

가람_진행상황

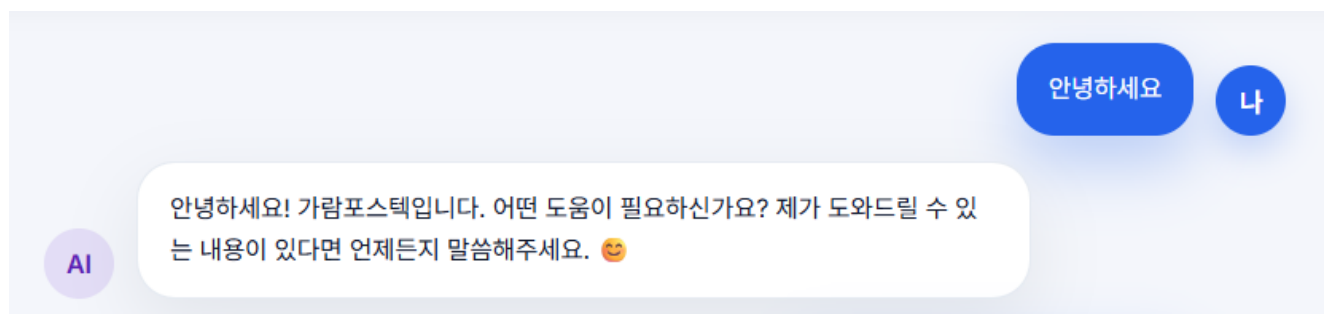
LLM RAG 테스트 과정

사용 사례

QA chain

1. 간단한 질문

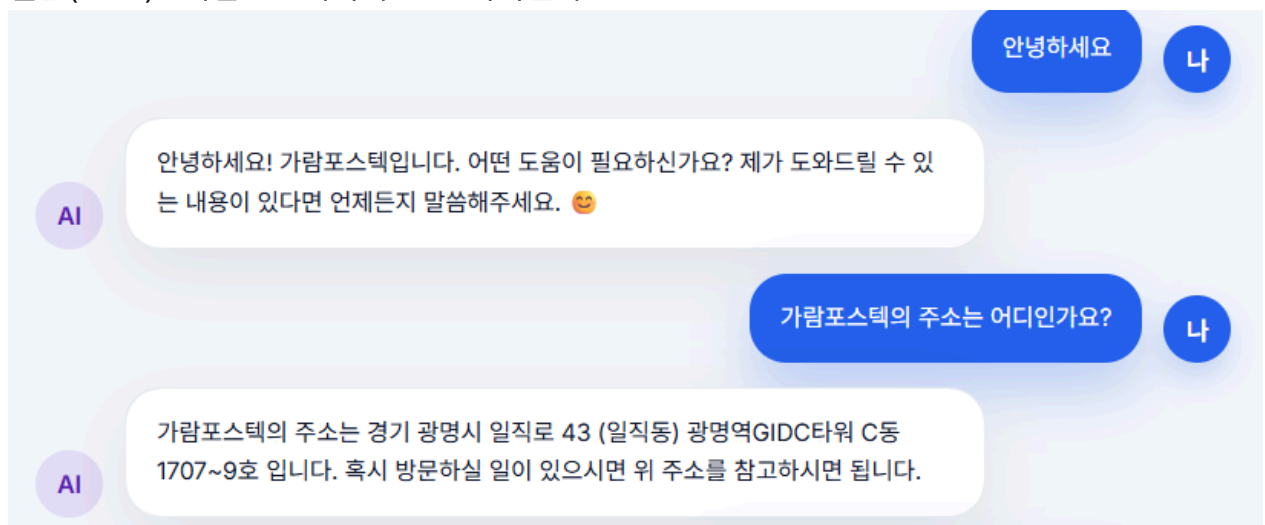
- 질문 : user : "안녕하세요"



- 답변 : output : "안녕하세요 가람 포스텍입니다. 어떤 도움이 필요..."

