내 집 마련 부담



20~24세 청년이 월급 전부 모아 서울 아파트 사는 데 걸리는 기간

2000년 16.05년

2022년 24.76년

자료: 김준형, 박순만 명지대 부동산학과 교수

# 프로젝트 개요

## “주거 관련 지원 사업에 대한 낮은 인지도”

### 주거복지 관련 사업 인지도

- 자료갱신일 : 2022-09-22 / 수록기간 : 2년 2020 ~ 2020 / 자료문의처 : 02-2133-4365
- 출처 : 서울특별시, 서울특별시 복지실태조사

항목[1/1] 특성별[22/22] 주거복지별[9/9] 인지도별[5/5] 시점[1/1]

새창보기 열고정 화면복사 주소보기 행열전환 부가기능설정 분석 차트 다운로드 인쇄 도움말

특성별(1)	특성별(2)	2020					
		사례수(명)	모름(%)	들어보았지만, 내용은 모름(%)	내용을 어느 정도 알고 있음(%)	비교적 자세히 알고 있음(%)	
전체	소계	3,027	5	26	51	18	
5권역별	도심권	163	8	34	42	17	
	동북권	944	1	21	63	16	
	서북권	365	2	22	52	25	
	서남권	951	8	35	38	19	
	동남권	604	7	20	55	18	
응답자성별	남성	779	4	24	51	20	
	여성	2,248	5	27	51	18	
응답자연령별	20대 이하	371	6	35	47	13	
	30~40대	1,259	4	25	50	21	
	50~60대	1,144	5	24	55	16	
	70대 이상	253	5	25	46	25	
연간총가구소득별	1천만원 미만	78	0	3	51	46	

출처: 서울열린광장

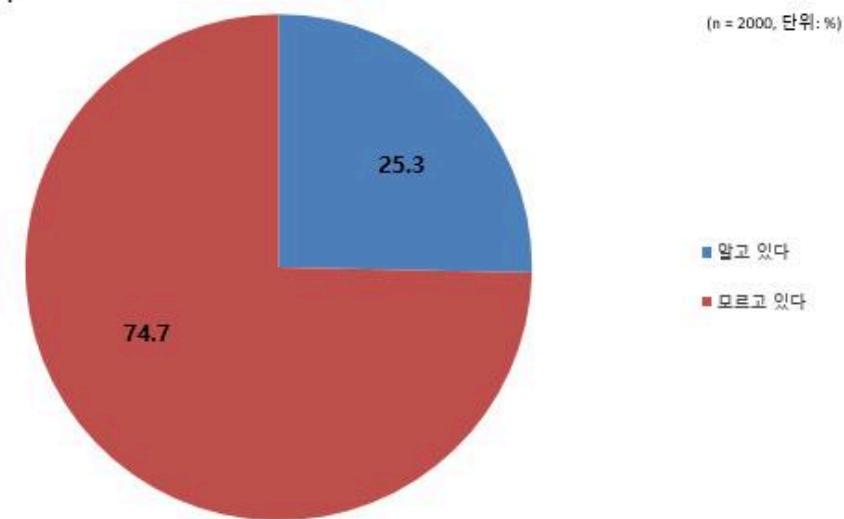
홈 > 시사 > 시사

### 서울시민 10명 중 7명 이상이 모르는 '1인 가구 지원사업'

손진석 기자 | 입력 2024.01.05 11:16 | 댓글 0

#### 서울시의원, 서울시 1인가구 지원사업 여론조사 발표

서울시 1인 가구 사업 인지 여부



출처: 네이트뉴스

나의 아이는  
나의 상표는

지역지식재산센터 16



Global Intelligent  
WiSDOM

Google Ads

# LH 임대주택

---

공공주택 특별법에 따라 무주택 서민의 주거안정을 위해 제공되는 임대주택

LH는 정부 위탁을 받아 전국적으로 임대주택을 건설·매입·관리

신혼부부, 청년, 고령자, 저소득층 등 생애주기별 맞춤형 공급

## 주요 수혜 대상

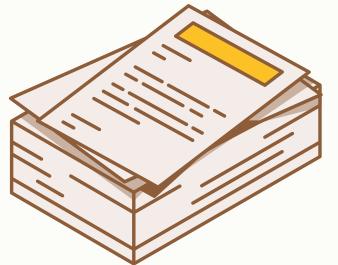
- 청년층: 사회초년생, 대학생, 취업준비생
- 신혼부부 및 한부모가구
- 고령자 및 장애인
- 저소득 무주택 세대

## 지원 자격 (기본 요건)

- 무주택 세대구성원일 것
- 소득 요건 충족 (도시근로자 가구소득 대비 일정 기준 이하)
- 자산 기준 충족 (부동산, 자동차, 금융자산 등)

# 편의성

---



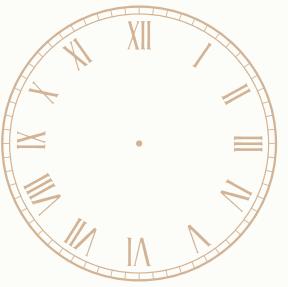
## 공고문

100건 이상 공고가 게시되며,  
유형·단지별로 분산되어 있음



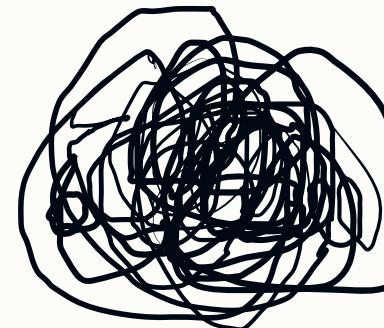
## 조건 확인의 비효율성

하나하나 PDF를 열어봐야 하며,  
직접 눈으로 조건을 대조해야 함



## 시간 소모가 큼

조건 분석과 판단에 1건당 10~15분 이상 소요,  
전체 검토 시 수 시간 이상 필요



## 내용의 난해함

소득, 자산, 우선공급 기준이 전문용어  
위주로 되어 있어 해석이 어려움



## 지원 기회를 놓침

지원 가능한데 몰라서 신청 못하거나  
신청했지만 조건 미충족으로 탈락

# 당첨 후 자금 문제

---

## 당첨됐지만, 돈이 없어 계약 못 하는 현실

- 신혼희망타운, 공공분양, 공공전세형 등  
→ 입주하려면 계약금만 최소 2,000만 원 ~ 2억 원 이상 필요
- 저소득·청년·신혼 대상자 대부분  
→ 청약은 가능하지만 입주 자금이 없어 포기
- LH 시스템 내 대출 안내 없음  
→ “누가, 얼마까지, 어떤 조건으로 대출 가능한가요?”에 대한 정보 전무
- 실제로 많은 사용자들이  
→ 계약 포기 또는 입주 지연 발생

# 해결 방안

## ① 사용자 조건 기반 공고 추천

- 소득, 지역, 가구, 자산 등 입력 시
- → 해당 조건에 부합하는 모집 공고만 선별 추천
- 더 이상 100개 PDF를 열어볼 필요 없음

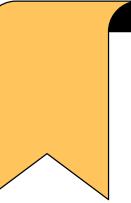
## ② 전문용어 및 정책 해석 - QnA 챗봇 탑재

- “무주택세대란?”, “우선공급이 뭐야?”
- 정책 전문용어를 쉬운 일상어로 설명
- 파인튜닝 기반 → 다양한 방식의 질문 대응 가능
- 사용자 이해력 증대 → 정보 격차 해소

