

하이퍼클로바 X와 RAG 기법을 활용한 개인 맞춤형 금융 비서

누리봄 | 강하람 김기남

01

주제 선정 배경

02

알고리즘 흐름도

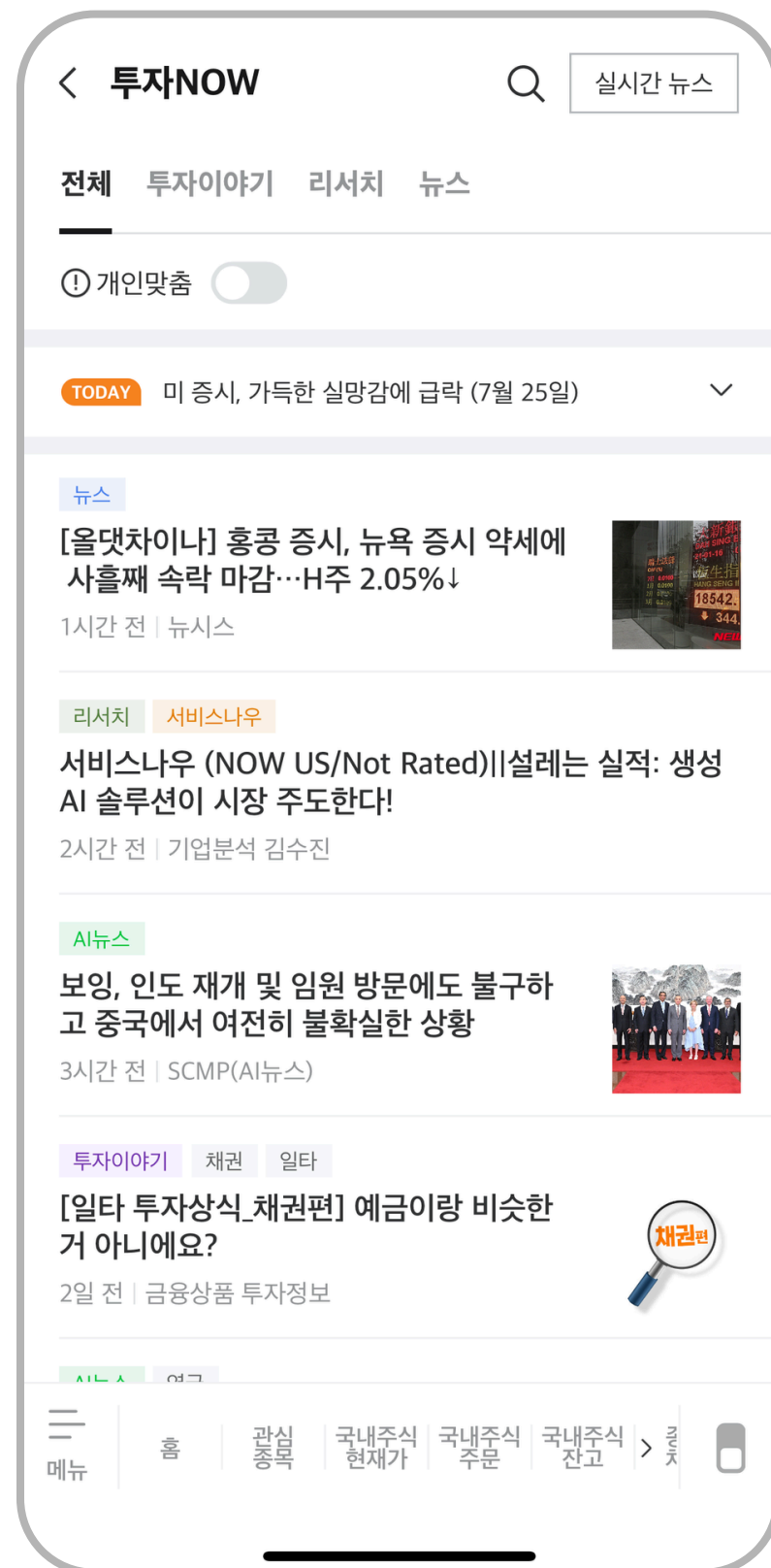
03

학습 과정

04

사용 예시

기존 서비스의 한계점으로 인한 필요



개인 맞춤형 조언의 필요성

현대 금융 환경에서는 수많은 일반적인 금융 콘텐츠가 제공되고 있지만, 이는 **대부분 단편적이고 일반화된 정보에 불과하다**. 개인의 소득, 재산, 포트폴리오 구성, 가족 관계 등 **다양한 요소를 종합적으로 고려한 맞춤형 조언이나 컨설팅 서비스는 매우 부족한 실정이다**. 이는 개인 투자자들이 자신에게 맞는 금융 전략을 세우기 어렵게 만든다. 따라서, 각 개인의 상황에 맞춘 **종합적인 금융 조언 서비스의 필요성**이 대두되고 있다.