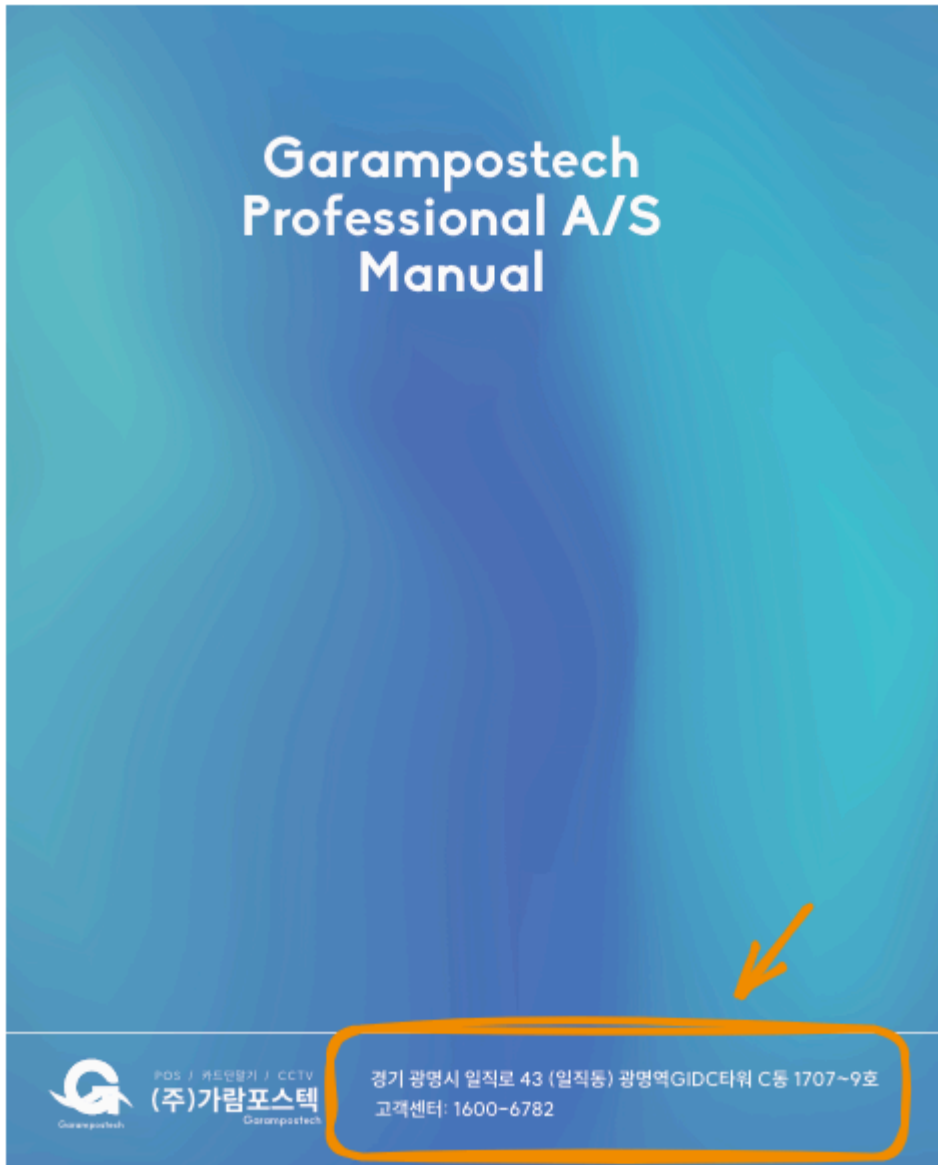
2. pdf 내 간단 질문 : 주소

- 질문(user) : 가람포스텍의 주소는 어디인가요?



- 답변 : output : "가람포스텍의 주소는 경기 광명시 일직로 43(일직동)..."

참조



- pdf 내의 위의 위치를 참조하여 답변합니다.

3. pdf 내 일반 질문 : 고객 응대원칙

- 질문(user) :



- 답변 : output : " 고객 응대 화법의 3대 원칙은 1. 공감이가 먼저다 2. 문장은 끝까지 말한다 ... 3. 말투는 단정하게 유지한다."

2-1. 고객 응대 화법의 3대 원칙

1. 공감이 먼저다

- 고객의 감정을 먼저 수용하는 말이 곧 서비스의 시작입니다.
- 기술적 설명 보다 감정 공감이 선행되어야 합니다.