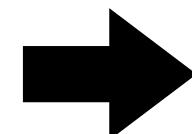
## ③ 대출 부담 대응 - 대출 챗봇 실시간 응답

- 사용자의 대출 관련 질문에
- → 자격 조건 / 한도 / 이자율 / 상환금 시뮬레이션 제공

# 설계도



사용자



임대 주택 추천

대출 추천

QnA, 용어 설명



사용자 조건 입력

용어 설명  
QnA



chroma

RAG

SQLite



임대 공고 추천

대출 상품 추천

# 타임라인

---

5월 26일

주제 선정



6월 9일

기술 스택 및 도구 선정



6월 2일

데이터 수집 및 전처리



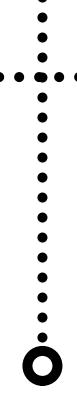
6월 17일

시스템 구현 및 통합



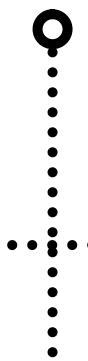
6월 13일

시스템 설계



7월 04일

최종발표



07월 01일

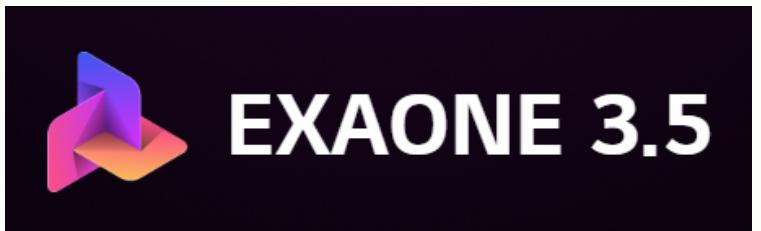
테스트 및 검증

# 기술 스택

Python



LLM



BGE-M3 임베딩 모델



분양 임대 API



HTML 형식으로 파싱



데이터 전처리



PDF 이미지 추출



벡터 DB 저장소



경량형 관계형 DB



LLM 기반 대화 흐름 구성



상태 기반 대화 에이전트 설계



PEFT (LoRA 방식) 기반 파인튜닝



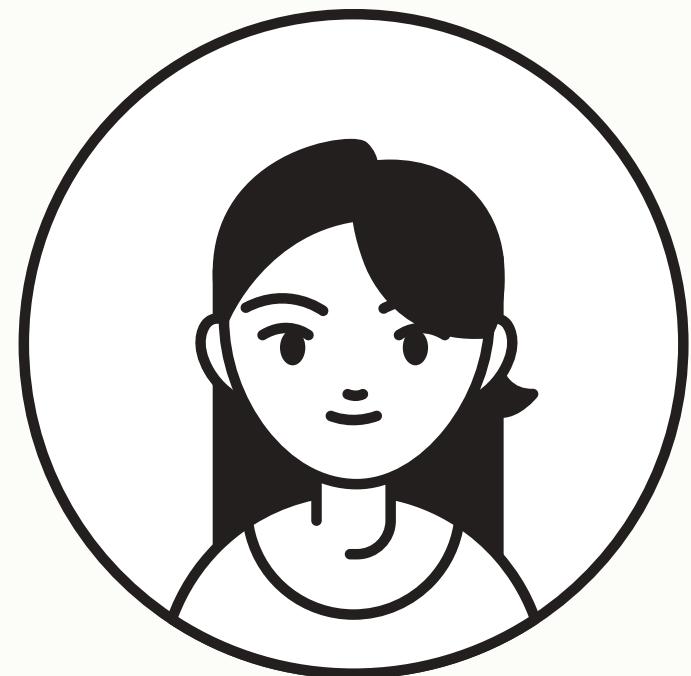
REST API 서버 구축

## 2. 팀 구성 및 역할



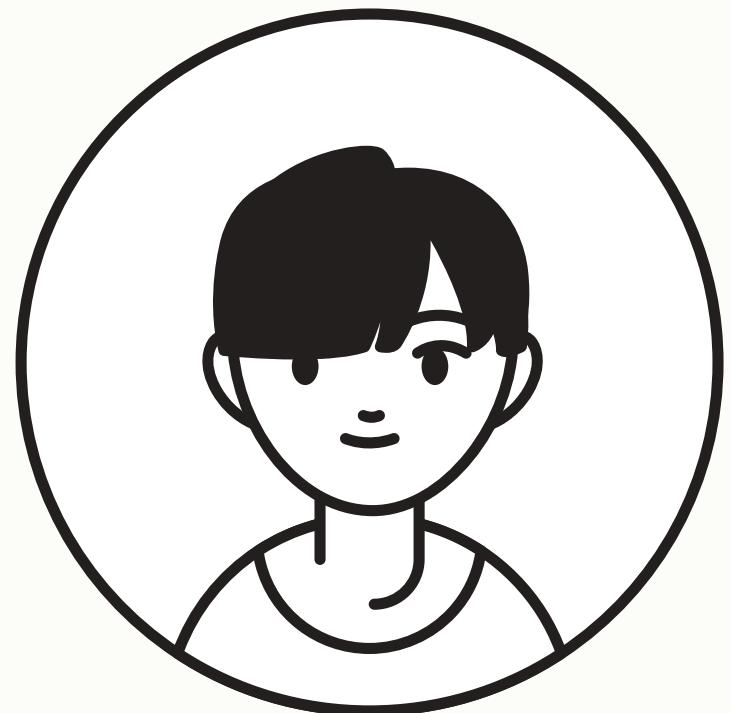
## 2. 팀 구성 및 역할

---



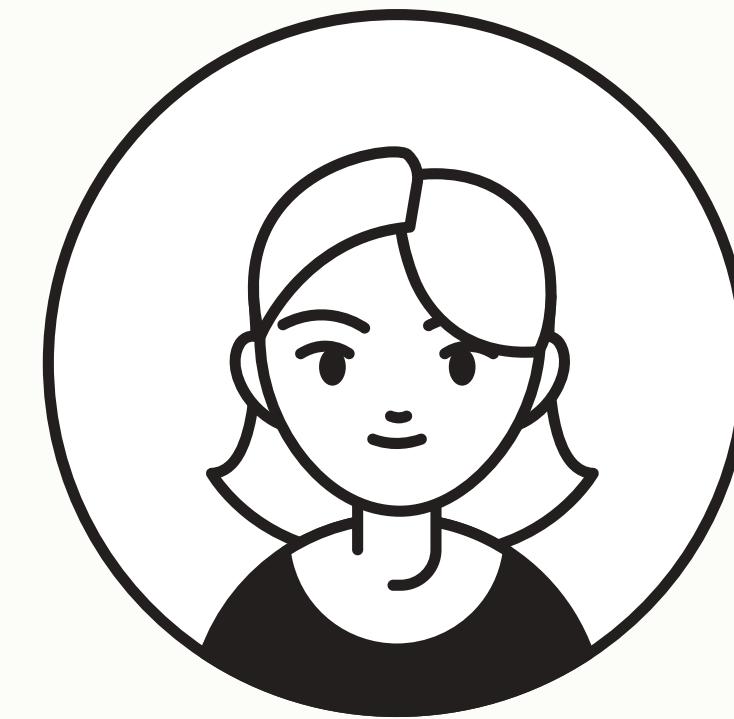
이유진

주택 대출 에이전트 개발  
데이터 수집 및 전처리  
데이터베이스 관리



이창민

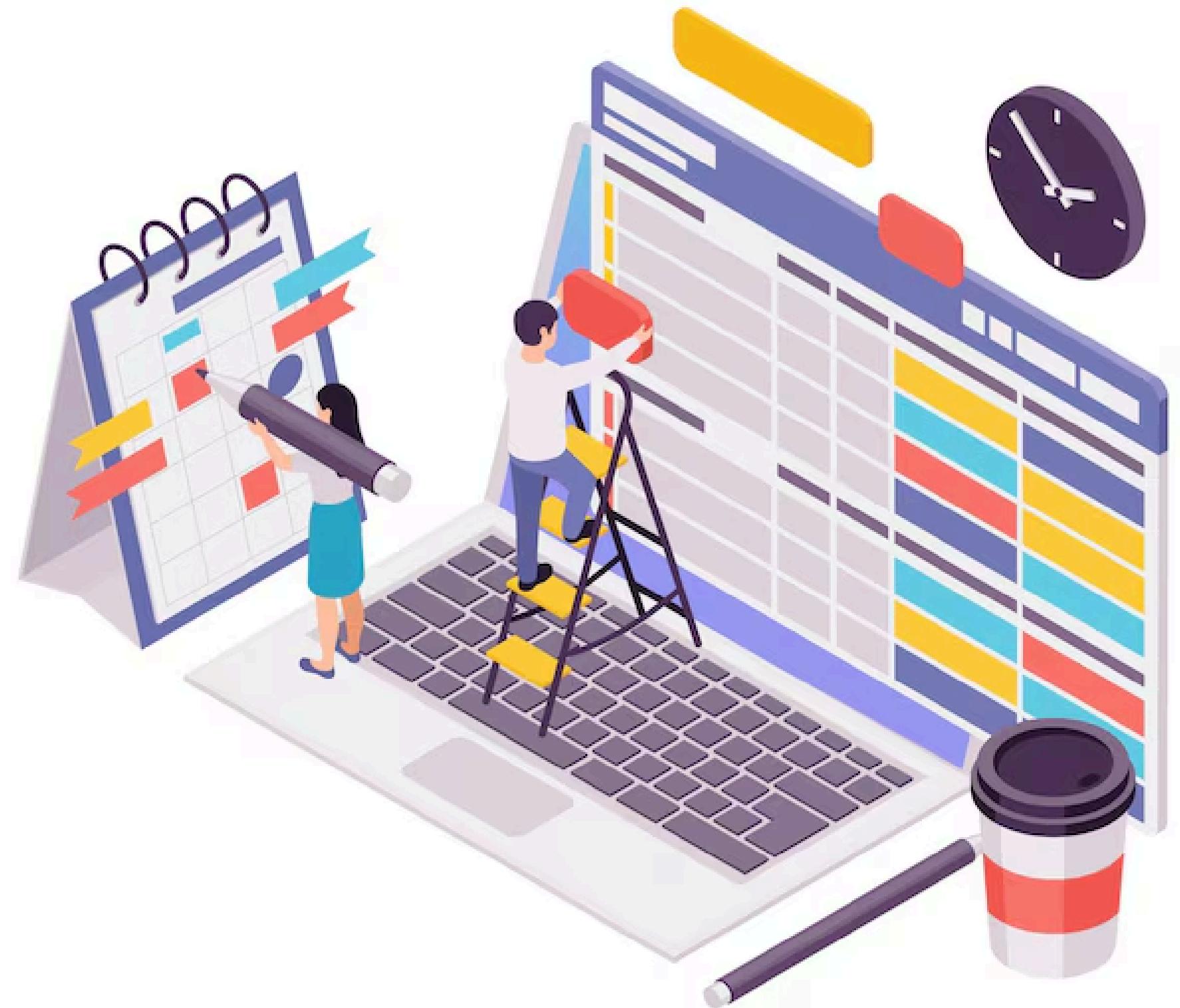
모델 파인튜닝  
데이터셋 준비  
모델 성능 최적화



전하연

프론트엔드 개발

### 3. 프로젝트 수행 절차 및 방법



# 사용 데이터

---

## 1. LH 청약 플러스 (<https://www.myhome.go.kr/hws/portal/wrd/selectWordDicView.do>)

LH 임대주택 공고문(PDF)