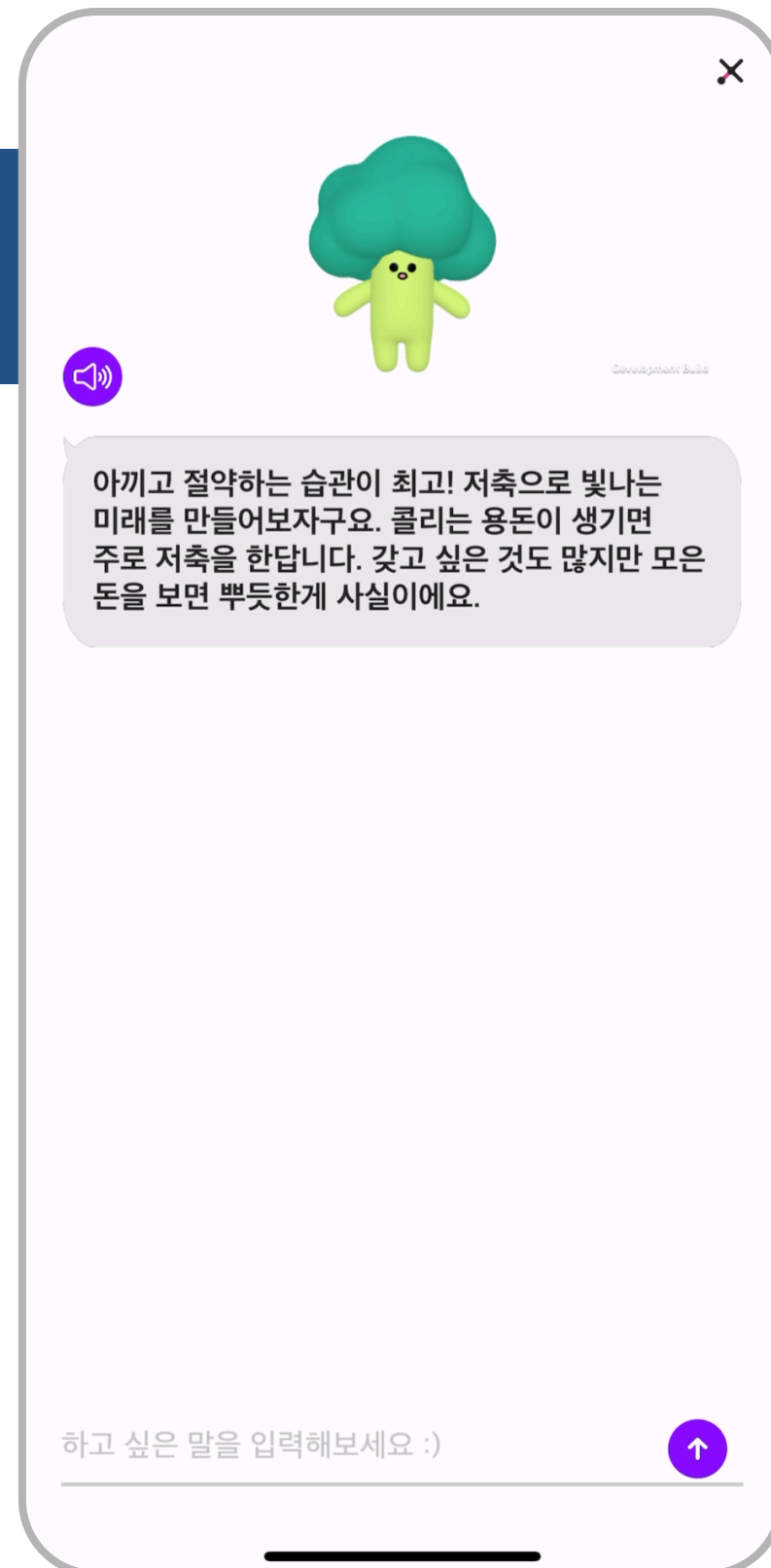
손쉬운 콘텐츠 접근의 필요성

미래에셋증권은 뉴스, AI를 이용한 섹터 분석, 투자 심리 분석, 리포트 등 다양한 양질의 콘텐츠를 제공하고 있다. 그러나 이러한 정보들은 양이 방대하고 복잡하여, 고객들이 **이를 이해하고 활용하는 데 어려움**을 겪을 수 있다. 따라서 고객이 **손쉽게 접근하고 이해할 수 있는 형태로 정보를 제공하는 것**이 중요하다. 이를 위해 챗봇을 활용하여 고객이 필요로 하는 정보를 빠르고 쉽게 접근할 수 있도록 지원하는 것이 필요하다.

금융 관련 정확한 답변의 필요성

금융 관련 질문에 대해 **정확한 답변**을 제공하는 것은 매우 중요하다. 부정확한 정보는 고객의 경제적 결정에 중대한 영향을 미칠 수 있으며, 이는 경제적 손실로 이어질 수 있다. 또한 금융 분야에는 다양한 법적 규제가 존재하므로, 이를 종합적으로 고려한 답변을 제공하는 것이 필수적이다. **정확하고 신뢰할 수 있는 금융 정보와 조언을 제공하는 서비스**는 고객의 신뢰를 확보하는데 아주 중요하다.