예 : "말씀 주셔서 감사합니다."

예 : "많이 답답하셨을 것 같습니다."

2. 문장은 끝까지 말한다

- '드렸고요, 했고요, 됐고요'와 같은 비완결 표현은 무책임하거나 반말처럼 느껴질 수 있습니다.
- 모든 문장은 끝맺는 말로 마무리합니다.

예 : "안내드렸습니다.", "처리 완료되었습니다."

3. 말투는 단정하게 유지한다

- 다짜체(습니다형)를 사용하여 신뢰감 있는 인상을 줍니다.
- '되셨어요?', '보셨죠?' 같은 요조체와 '쫌만요', '아까' 등의 구어체는 사용을 지양합니다.

예 : "되셨습니까?", "확인하셨습니까?"

2-2. 부적절한 말투 vs 바람직한 표현

1. 무례·직설 > 공감형 완곡 표현

부적절 표현	바람직한 표현	설명
"그건 안 돼요"	"안타깝지만, 해당 방식은 지원이 어렵습니다"	단호함 대신 유감 표현
"몰라요"	"확인 후 다시 안내드리겠습니다"	무지 표현 금지
"고객님이 잘못하셨어요"	"설정 중 누락이 있었던 것 같습니다"	비난 대신 분석

지식베이스 전체 설정

- 지식베이스를 전체로 설정한 경우 업로드 한 파일을 전부 검토하여 해당되는 내용을 검색하여 답변합니다.
- 지식베이스 "전체" 설정

4. pdf 전체 설정 시 다른.pdf 내용 검색

- 질문(user) : "정부의 AI산업 투자금액은 어느 정도 인가요?"

+ 새 대화

AI 상담 챗봇

세션 #56

API 연결 성공

메뉴얼 상담

API http://localhost:5002

연결 확인

가람포스텍의 주소는 어디인가요?

가람포스텍의 주소는 경기 광명시 일직로 43 (일직동) 광명역GIDC타워 C동 1707~9호 입니다. 혹시 방문하실 일이 있으시면 위 주소를 참고하시면 됩니다.

가람포스텍의 3대 응대 원칙이 뭔가요?

가람포스텍의 3대 응대 원칙은 '문제보다 감정을 먼저 다룬다', '기술보다 신뢰를 먼저 전달한다', '속도보다 정확하고 따뜻한 말로 안내한다' 입니다. 이 원칙들을 따르면 고객과의 상호작용에서 더 나은 경험을 제공할 수 있습니다.

그럼 고객 응대 화법의 3대원칙은 뭐죠?

고객 응대 화법의 3대 원칙은 다음과 같아요:

1. 공감하기 먼저다: 고객의 감정을 먼저 이해하고 수용하는 것이 서비스의 시작입니다.
2. 문장은 끝까지 말한다: 모든 문장은 완결된 말로 마무리하여 무책임한 느낌을 피해야 합니다.
3. 말투는 단정하게 유지한다: 다짜체(습니다형)를 사용하여 신뢰감을 주는 것이 중요해요.

정부의 AI 산업 투자금액은 어느정도인가요?

정부의 AI 산업 투자금액은 100조원입니다. 이는 국민, 기업, 국민연금이 참여하는 50조원 첨단산업 국민펀드와 정부 예산 및 민간 투자로 구성되어 있습니다. 이 투자를 통해 AI 데이터센터 및 데이터집적 클러스터를 구축하고 AI 전용 NPU 개발과 실증을 지원하여 기술 주권을 강화할 예정입니다.

전체

#6 · 기본_고객응대 매뉴얼.pdf

#5 · 고객응대관련법규.pdf

#4 · 고객응대근로자.pdf

#2 · 정부AI정책방향.pdf

#1 · 가람포스텍전문AS매뉴얼_20250714.pdf

메시지를 입력하고 Enter 키로 전송

검색 Top-K 5

지식베이스 #2 · 정부AI정책방향.pdf

음성

전송

응답 언어 한국어

- 답변(output) : "정부의 AI 산업 투자금액은 100조원 입니다. 이는 국민,기업 ..."
 ◦ 정부의 AI정책방향.pdf 에서 찾아 답변을 제공합니다.