## 2. 공공데이터 포털([data.go.kr](http://data.go.kr))

- 한국토지주택공사\_분양임대공고문 조회 서비스
- 한국토지주택공사\_분양임대공고별 상세정보 조회 서비스
- 한국토지주택공사\_분양임대공고별 공급정보 조회 서비스

## 3. 마이홈 용어 사전(<https://www.myhome.go.kr/hws/portal/wrd/selectWordDicView.do>)

공고문에 쓰이는 전문용어(용어 사전)

## 4. 은행연합회 소비자포털(<https://www.bokjiro.go.kr/>)

대출 상품 관련 데이터 수집

# PDF 데이터 추출 과정 및 도구 선택 배경

## 1. 텍스트 추출의 한계

PDF에서 텍스트를 그대로 추출할 경우 다음과 같은 문제가 발생

- 표 형식 데이터의 구조가 무너짐
- 열 제목의 의미가 사라져 문맥 파악이 어려움

2. 임대대상 및 임대조건																	
공급 형별	공급 대상	건설 호수	현재 예비 자수	모집 예비 자수	세대 당 계약면적(m <sup>2</sup> )			임대조건				전환 가능 보증금 한도액 (천원)	최대전환시 임대조건	구조 및 난방			
					그 밖의 공용면적		합계	임대보증금(천원)									
					기타 공용	주차장		계	계약금	잔금	월 임대료 (원)						
16A	대학생	35	6	15	16.77	9.4776	7.3900	7.6709	41.3085	33,320	1,666	31,654	133,280	(+12,000)	45,320	63,280	
	청년 (소득없음)									(-28,000)		5,320	214,940				
	청년 (소득있음)									35,280	1,764	33,516	141,120	(+13,000)	48,280	65,280	
19A	대학생	8	0	8	19.38	10.9527	8.5401	8.8647	47.7375	38,420	1,921	36,499	153,680	(+15,000)	53,420	66,180	
	청년 (소득없음)									(-32,000)		6,420	247,010				
	청년 (소득있음)									40,680	2,034	38,646	162,720	(+16,000)	56,680	69,380	
최대전환시 임대조건 구조 및 난방 주거 전용 주거 공용 그 밖의 공용면적 합계 임대보증금천원 ( ) 세대 당 계약면적 ( ) m <sup>2</sup> 임대 조건 전환 가능 보증금 한도액 천원 월 임대료 원 ( ) 기타 공용 주차장 계 계약금 잔금 임대																	

# PDF 데이터 추출 과정 및 도구 선택 배경

## 2. 테이블 구조 보존 시도

Camelot, pdfplumber 등 라이브러리 활용

- 표 구조를 최대한 보존하기 위해 다양한 도구 시도

0	단지명	주 n택 n형	임대조건	None	None	None	전환가능 n보증금 n한도액(원)	최대전환시 임대조건	None
1	None	None	임대보증금(원)	None	None	월임대료 n(원)	None	None	None
2	None	None	None	None	None	None	보증금 n(원)	월임대료 n(원)	None
3	None	None	계	계약금	잔금	None	None	None	None
4	대전도안 1단지	33	19,648,000	982,000	18,666,000	148,000	(+) 14,000,000	33,648,000	66,330

### 병합 셀 처리 : 결측치 보정 규칙 적용

- 첫 번째 행은 타이틀로 간주
- 0번째 열(x축)에서 셀 값이 비어 있을 경우 → 왼쪽 셀의 값을 복사하여 채움
- 1행 이후의 행(x축)에서 셀 값이 비어 있을 경우 → 바로 위 셀의 값을 복사하여 채움

0	단지명	주 n택 n형	임대조건	임대조건	임대조건	전환가능 n보증금 n한도액(원)	최대전환시 임대조건	최대전환시 임대조건
1	단지명	주 n택 n형	임대보증금(원)	임대조건	임대조건	월임대료 n(원)	전환가능 n보증금 n한도액(원)	최대전환시 임대조건
2	단지명	주 n택 n형	임대보증금(원)	임대조건	임대조건	월임대료 n(원)	전환가능 n보증금 n한도액(원)	보증금 n(원)
3	단지명	주 n택 n형	계	계약금	잔금	월임대료 n(원)	전환가능 n보증금 n한도액(원)	보증금 n(원)
4	대전도안 1단지	33	19,648,000	982,000	18,666,000	148,000	(+) 14,000,000	33,648,000

### 한계점

- 공고문마다 표 구성 방식이 달라 일반화 어려움

유형	대상주택		세대당 계약면적(m <sup>2</sup> )			임대조건			입주 예정	
	호수	방	전용	공용	합계	임대보증금(원)		월임대료 (원)		
						계	계약금	잔금		
303	1	15.4053	6.2365	21.6418	2,748,000	274,000	2,474,000	229,010		
	304	1	15.2915	6.1905	21.4820	2,726,000	272,000	2,454,000	227,240	

단지명	주택형	임대조건			전환가능 보증금 한도액(원)	최대전환시 임대조건		
		임대보증금(원)		월임대료 (원)		보증금 (원)	월임대료 (원)	
		계	계약금			잔금		
정선신동	29A	5,397,000	260,000	5,137,000	116,570	(+) 9,000,000	14,397,000	64,070
	37A	10,038,000	500,000	9,538,000	158,670	(+) 15,000,000	25,038,000	71,170
	37B (주거약자)					(-) 5,000,000	5,038,000	173,250
46		14,464,000	720,000	13,744,000	192,130	(+) 19,000,000	33,464,000	81,290
						(-) 8,000,000	6,464,000	215,460

공급형별 (m <sup>2</sup> )	가군 (생계·의료급여수급자 등)			나군 (일반 등)		
	임대보증금(원)		월 임대료 (원)	임대보증금(원)		월 임대료 (원)
	계	계약금 (계약 시)		잔금 (입주 시)	계	
26	2,409,000	120,000	2,289,000	47,910	12,054,000	602,000
					11,452,000	103,310

# PDF 데이터 추출 과정 및 도구 선택 배경

## 3. LLM 기반 요약 시도

한계점

PDF 공고문 내 명시적인 목차가 존재하지 않음

문서 내 주요 항목(예: 주택단지 개요, 모집대상 등)은 존재하지만

- 숫자 없이 시작하거나
  - 로마자, 특수기호 등 서식이 문서마다 상이함

그 결과, 각 항목의 시작과 끝 범위를 명확히 파악하기 어려운

■ 원주혁신 푸른숲 9단지 A-2BL : 10년 분양전환 공공임대주택 756세대 중 예비입주자 190세대														
단지명	주택형	빌코니 유형	세대별 주택면적(m <sup>2</sup> )					공유 대지 면적 (m <sup>2</sup> )	최고 층수	건설 호수	기준 예비자수	공가 호수	모집 호수	
			공급면적		그 밖의 공동면적		계약 면적 (계)							
푸른숲9 [A-2BL]	51.9500A	학장형	51.95	21.9539	73.9039	4.4564	24.1486	102.5089	49	15	200	16	5	80
	59.8100A	학장형	59.81	25.2755	85.0855	5.1307	27.8023	118.0185	56	20	479	13	16	80
	59.9600B	학장형	59.96	25.3389	85.2989	5.1435	27.8720	118.3144	56	20	77	-	3	30

\* 청약신청은 단지내 주택형별로 구분하여 1세대내 1건에 한해 신청 접수할 수 있으며, 중복신청시 전무 무효처리합니다. (청약신청한 신청은 추후 다른 주택형별로 변경 불가)

\* 기준 예비자수 및 공가호수는 2025.05.15.일 현재 기준이며, 기존 입주자의 임대차계약 해지 및 선 순위 예비입주자의 추가계약 또는 계약포기 등  
인하여 변동 될 수 있습니다.