타사 서비스의 한계점

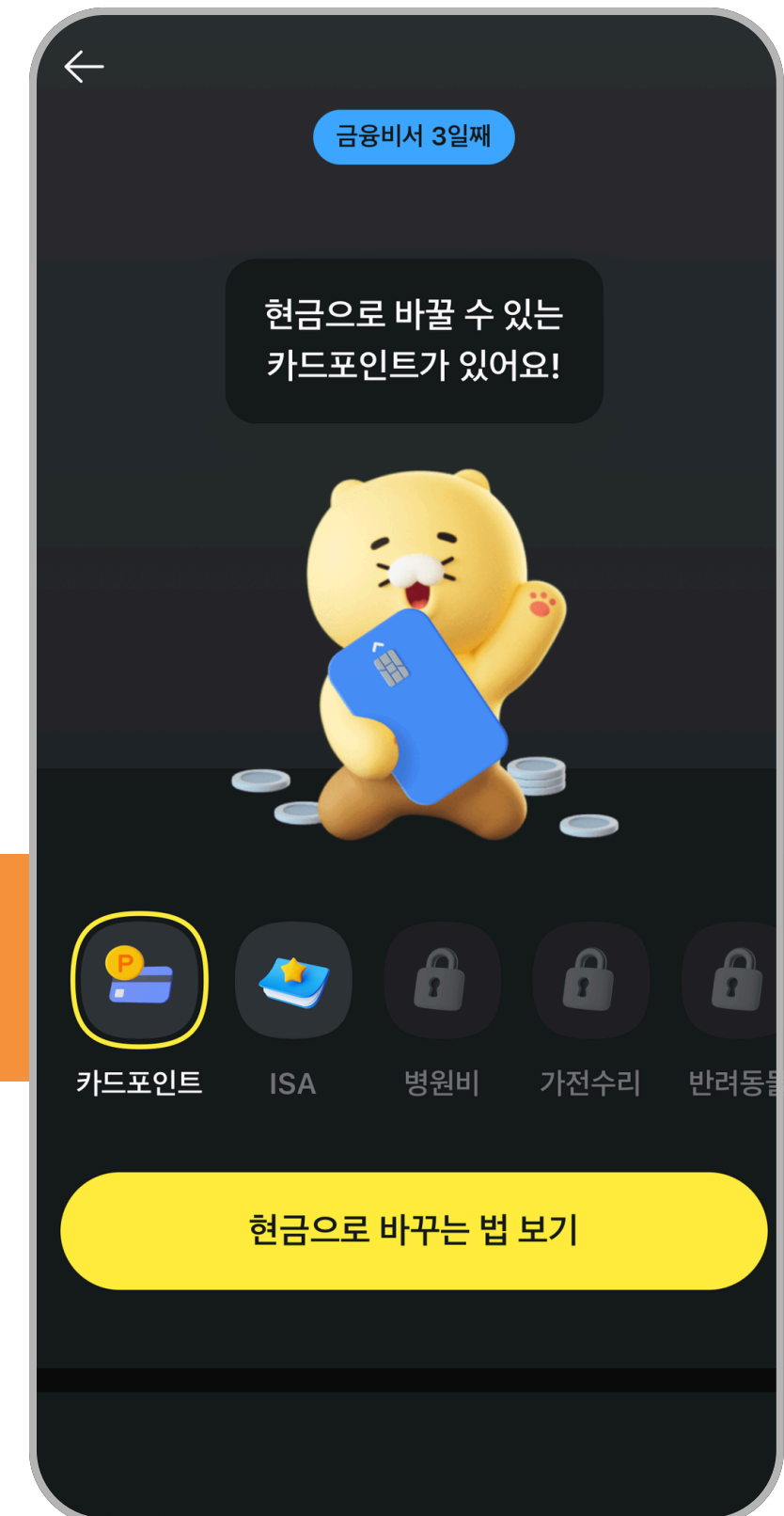


K사의 인공지능 서비스

“난 저축을 얼마나 해야할까?”라는 질문에 “아끼고 절약하는 습관이 최고!”라며 영양가 없는 대답을 하는 등 구체적이고 도움이 되는 내용을 출력하지 못하는 모습

K사의 금융비서 서비스

원하는 내용을 능동적으로 질문하고 찾을 수 있는 형식이 아니고, 몇가지 제안사항 중에 관심있는 내용을 선택해서 보는 제한적 서비스



제안하는 서비스

하이퍼클로바 X와 RAG 기법을 활용한

개인 맞춤형 금융 비서

RAG 기법이란? LLM의 한계를 극복하기 위해 ‘지식 검색’과 ‘언어 생성’을 결합한 프레임워크이다. RAG의 기본 아이디어는 질문에 답하기 위해 필요한 지식을 외부 데이터베이스에서 검색하여 활용하는 것이다.

개인의 정보 바탕