미래산업: AI 산업 100조 투자 시대 개막

AI 산업에 100조원 투자

- 이재명 대통령은 2025년 대선을 앞두고 "AI(인공지능) 투자 100조원 시대를 열겠다"는 대규모 공약을 발표
- AI 핵심 자산인 GPU를 최소 5만 개 이상 확보하고, AI 전용 NPU 개발과 실증을 적극 지원해 기술 주권 확보에 역량을 쏟을 전망
- '모두의 AI' 프로젝트 : 한국형 ChatGPT를 개발해 전 국민이 무료로 사용할 수 있도록 하여, 데이터 축적과 AI 접근성 확대 동시 추진

AI 산업 100조원 투자 관련 세부 내용 및 실행 방식

구분	세부 내용 및 목표	실행 방식 및 특징
투자 규모 & 재원	<ul style="list-style-type: none">정부-민간 합동 100조원 투자국민·기업·국민연금이 참여하는 50조원 첨단산업 국민펀드 조성나머지는 정부 예산 및 민간 투자 유치	<ul style="list-style-type: none">국민펀드 투자자 세제 혜택 제공정책 펀드는 국내 전략산업 주식 채권 등에 집중 투자글로벌 AI 기업·투자자 유치
연계 정책	<ul style="list-style-type: none">고성능 GPU 5만개 이상 확보전국 단위 AI 데이터센터 및 데이터 집적 클러스터 구축	<ul style="list-style-type: none">정부가 직접 인프라 구축 주도민간 클라우드·데이터센터 연계

주요 업체별 LLM 모델 및 사용 GPU 개수와 대한민국 GPU 투자 비교

회사/모델명	사용 GPU 종류	투입 GPU 개수	비고
OpenAI GPT-4o	NVIDIA H100	약 15,000개	GPT-4o 기준
xAI Grok-3	NVIDIA H100	약 200,000개	20만개, 세계 최대 규모
Meta Llama 4	NVIDIA H100	100,000개 이상	10만개 이상 클러스터 구축 중

5. pdf 전체 설정 시 유사한.pdf 내용 검색

- 질문(user): "전화벨이 얼마나 울리기 전에 받아야 할까요?"

전화벨이 얼마나 울리기 전에 받아야 할까요?

나

전화벨이 울리기 전에 받으려면 최대한 빨리 받는 것이 좋습니다. 일반적으로 3회에서 4회 울리기 전에 받는 것이 일반적이며, 이렇게 하면 상대방과의 연결이 빠르고 원활해질 수 있습니다.

AI

- 답변(output) : (전화벨이 울리기 3회 이상 울리기 전) 관련 내용

고객 응대 매뉴얼

■ 전화 응대

1. 첫 인사 및 용건 확인

- 전화벨이 3회 이상 울리기 전에 응답. (늦게 받았다면 '늦게 받아 죄송합니다.')
- (인사) 감사합니다. 『 상호 』입니다.

2. 통화중 응대

- 고객의 말을 끝까지 경청한 후 답변한다. (중요한 내용은 메모한다.)
- 단답형 답변보다 고객을 배려하는 표현을 사용한다.

잘못된 표현	고객을 배려하는 표현
잠깐만요.	잠시만 기다려주시겠습니까?
모르겠네요.	제가 알아보고 말씀드려도 괜찮을까요?
없는데요.	부재중이십니다. 메모 남겨드릴까요?
예? 잘 안 들리는데요? 뭐라고요?	죄송하지만 다시 한 번 말씀해 주시겠습니까?