\* 금회 예비자모집 대상 단지의 기계약자 및 그 세대원은 동일단지·동일유형·동일면적(소수점이하 절사)의 주택형으로 신청불가하여, 신청할  
경우처리하고 계약체결 후라도 확인시 계약해제됨

\* 금회 모집에 공기세대를 초과하여 선정된 예비입주자는 현재 공가 및 향후 임대주택의 혜택이 발생할 경우를 대비하여 모집하는 것으로, 예비입주  
선정되었다 하더라도 예비순위에 따라 실제 입주시기까지는 많은 기간이 소요될 수 있으며 예비입주자의 지원은 임대기간 내에만 그 자격이 유지  
된다. 단, 계약시점에 예비입주권으로서 지정되는 수령권만이다.

\* 금회 공급되는 주택은 전 세대 발코니 확장형으로 시공되며, 팝플렛 등으로 동호배치도, 평면도 등을 확인하신 후 신청하시기 바랍니다.

- \* 주택규모 표시방법은 법정계평단위인 제곱미터(m<sup>2</sup>)로 표기하였습니다.(m<sup>2</sup>를 평으로 환산하는 방법: m<sup>2</sup> × 0.3025 또는 m<sup>2</sup> ÷ 3.3058)
- \* 주거면적은 주거의 유통크기 쌍이거나 면적이나 주거고용면적은 계단, 복도, 현관, 등 고도증대의 치사층에 있는 고용면적이고 그 밖의 고용면적은 주거면적으로 표기합니다.

\* 주거공용면적은 주거의 중요도로는 드물어 전용면적, 주거공용면적은 개인, 혼인 등 경제구조의 시장성이 있는 경제면적이고, 그 뒤로 경제주거공용면적을 제외한 지하층, 관리사무소, 노인정 등의 공용면적입니다.

※ 각 세대별 주거공용면적은 단지전체의 주거공용면적을 세대별 전용면적 비율에 따라 배분한 것으로 계약상 주거공용면적이 해당세대 또는 동의 공용면적과 일치하는 것은 아니며, 동일한 주택형의 경우라도 해당세대 주거공용부분은 동·호별로 형태 및 면적 등이 다소 차이가 있을 수 있습니다.

\* 금회 공급되는 주택은 전세대 발코니 확장형으로 시공되며, 인터넷 공고시 첨부된 험플릿 등으로 동·호배치도, 평면도 등을 확인하신 후 신청하

\* 최고층수는 해당주택형의 최상층 층수이며, 난방방식은 개별난방, 구조는 철근콘크리트벽식 구조, 지붕은 경사지붕으로 시공됩니다.

\* 입주지정기간은 계약체결일로부터 30일 이내입니다.

## 임대기간 및 임대조건

■ 임대기간  
이 주택의 임대기간은 최초 입주일로부터 5년(5년마다 1회) 이내로 하여금 임대료를 조정하는 권리가 있는 경우에 해당된다.

- 이 주택의 임대기간은 최초 입주기간 종료일(17.01.16)이 축하는 월의 다음을 1일부터 10년이며, 임대기간 종료 후 분양전환되는 주택입니다.
- 이 주택의 임대차계약기간은 2년이며, 계속 거주를 희망하는 경우에는 「공공주택 특별법」 등 관계법령에서 정한 입주자요건을 충족하는 분에 한

## 1. 주택단지 개요

단지명	단지위치	건설호수	최초입주일자
대전도안 1단지	대전광역시 유성구 도안대로 560 (봉명동 1024)	14개동 1,668호	'10.09
대전도안 4단지	대전광역시 유성구 상대로 40 (상대동 465)	9개동 1,216호	'10.11
대전도안 10단지	대전광역시 유성구 원신흥로 37 (원신흥동 495)	11개동 1,647호	'10.10
대전노은3지구 1단지	대전광역시 유성구 지죽로 190번길 15 (지죽동 1040)	5개동 600호	'15.08
대전노은3지구 4단지	대전광역시 유성구 지죽로 148번길 40 (지죽동 1051)	7개동 624호	'12.06
대전복산 2단지	대전광역시 유성구 애월로 136번길 75 (복산동 298)	8개동 853호	'09.07

- 주택단지별로 단지여건·주변환경 등이 다르니, 주택단지별 안내사항을 확인하시기 바랍니다.
  - 본 공고문으로 안내하지 못한 단지여건·주변환경 등이 있을 수 있으며, 매연·분진·악취·소음·폐기물 등 주변의 생활환경 오염으로 인한 환경오염에 대한 안내사항은 주택단지별 안내사항을 참조하시기 바랍니다.

2. 모집대상 주

단지명	주택형	세대당 계약면적(m <sup>2</sup> )					구조/난방	건설호수	대기중인 예비자수	모집할 예비자수(570호)
		주거 전용	주거 공용	기타 공용	주차장	합계				
대전도안 1단지	33	33.8200	17.2449	0.8830	11.3321	63.2800	복도식 지역난방	539	1	70
		33.9100	17.2908	0.8854	11.3623	63.4485				
	39	39.9100	20.3502	1.0421	13.3727	74.6750		640	-	100
		39.9400	20.3655	1.0428	13.3827	74.7310				
		39.5200	20.1514	1.0319	13.2420	73.9453				
	46	39.8000	20.2941	1.0392	13.3358	74.4691		260	-	50
		46.6900	23.8074	1.2191	15.6445	87.3610				
		46.8400	23.8839	1.2230	15.6947	87.6416				
대전도안 4단지	41	46.9700	23.9502	1.2264	15.7383	87.8849		362	5	60
		41.5800	17.4483	1.6404	15.0193	75.6880				
		41.1100	17.2511	1.6218	14.8495	74.8324				
		41.3900	17.3686	1.6328	14.9506	75.3420				

2 익대대상 및 익대조거

# PDF 데이터 추출 과정 및 도구 선택 배경



## 4. Upstage API 활용

HTML 구조로 변환하여 추출 한결과 텍스트/위치/계층 정보  
를 잘 보존한 상태로 가져오는것을 확인

```
# Upstage API 활용
loader = UpstageLayoutAnalysisLoader(
    file_path,
    output_type="html",
    split="element",
    use_ocr=True,
    exclude=["header", "footer"]
)
docs = loader.load()
```

```
page_content=
'<br>
<h1 id='1' style='font-size:22px'>
살고 싶은 집과 도시로 국민의 희망을 가꾸는 기업</h1>'
metadata=
{'page': 1, 'id': 1,
'bounding_box':
'[{"x": 288, "y": 130},
 {"x": 985, "y": 130},
 {"x": 985, "y": 173},
 {"x": 288, "y": 173}]',
'category': 'heading1'}
page_content=
'<br>
<h1 id='2' style='font-size:22px'>
행복도시 6-3M4BL 신혼희망타운 통합공공임대주택 입주자 추가모집 공고
<br>
입주자 모집공고일
[2025.03.14]
</h1>'
metadata=
{'page': 1,
'id': 2,
'bounding_box':
'[{"x": 73, "y": 189},
 {"x": 1165, "y": 189},
 {"x": 1165, "y": 270},
 {"x": 73, "y": 270}]',
'category': 'heading1'}
```

## 4. Upstage API 활용

- PDF 내용을 HTML 형태로 파싱하여 LLM에 입력 후 테스트 진행

```
user_prompt = """
다음은 주거 임대 조건을 정리한 HTML 테이블입니다.
이 표의 내용을 이해하고 아래 질문에 답해 주세요.
```

[표]

공급 형별	공급 대상	건설 호수	현재 예비 자수	모집 예비 자수	세대 당 계약면적 (m <sup>2</sup> )	임대조건	전환 가능 보증금 한도액 (천원)	최대전환시 임대조건	구조 및 난방	주거 전용	주거 공용	그 밖의 공용면적	합계	임대보증금 (천원)	월 임대료 (원)
16A	16세 미만 아동이 있는 가족	1~3층	1~3호	1~3호	25~30m <sup>2</sup>	1~3년	50~100	1~3년	1~3층	30~40m <sup>2</sup>	1~2호	5~10m <sup>2</sup>	3~4호	50~100	1~3년
19A	19세 미만 아동이 있는 가족	1~3층	1~3호	1~3호	25~30m <sup>2</sup>	1~3년	50~100	1~3년	1~3층	30~40m <sup>2</sup>	1~2호	5~10m <sup>2</sup>	3~4호	50~100	1~3년
40A, B	40세 미만 아동이 있는 가족	1~3층	1~3호	1~3호	25~30m <sup>2</sup>	1~3년	50~100	1~3년	1~3층	30~40m <sup>2</sup>	1~2호	5~10m <sup>2</sup>	3~4호	50~100	1~3년

-이하 생략-

[질문]

1. 공급형 16A의 "청년(소득 없음)" 대상의 모집 예비 자수는 몇 명인가요?
2. 공급형 19A의 "청년(소득 있음)"과 29형의 "청년(소득 있음)" 중, 세대 당 계약면적(합계) 가 더 큰 것은?
3. 공급형 40A,B "신혼부부"의 경우, 전환 가능한 보증금을 최대로 전환했을 때의 월 임대료는 얼마인가요?
4. 전체 공급형 중 임대보증금(계)이 가장 높은 대상은 누구이며, 그 금액은 얼마인가요?
5. 공급형별 공급대상별 임대보증금 알려주세요.