단순한 사용금액이나 통장 잔고만을 고려한 분석이 아닌, 개인의 직업, 직장, 연봉, 자산 규모, 자산 구성 등을 종합적으로 분석한다. 이를 통해 개인 맞춤형 금융 조언을 제공하여 고객의 재정 상태를 개선하고, 최적의 금융 전략을 제안한다.

정확한 답변

금융 분야에서 정확성은 매우 중요하다. RAG 기법을 활용하여 LLM이 답변을 생성할 때 외부 데이터베이스에서 정확한 정보를 검색하고 이를 기반으로 답변을 제공한다. 이렇게 함으로써 허위 정보를 제공할 가능성을 최소화하고, 답변의 출처를 확인할 수 있게 되어 신뢰성을 높인다.

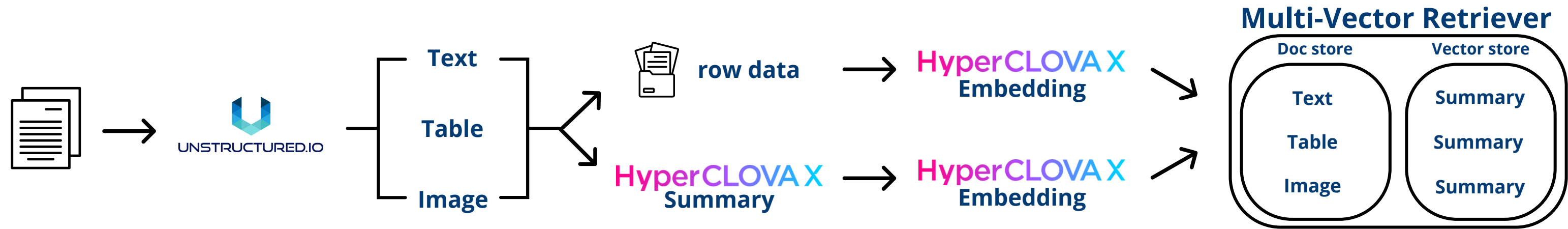
유연한 답변

정해진 답변만 가능한 챗봇들과 달리, 생성형 AI를 사용하여 다양한 질문에 유연하게 답변할 수 있다. 고객의 다양한 금융 관련 질문에 대해 포괄적이고 상세한 답변을 제공하여 고객 만족도를 높인다.