3. 통화 종료 시 응대

- 마케팅 참고용 질문 : 저희 『 상호 』은 어떻게 알고 전화 주셨나요?
- 추가 요청 사항 문의 : 더 필요한 것은 없으신가요?
- 끝인사 : 감사합니다. 좋은 하루 되세요. (먼저 끊지 않도록 주의)

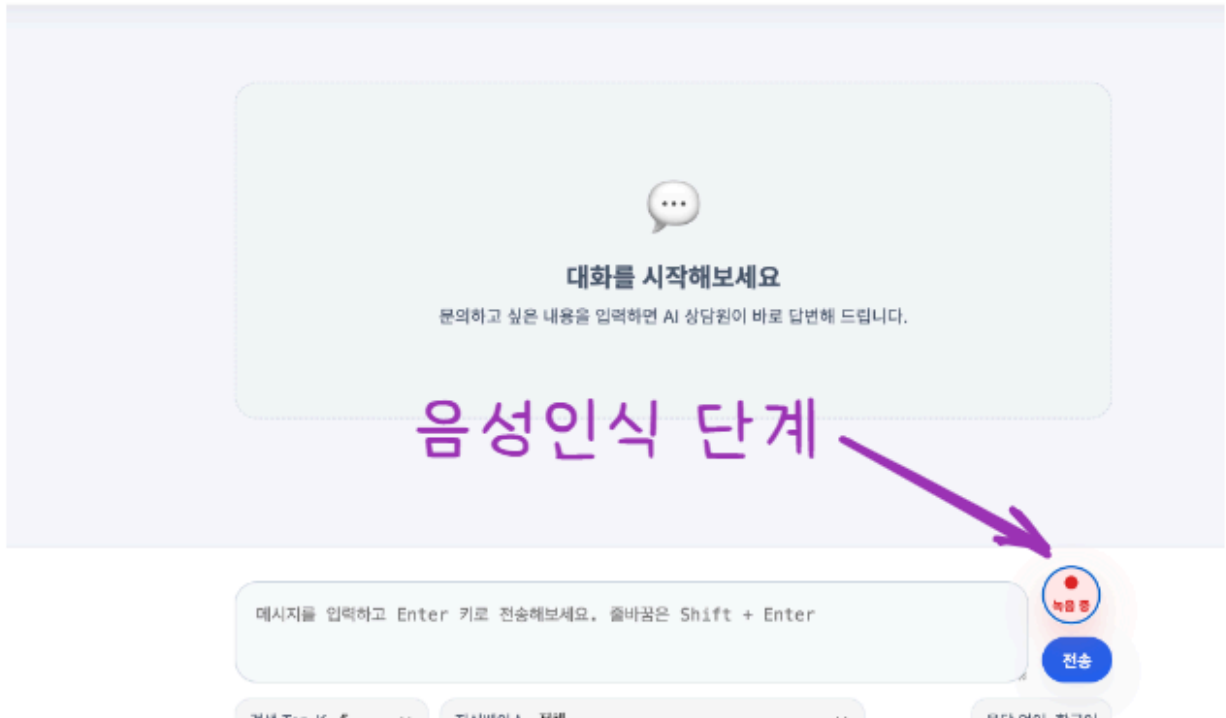
새로운 PDF 적용시 전체 지식베이스에서 알맞은 내용을 Retriever(검색)해서 답변합니다.

음성인식 사용(한국어)