""

# PDF 데이터 추출 과정 및 도구 선택 배경



## 1. HTML 기반 문서 파싱 후 LLM 응답 비교

qwen3:8b

응답 시간: **71.81초**

5개의 질문중에 2개만 응답

1. 29형의 청년(소득 있음)  
-세대 당 계약면적(합계) : 58.08 m<sup>2</sup> (표에서 "세대 당 계약면적 (m<sup>2</sup>)"의 "합계" 열)  
  
19A의 청년(소득 있음)  
- 데이터가 없습니다. 19A는 신혼부부를 대상으로 하며,  
청년(소득 있음)은 다른 섹션에서만 나타납니다.  
  
결론:  
29형의 청년(소득 있음)이 19A의 청년(소득 있음) 대신 가능한 유일한 항목입니다.  
따라서 29형의 청년(소득 있음)이 58.08 m<sup>2</sup>로, 더 큰 계약면적을 갖습니다.
  
2. 16A의 '청년(소득 없음)'  
-세대 당 계약면적(합계) : 58.08 m<sup>2</sup>  
  
19A의 신혼부부  
-세대 당 계약면적(합계) : 99.3180 m<sup>2</sup>  
  
29형의 청년(소득 있음)  
-세대 당 계약면적(합계) : 58.08 m<sup>2</sup>  
  
31형의 '청년(소득 있음)'  
-세대 당 계약면적(합계) : 58.08 m<sup>2</sup>  
  
- 결론:  
19A의 신혼부부가 99.3180 m<sup>2</sup>로, 가장 큰 계약면적을 제공합니다.  
  
-이하 생략

# PDF 데이터 추출 과정 및 도구 선택 배경



## 2. HTML 기반 문서 파싱 후 LLM 응답 비교

gemma3:12b

응답 시간: 19.22초

대체로 응답 잘함

1. 16A 공급형의 청년(소득 없음) 대상의 모집 예비 자수는 15명입니다. 테이블의 16A 항목을 살펴보면 모집 예비 자수 란에 15라고 명시되어 있습니다.
2. 19A 청년(소득 있음) : 세대 당 계약면적(합계)은  $47.7375 \text{ m}^2$ 입니다. 29형 청년(소득 있음) : 세대 당 계약면적(합계)은  $72.3701 \text{ m}^2$ 입니다. 따라서 29형의 청년(소득 있음)이 19A 청년(소득 있음) 보다 세대 당 계약면적(합계)이 더 큽니다.
3. 40A, B 신혼부부의 경우 전환 가능한 보증금을 최대로 전환했을 때의 월 임대료는 10,220 원입니다. 표에서 전환 가능 보증금 한도액이 (+34,000) 이므로 34000 원에서 전환 수수료를 제외한 값이 월 임대료가 됩니다.
4. 전체 공급형 중 임대보증금(계)이 가장 높은 대상은 36A, B 40A, B 신혼부부이며 그 금액은 590,980 천원입니다.
5. 각 공급형별 공급대상별 임대보증금은 다음과 같습니다:

공급형	공급 대상	임대보증금 (천원)
---   ---   ---		
16A	대학생 청년 (소득 없음) 청년 (소득있음)	33,320, 35,280
19A	대학생 청년 (소득없음)	38,420
19A	청년 (소득있음)	40,680
29	대학생 청년 (소득없음)	57,460
29	청년 (소득있음)	60,840
29	주거급여 수급자	50,700
29	고령자	64,220
신혼부부	신혼부부	82,800
36A, B 40A, B	신혼부부	590,980

# PDF 데이터 추출 과정 및 도구 선택 배경



## 3. HTML 기반 문서 파싱 후 LLM 응답 비교

exaone3.5:7.8b

응답 시간: **14.6초**

속도가 가장 빠르며 대체로 응답 잘함

1. 공급형 16A의 청년(소득 없음) 대상의 모집 예비 자수는 15명입니다.
2. 공급형 29형의 청년(소득 있음)의 세대 당 계약면적 합이 더 큽니다.  
29형의 세대 당 계약면적 합: 50,700 m<sup>2</sup>  
19A의 청년(소득 있음) 세대 당 계약면적 합: 36.48 m<sup>2</sup>
3. 공급형 40A,B 신혼부부"에서 전환 가능한 보증금을 최대로 전환했을 때의 월 임대료는 128,600원입니다.  
이는 (-) 77,000 전환 후의 값으로 해석됩니다.
4. 임대보증금(계)이 가장 높은 대상은 주거급여 수급자로,  
그 금액은 325,300 천원입니다.
5. 각 공급형 및 공급대상별 임대보증금은 다음과 같습니다
  - 16A  
- 청년 (소득 없음): 33,320 천원
  - 청년 (소득 있음): 5,320 + (-) 28,000 = 214,940 천원
  - 대학생 청년: 33,320 천원
  - 대학생 청년 (소득 없음): 33,516 천원
  - 청년 29 (소득 있음): 56,680 천원
  - 청년 29 (소득 없음): 60,840 천원
  - 주거급여 수급자: 48,165 천원
  - 고령자: 61,009 천원
  - 이하 생략-

# Chunk 전략

## 문맥 보전을 위한 단위 청크 구성 설계

PDF 내 테이블이나 문단의 흐름을 유지하기 위해 HTML element 단위로 청크를 구성하며,  
LLM 문맥 해석 한계를 고려해 청크 길이는 **2000자**로 제한

청크 단위	doc[i].page_content(HTML element 단위)
문자 수 제한	청크는 최대 2000자 이하로 구성
결합 조건	다음 요소를 추가해도 2000자 이하면 결합
예외 처리	하나의 요소가 2000자 초과 시 단독 청크로 처리

```
# 예외 처리: 하나의 요소가 2000자 초과인 경우 → 단독 청크로 처리
if content_len > max_chars:
    flush_chunk()
    chunks.append({
        "content": content,
        "category": [category],
        "page": [page],
        "id": [element_id]
    })
    continue

# 결합 조건: 현재 청크에 넣어도 2000자 이하일 경우만 추가
if current_len + content_len + 1 <= max_chars:
    if current_chunk["content"]:
        current_chunk["content"] += "\n" + content
    else:
        current_chunk["content"] = content

    current_chunk["category"].append(category)
    current_chunk["page"].append(page)
    current_chunk["id"].append(element_id)
    current_len += content_len
else:
    # 초과 시 현재 청크 마감하고 새 청크 시작
    flush_chunk()
    current_chunk["content"] = content
    current_chunk["category"] = [category]
    current_chunk["page"] = [page]
    current_chunk["id"] = [element_id]
    current_len = content_len
```

# RAG

---

## 1. 임베딩 모델

- **bge-m3**
- <table>, <section>, <ul> 태그들을 구문적으로 무시하지 않고 의미 덩어리로 파악 → chunk 간 검색 정확도 향상
- 한국어 자연어 쿼리에 강력한 의미 매칭 성능 보장

## 2. 벡터DB

- ChromaDB
- LangChain과의 완벽한 통합
- 설치 및 사용이 매우 간편함

## 3. LLM 모델

- **exaone3.5:7.8b**
- 벡터 DB에서 검색한 결과는 메타정보 chunk\_id, page, element\_id 기준으로 정렬하여 LLM에게 전달
- 테스트 결과 가장 좋은 성능을 보임

# 파인튜닝

## 1. 프로젝트 개요

- 모델명: naver-hyperclovax/HyperCLOVAX-SEED-Text-Instruct-1.5B
- 선정 이유:
- QnA 용도로 적합한 구조 (Instruction 기반 대화 최적화)
- 1.5B 파라미터로 메모리/응답속도 대비 성능 우수
- Naver 내부에서도 한국어 QnA 베이스모델로 채택된 사례 있음
- 실제 서비스에 탑재 가능한 경량화된 구조로 LoRA 적용 용이
- 목표: 복잡한 LH 청약 제도와 임대 정책을 쉽게 설명해주는 한국어 QA 챗봇 구축
- 기술: LoRA 기반 PEFT 파인튜닝 적용

## 2. 파인튜닝 데이터셋 구성

- 출처: 실제 LH 공고문 기반 사용자 질문 시나리오 구성, 마이홈 용어 사전
- 도메인 분류:
- 통합공공임대 / 국민임대 / 행복주택 / 공공임대 5년·10년형 / 장기전세 / 매입임대 / 영구임대 등 용어 기반 설명
- 도메인 특화형 QA
- 다양한 질문 표현(instruction)
- 통일된 정책 설명(output)
- 총 1,363개

"**instruction**": "통합공공임대가 뭐야?",

"**output**": "통합공공임대는 저소득층 및 취약계층을 위한 공공임대주택입니다. 시세의 35~90% 수준의 임대료로 공급됩니다."

## 데이터셋

"**instruction**": "WELHOME은 어떤 서비스인가요?"

"**output**": "WELHOME은 개인 정보 기반으로 맞춤형 청약 정보를 추천하고, 관련 문서를 찾아주는 청약 안내 플랫폼입니다."