하이퍼클로바 X

고객사의 데이터를 자체 클라우드에 안전하게 저장할 수 있는 ‘소버린 AI’ 전략을 채택하여 데이터 보안과 프라이버시를 철저히 보호한다. 고객의 민감한 금융 데이터를 안전하게 관리하고, 데이터 유출의 위험을 최소화한다.

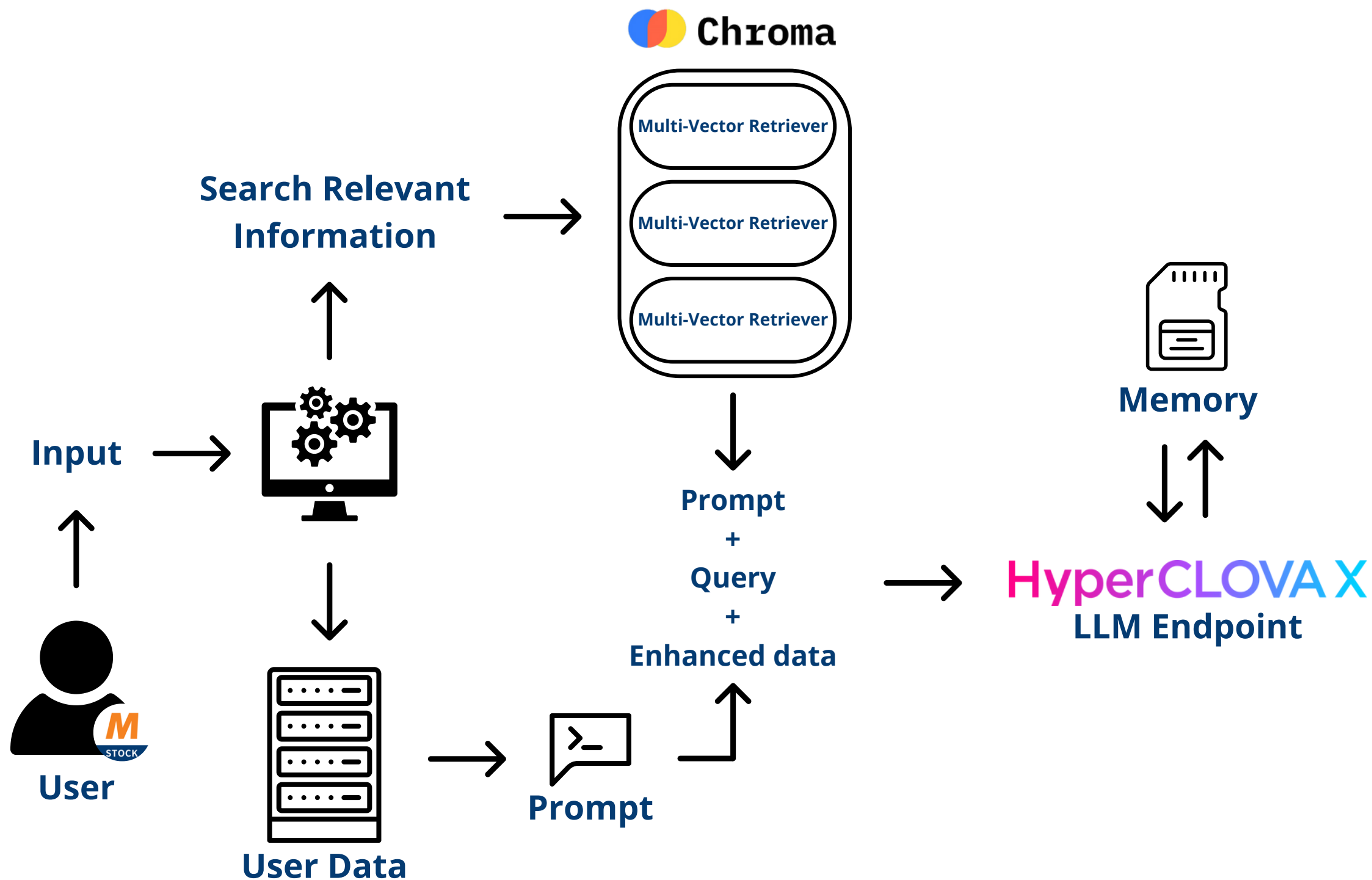
데이터 처리 및 DB저장 구조



1. PDF 파일 형식 원본 데이터 확보
2. 데이터를 UNSTRUCTURED.IO 로 문서의 레이아웃을 분석하여 텍스트, 테이블, 이미지로 구분하여 파싱
3. 데이터를 하이퍼클로바 요약 API로 요약
4. 요약 데이터와 원 데이터를 각각 하이퍼클로바 임베딩 API로 임베딩
5. 임베딩 벡터를 각각 DOCSTORE(원 데이터)와 VECTORSTORE(요약 데이터)에 저장

알고리즘 흐름도

RAG 기법



1. 사용자의 질문 입력
2. 질문의 카테고리 분류
3. 카테고리에 해당하는 리트리버 선택
4. 리트리버로 요약된 데이터를 보고 필요한 데이터 선택
5. 서버에 저장된 사용자의 데이터를 가명화해서 프롬프트에 추가
6. 선택된 데이터와 사용자의 질문과 프롬프트를 하이퍼클로바 LLM에 제공
7. LLM은 새로운 지식과 학습 데이터를 사용하여 더 나은 응답을 생성
8. 메모리로 사용자의 채팅 히스토리 기록

데이터 선정

경제

기획재정부, 한국개발연구원(KDI)
생애주기별 경제교육

유아기부터 노년기까지 주기별 경제상식이 들어있어 사용자의 나이에 맞는 조언 가능

기업

미래에셋증권 리서치 리포트
NAVER SK 하이닉스

미래에셋 서버에 있는 다양한 투자관련 자료를 이용한다면 더 좋은 서비스가 가능할 것

세법