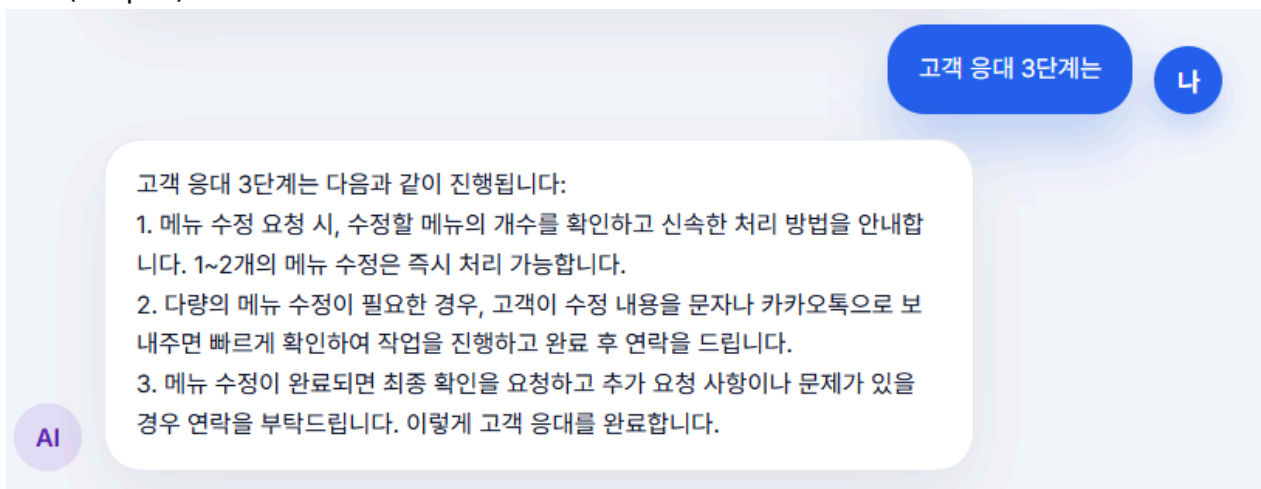
- 마이크 사용



- 말하면 "음성인식" 단계로 전환(Web-speech 사용)
- 질문(user) : (음성으로) "고객 응대 3단계는?"



- 답변(output) : 고객 응대 3단계는 다음과 같이...



데이터 처리 과정

각 대화들은 DATABASE에 서 관리 됩니다

- message 테이블 중 일부

role	character	varying	content	text	response_latency_ms	vector_memory	integer	extra_d	created_at
user			안녕			[null]			
assistant			안녕하세요! 궁금한 점 있으신가요? 부탁하시면 최대한 도와드리겠습니다. 🍷		3740	[0.013785294,-0.010510956,0.008625731,-0.016656127,-0.0...			2025-10-16 0...
user			안녕하세요			[null]			
assistant			안녕하세요! 가람포스텍입니다. 어떤 도움이 필요하신가요? 제가 도와드릴 수 있는 내용이 있다면 언제든지 말씀해...		4301	[0.013639058,-0.009446357,-0.005943919,-0.022944806,-0.0...			2025-10-16 0...
user			가람포스텍의 주소는 어디인가요?			[null]			
assistant			가람포스텍의 주소는 경기 광명시 일직로 43 (일직동) 광명역GiDC타워 C동 1707~9호 입니다. 혹시 방문하실 일이...		4723	[0.0006492223,-0.0064711496,0.007210339,-0.036700107,-0.0...			2025-10-16 0...
user			가람포스텍의 3대 응대 원칙이 뭐예요?			[null]			
assistant			가람포스텍의 3대 응대 원칙은 '문제보다 감정을 먼저 다룬다', '기술보다 신뢰를 먼저 전달한다', '속도보다 정확하고 ...		3865	[0.015758099,-0.0050120656,0.018975722,-0.023087127,-0.0...			2025-10-16 0...
user			그럼 고객 응대 화법의 3대원칙은 뭐죠?			[null]			
assistant			고객 응대 화법의 3대 원칙은 다음과 같아요:		3519	[0.015889779,-0.011325978,0.007944889,-0.017760737,-0.0...			2025-10-16 0...
user			정부의 AI 산업 투자금액은 어느정도인가요?			[null]			
assistant			정부의 AI 산업 투자금액은 100조원입니다. 이는 국민, 기업, 국민연금이 참여하는 50조원 첨단산업 국민펀드와 정...		3372	[null]			2025-10-16 0...
user			전화 응대시 주의사항이 있을까요?			[null]			
assistant			전화 응대시 주의사항은 다음과 같습니다: 1. **친절하고 명확한 응대** ** 고객의 문제를 이해하고 친절하게 안내해야...		6314	[0.002950744,-0.018245133,0.024326844,-0.040763192,-0.0...			2025-10-16 0...
user			전화벨이 얼마나 울리기 전에 받아야 할까요?			[null]			
assistant			전화벨이 울리기 전에 받으려면 최대한 빨리 받는 것이 좋습니다. 일반적으로 3회에서 4회 울리기 전에 받는 것이 일...		3198	[0.013042851,-0.026313232,0.01742616,-0.028213782,-0.00...			2025-10-16 0...

남은 과정

음성인식 고도화

RAG 시스템 고도화

DB 관련 작업 마무리