# 파인튜닝

## 3. 모델 파인튜닝 설정

**토크나이저:** Chat Template 기반 Prompt 구성

**모델 구조:**

$r=8 \rightarrow$  파라미터 효율성과 표현력 간 균형. 실무 검증에서 가장 안정적인 기본 값 중 하나로 평가됨

**lora\_alpha=16** → 상대적으로 작은 r에 비례해 scaling factor 설정 (일반적으로 alpha =  $2 \times r$ 이 기준)

**dropout=0.05** → **과적합** 방지 목적, 적은 데이터셋에 적합한 경량 설정

**학습 설정:**

TrainingArguments(

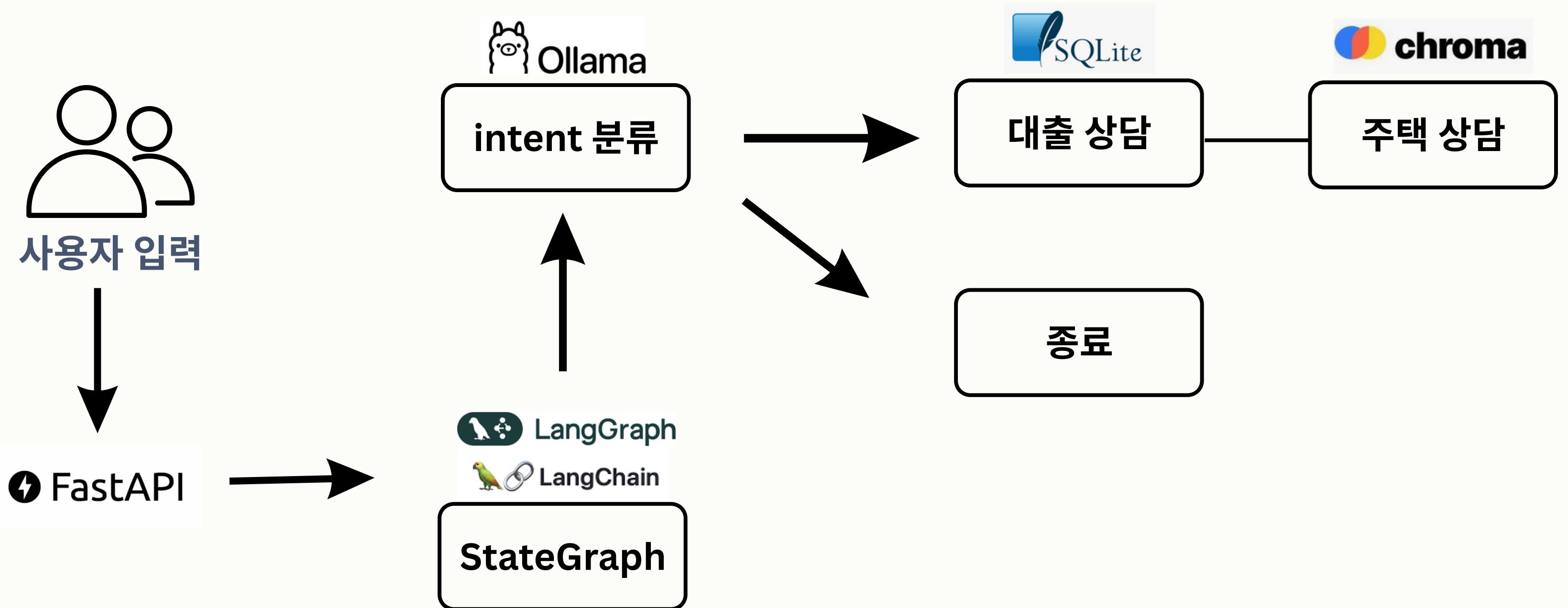
`per_device_train_batch_size=2,`  
`gradient_accumulation_steps=4,`  
`num_train_epochs=30,`  
`learning_rate=5e-5,`  
`bf16=True`

)

질문	30Epoch	50Epoch
WELHOME은 어떤 서비스인가요?	WELHOME은 주택 관련 정보를 실시간으로 제공하는 챗봇 기반 플랫폼입니다.	WELHOME은 청약, 임대, 공공지원 주거 정보 통합 플랫폼입니다. 여러 공고를 한 번에 조회할 수 있습니다.
신혼부부도 지원 가능한가요?	네, 신혼부부 특별공급은 나이와 소득 기준을 달리 적용합니다.	질문 내용에 따라 신혼부부 전용 주거지원 항목 여부가 답변됩니다.
"공고 이미지도 제공되나요?	조건에 맞는 공고만 이미지 포함 형식으로 제공합니다.	일부 공공임대복달은 이미지와 상세 설명과 함께 첨부됩니다.
임대조건은 어떤 기준인가요?	임대 대상자의 소득 수준이나 가족 구성에 따라 설정됩니다.	국저왕 주택, 소득수준이 낮은 가구를 위한 공공임대를 기준으로 설정합니다.

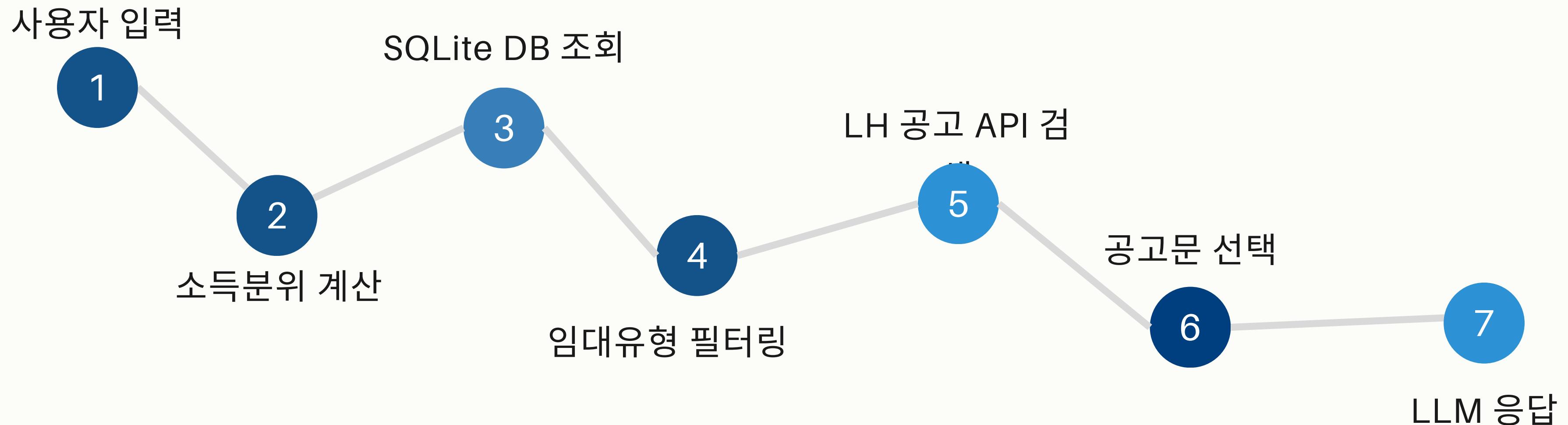
30Epoch	30 Epoch에서 더 안정적이고 예측 가능한 응답 생성
50Epoch	50 Epoch에서 국저왕, 공공임대복달 등 비정상 토큰 생성 확인
30 Epoch는 응답의 안정성과 신뢰도를 보였습니다. 따라서, 30 Epoch 모델을 선정하게 되었습니다.	