나무경영아카데미_2024개정세법.PDF

주식관련 세금 조언은 물론, 상속세, 종합소득세 등 다양한 세금 문의를 답변하고자 구축

미래에셋

미래에셋증권_2024지속가능경영보고서.PDF

특pecially 미래에셋에 대한 다양한 질문을 답변하고자 미래에셋 카테고리를 구성

RAG 기법을 사용하기 위해서는 Vector DB 구축이 필요하다.

미래에셋의 개인 맞춤형 금융 비서라는 명목에 맞게 위와같이 4개의 카테고리를 나누고 정확도가 높은 자료를 사용하고자 했다.

다만 이는 예시일뿐 실제로는 미래에셋증권에서 제공하는 다양한 자료를 활용하는 것을 전제로 한다.

자료 선정의 기준은 다음과 같다.

- 1) 하이퍼클로바가 학습하지 못한 최신의 데이터인가?
- 2) 신뢰있는 기관에서 발행한 신뢰성있는 자료인가?
- 3) 테이블과 텍스트가 함께 들어있는 자료인가?
- 4) 충분히 전문적이고 깊은 내용이 들어있는 자료인가?

데이터 처리 및 DB저장

멀티모달 정보 추출 및 청킹



문서로부터 텍스트, 테이블 그리고 이미지 정보를 추출한다.

금융과 관련한 PDF자료는 숫자를 기반으로한 테이블을 많이 포함하고 있다. 따라서 텍스트뿐만 아니라 테이블을 잘 이해하고 추출하는 것이 질 좋은 데이터 베이스를 구축하는 핵심요소라고 할 수 있다.

긴 문서를 작은 단위인 청크(CHUNK)로 나눈다. 이렇게 문서를 작은 조각으로 나누는 이유는 LLM 모델의 입력 토큰의 개수가 정해져 있기 때문이다.

요약

HyperCLOVA X Summary

각 문서에 대한 요약을 생성하고, 그것을 문서와 함께 임베딩한다. 요약 임베딩은 문서의 핵심 내용을 신속하게 파악하는 데 도움을 준다.