# 에이전트



# 주택 추천 흐름

---





## SQLite DB 구조

### 1. program

임대주택 유형에 대한 신청 자격 기준 정보 저장

### 2. income\_rule

임대주택 유형별 소득 제한 규칙을 저장

예: 국민임대의 경우 소득 기준이 70%이하

### 3. income\_reference

가구원수별 기준 소득을 저장하는 테이블

### 4. asset\_rule

자산 및 차량가액 한도를 관리하는 테이블

### 5. bonus\_rule

소득 기준을 완화 해주는

추가 가점 규칙을 관리하는 테이블

### 6. region\_code

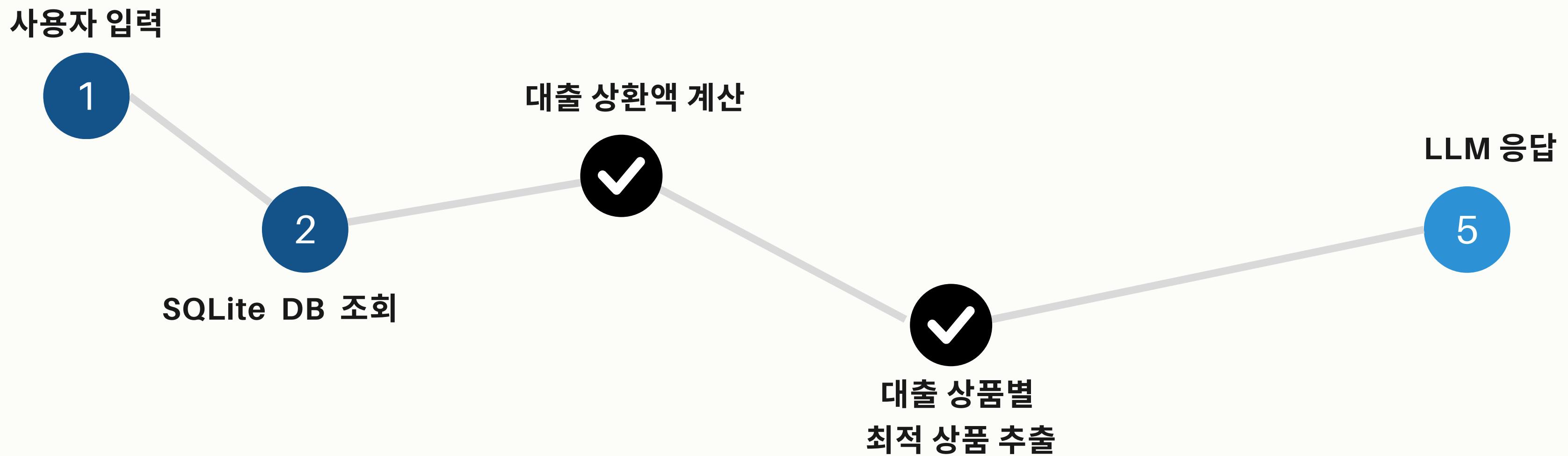
지역 코드와 지역명을 매핑하는 테이블

lh api 호출 시 지역 코드로 특정 지역 공고 검색

사용자의 입력 조건을 정밀하게 검증하고, LH API에서 받아온 공고문 조건과 정확히 매  
칭시켜 신뢰성 있는 공고 추천이 가능하도록 합니다.

# 대출 추천 흐름

---



# SQLite DB 구조



전국은행연합회 소비자포털에서 은행별  
금리방식, 상환방식, 당월 최저금리, 당월  
최고금리, 전월취급 평균금리, 월평균 상  
환액, 총 대출비용

각 은행별 상품별 상환 방식에 따라  
총 상환액 계산

```
def calculate_cost_total(row, amount, years):
    rate = row["rate_avg_prev"]
    repay_type = row["repay_type"]
    r = rate / 100

    if repay_type == "만기일시상환":
        total = amount + (amount * r * years)
        return round(total)

    elif repay_type == "원리금분할상환":
        monthly_rate = r / 12
        n = years * 12
        if monthly_rate == 0:
            monthly_payment = amount / n
        else:
            monthly_payment = amount * monthly_rate * (1 + monthly_rate) ** n / ((1 + monthly_rate) ** n - 1)
        total = monthly_payment * n
        return round(total)

    elif repay_type == "원금분할상환":
        n = years * 12
        monthly_principal = amount / n
        total = 0
        for i in range(1, n + 1):
            remaining_principal = amount - monthly_principal * (i - 1)
            interest = remaining_principal * (r / 12)
            total += monthly_principal + interest
        return round(total)
    else:
```

# 대출 상담

Pandas로 각 은행별 상환액이 낮은 상품 2개

씩 추출

마크다운 표를 LLM에 전달

```
def get_table_text(loan_amount, loan_years, db_path):
    conn = sqlite3.connect(db_path)
    query = """
        SELECT
            bank,
            product,
            repay_type,
            rate_avg_prev,
            limit_amt
        FROM loan_products
    """
    df = pd.read_sql_query(query, conn)
    conn.close()

    df["cost_total"] = df.apply(
        lambda row: calculate_cost_total(row, loan_amount, loan_years),
        axis=1
    )

    df_filtered = df[df["limit_amt"] >= loan_amount]
    df_sorted = df_filtered.sort_values(["bank", "cost_total"])
    best_two = df_sorted.groupby("bank").head(2).reset_index(drop=True)

    if best_two.empty:
        return None

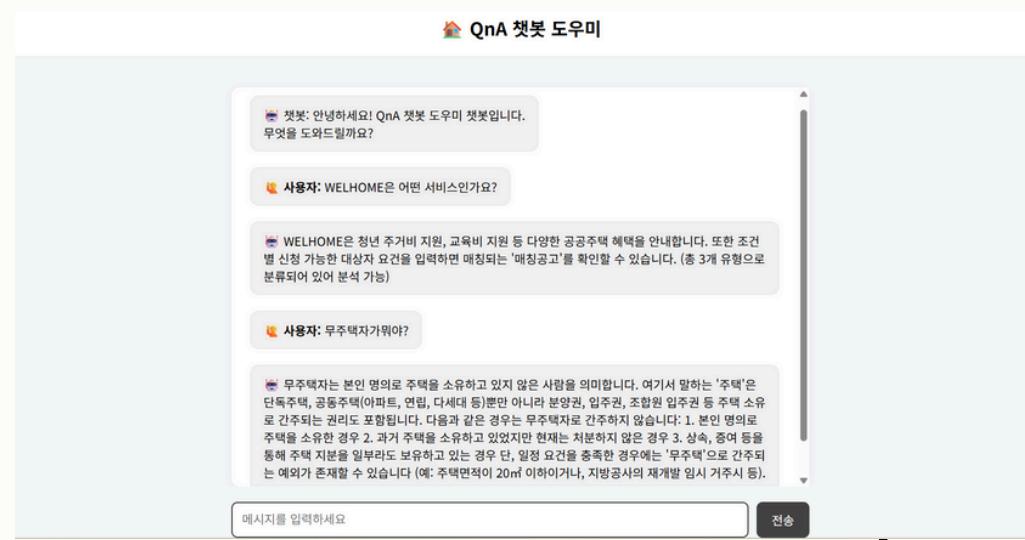
    return best_two.to_markdown(index=False)
```

## 4. 프로젝트 수행 경과



# 페이지 구성

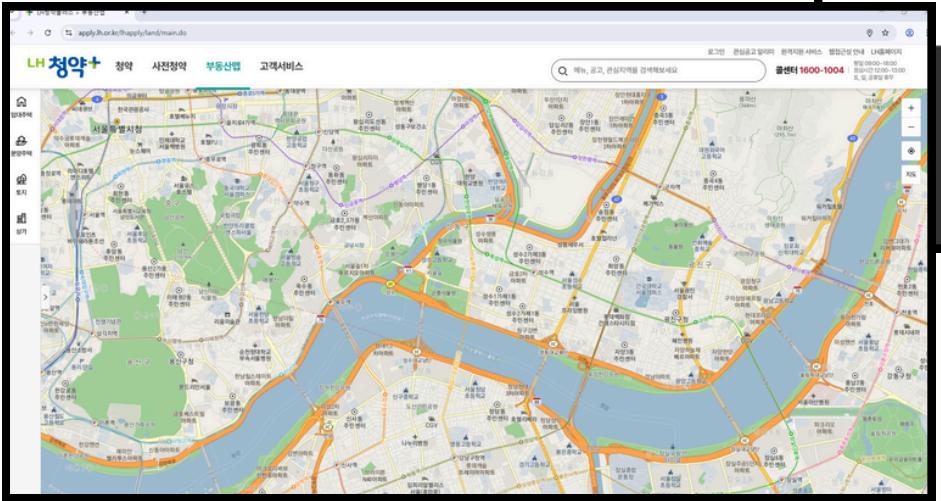
## QnA 챗봇 연결



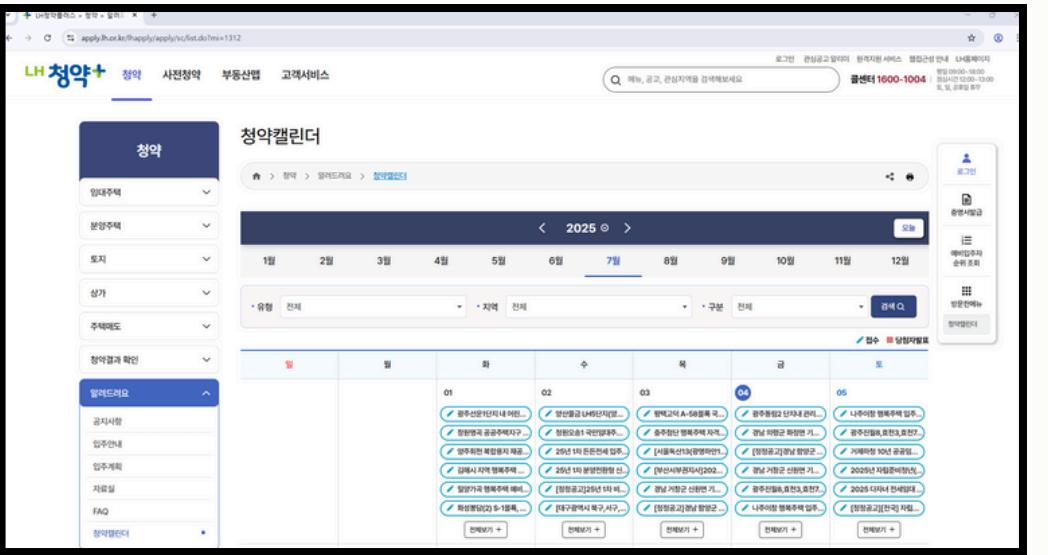
## 본 프로젝트 메인페이지

The screenshot shows the main page for 'Well Home'. It features a sidebar with navigation links like '공지사항', '공고달력보기' (highlighted with a red dashed box), '용역사전', and '고객센터'. Below the sidebar, there are three main content boxes: '자주 묻는 질문' (Frequently Asked Questions), '주택/대출 상담' (Housing/Loan Consultation) which includes a house icon, and 'FAQ' which includes a Q&A icon. A large red dashed box surrounds the '주택/대출 상담' and 'FAQ' sections.

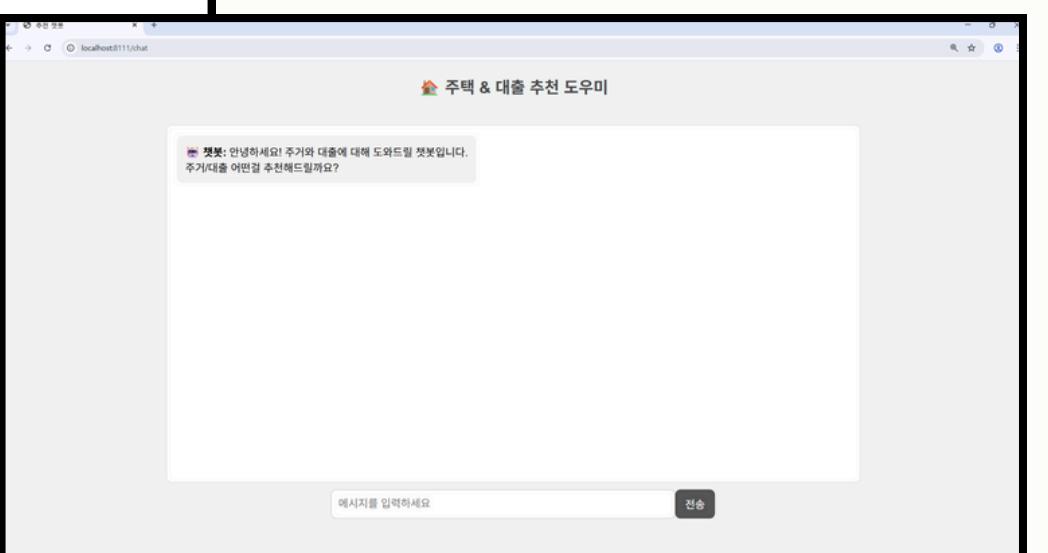
## LH 지도뷰 연결



## LH 캘린더 연결



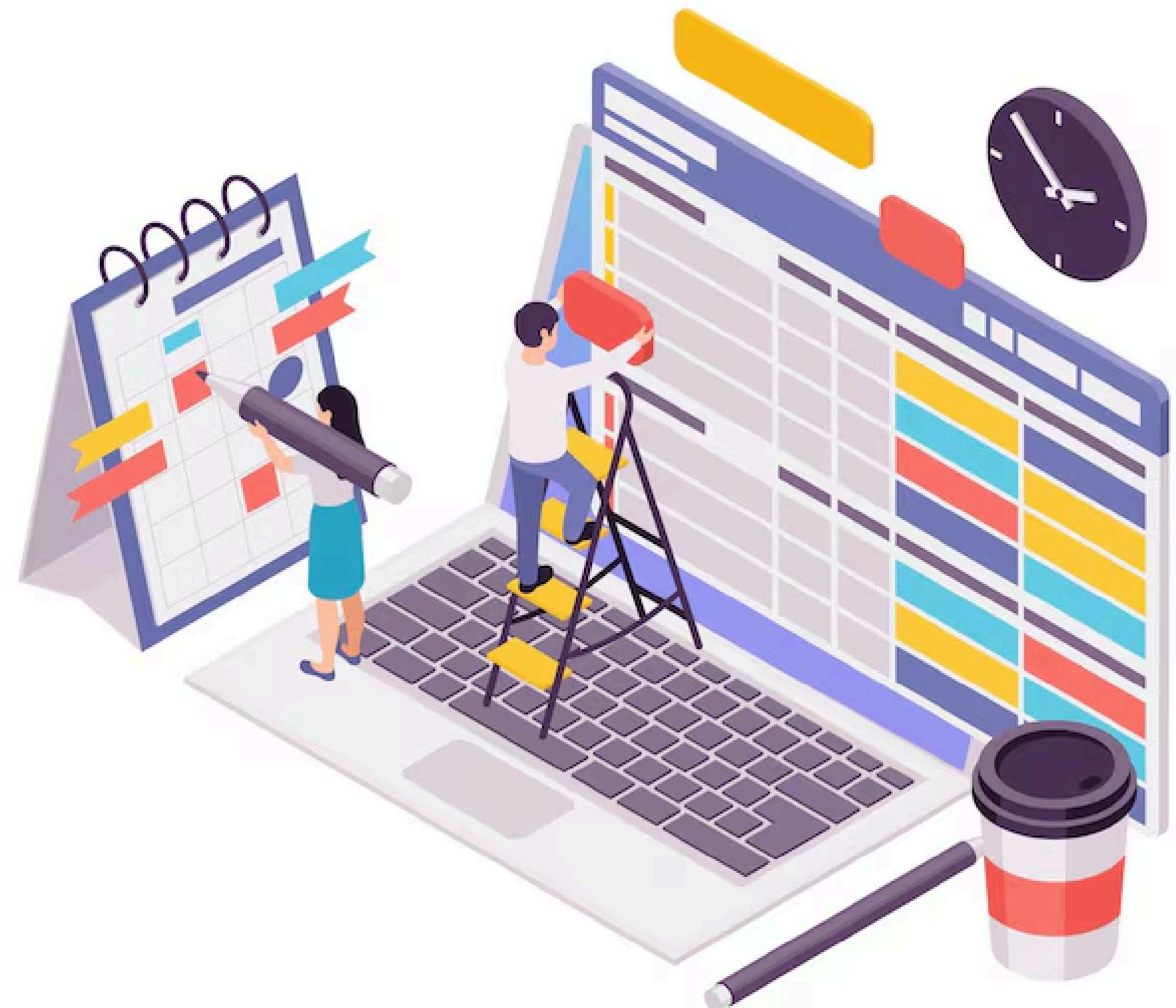
## 추천 챗봇 연결

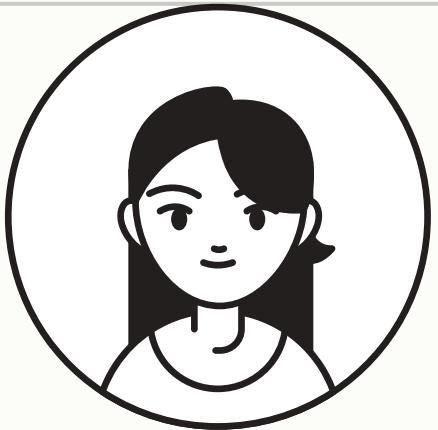


# 프로젝트 수행 경과



## 5. 자체 평가 의견





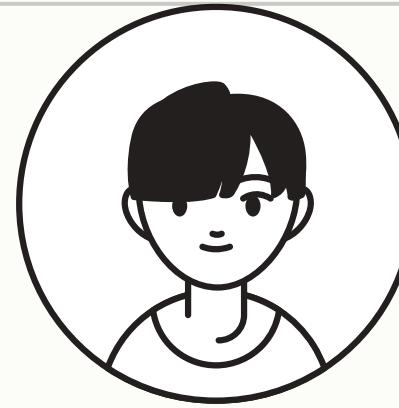
이유진

프로젝트를 진행하면서 가장 어려웠던 부분은 대화형 서비스의 전체 구조를 설계하는 과정이었습니다.

단순히 코드를 작성하는 것과 달리,  
전체적인 서비스 흐름과 상태 관리,  
워크플로우 설계까지

직접 고민하고 구현하는 일이  
생각보다 훨씬 복잡하고 까다롭게 느껴졌습니다.

이번 경험을 통해 설계 단계의 중요성을 실감했고,  
앞으로는 더 체계적으로 구조를 잡을 수 있도록 준비해야  
겠다는 다짐을 하게 되었습니다.



이창민

이번 프로젝트의 가장 큰 도전은 대화형 시스템의 아키텍처를 처음부터 설계하는 과정이었습니다.  
서비스 흐름, 상태 관리, 응답 전략을 유기적으로 연결하는 작업은 단순 구현과는 차원이 다른 복잡성을 요구했습니다.

시행착오를 통해 설계 중심의 접근이 시스템의 일관성과 안정성을 좌우한다는 사실을 체감했고,  
이후에는 흐름도와 결정 기준을 명확히 정리한 후  
개발에 착수하는 방식으로 개선해 나갔습니다.



감사합니다.

---