임베딩

HyperCLOVA X Embedding

임베딩이란 자연어처리에서 사람이 쓰는 자연어를 기계가 이해할 수 있도록 숫자형태인 vector로 바꾸는 과정을 말한다.

이번 프로젝트에서는 네이버 하이퍼클로바 임베딩 V2 api를 이용해 청킹된 데이터를 벡터로 임베딩했다.

멀티 벡터 리트리버



Chroma는 임베딩 벡터를 저장하기 위한 오픈소스 소프트웨어이다. LangChain에서 제공하는 Retriever로 Chroma db에 저장된 벡터스토어를 탐색할 수 있다.

다만 테이블을 임베딩하는 것은 의미적 유사성 검색에 어려움을 초래할 수 있다. 따라서 Multi Vector Retriever를 사용하여 원본 테이블(raw tables), 텍스트와 함께 요약된 테이블(table summaries)을 각각 저장 및 임베딩을 한다.

이를 통해 의미적 유사성 검색을 용이하게 하면서 원본 데이터까지 보존할 수 있게 된다.

카테고리 분류

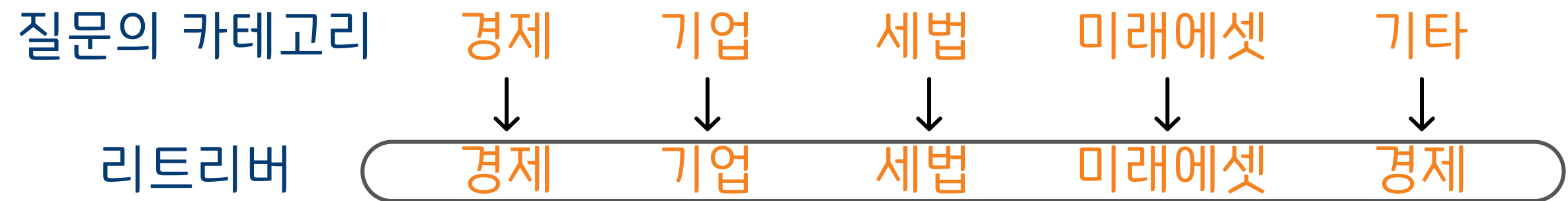
멀티플 리트리버

검색기(RETRIEVER) 단계는 저장된 벡터 데이터베이스에서 사용자의 질문과 관련된 문서를 검색하는 과정이다. 이 단계는 사용자 질문에 가장 적합한 정보를 벡터 데이터베이스 내에서 신속하게 찾아내는 것이 목표이며, RAG 시스템의 전반적인 성능과 직결되는 매우 중요한 과정이다.

여러개의 리트리버를 생성한 후 질문의 의도에 따라 적합한 리트리버를 선택하는 방식을 멀티플 리트리버라고 한다.

리트리버 선택을 위해 질문의 **카테고리 분류**가 필요

HyperCLOVA X | Few-shot 프롬프트 엔지니어링으로 질문을 넣으면 카테고리를 분류하도록 훈련



*기타 카테고리의 경우 가장 일상과 밀접하고 다양한 내용을 가진 경제 리트리버 사용

리트리버를 통해 질문과 가장 가까운 정보를 찾아 주어진 질문이나 문제에 대해 더 정확하고 풍부한 정보를 기반으로 답변을 생성

프롬프트 엔지니어링

프롬프트는 특정 작업을 수행하도록 생성형 AI에 요청하는 자연어 텍스트이다.

LLM이 유용한 출력을 생성할 수 있도록 제공하는 추가 정보인 프롬프트를 잘 조정하는 것이 프롬프트 엔지니어링이다.

사용자에 맞춤형 답변을 제공하기 위해 사용자의 정보를 서버에서 가져와 익명처리하여 프롬프트에 넣는 방식을 사용하고자 했다. 그러나 사용자 정보 데이터에 접근할 수 없었기에 임의로 사용자의 데이터를 입력하면 이를 이용해서 프롬프트에 넣는 형식으로 과정을 구성했다.

또한 질문하는 어미의 형태가 달라지면 답변을 안하는 경우가 확인되어 퓨샷 프롬프트 엔지니어링으로 답변을 안하는 어미를 답변하도록 학습했다.

메모리

대화를 기억하고 맥락을 이해하기 위해 대화 내용을 저장하도록 코드를 작성하였다. 이를 통해 아래 예시처럼 LLM이 대화 기록을 토대로 이어 답변할 수 있게 되었다.

질문 : 최근 유망한 주식 섹터에 대해 키워드만 제시해줘.

최근 유망한 주식 섹터로는 다음과 같은 키워드가 있습니다.

1. 인공지능(AI)
2. 빅데이터
3. 전기차
4. 수소차
5. 바이오

해당 섹터에 속한 기업들의 주식을 분석하여 투자를 결정하는 것이 좋습니다. 단, 주식 시장은 변동성이 있으므로 투자 전에 충분한 조사와 분석이 필요합니다.

질문 : 그것들을 제외하고 하나만 더 제시해줘. 이번에는 설명도 함께 제공해줘.

최근 유망한 주식 섹터 중에서 앞서 제시한 키워드를 제외하고 하나를 더 제시하자면, '우주산업' 섹터를 추천드립니다. 우주산업은 미래 성장 가능성이 매우 높은 분야 중 하나입니다. 우주산업은 크게 위성산업, 우주여행산업, 우주자원개발산업 등으로 나눌 수 있습니다.

- 위성산업은 인공위성을 이용하여 다양한 서비스를 제공하는 산업입니다. 대표적인 예로는 통신위성, 방송위성, 기상위성 등이 있습니다. 최근에는 자율주행차, 드론, IOT 등의 기술 발전으로 인해 위성산업의 중요성이 더욱 커지고 있습니다.

- 우주여행산업은 우주를 여행하는 산업입니다. 아직은 초기 단계이지만, 우주여행에 대한 수요가 증가하면서 우주여행산업의 규모도 점차 커질 것으로 예상됩니다.

- 우주자원개발산업은 우주에서 자원을 개발하는 산업입니다. 대표적인 예로는 달 탐사, 화성 탐사 등이 있습니다. 우주자원개발산업은 인류의 미래를 위한 중요한 산업 중 하나입니다.

우주산업은 아직까지 발전 초기 단계이기 때문에, 투자 시에는 기업의 기술력과 경쟁력, 시장 동향 등을 충분히 파악하고, 장기적인 관점에서 투자하는 것이 좋습니다.

사용 예시 - 실시간 정보 제공

일반 CLOVA X

RAG 적용 모델

MIRAE ASSET
미래에셋증권

질문 : SK하이닉스의 주가 전망 알려줘

저는 주가 예측과 같은 미래 예측형 정보는 제공하지 않습니다.

주식 시장은 다양한 요인에 영향을 받아 변동성이 크기 때문에 예측하기 어렵습니다. 주식 투자를 하실 때는 기업의 재무 상태, 경영 전략, 시장 동향 등을 종합적으로 고려하여 신중하게 결정하시는 것이 좋습니다.

투자 전에 전문가의 조언을 듣거나, 금융기관의 리서치 자료를 참고하시는 것도 좋은 방법입니다.

다른 도움이 필요하시면 말씀해 주세요.

SK하이닉스의 주가 전망은 긍정적입니다. 미래에셋증권은 SK하이닉스에 대한 투자 의견을 '매수'로 유지하고, 목표주가를 기존 240,000원에서 260,000원으로 8% 상향했습니다. 이는 예상보다 강한 메모리 가격 전망에 따라 2024년과 2025년 영업이익 추정치를 각각 23.4%, 62.5% 상향 조정했기 때문입니다.

특히, 2025년 DRAM 수요 비트그로스(BIT GROWTH)는 +23.0%로 전망되고 있으며, 경기 회복에 대한 명확한 전망이 어려운 상황에서 올해 비트그로스(+17.1%)보다 상승폭이 확대되는 부분이 고무적입니다. 수요 증가의 주요 동력원은 컨벤셔널 서버 수요 회복과 HBM 수요 강세 지속입니다.

더불어, 내년 DRAM 업계의 평균 WAFER CAPA 증가율은 14.5%에 그치며 수요 성장률보다 낮을 것으로 전망되어, 상기 수급 동향에 따라 2025년 DRAM의 공급초과율은 -9.3%로 올해(-5.1%)보다 타이트한 수급이 심화될 것으로 보입니다. 이에 따라 2025년 DRAM 예상 가격은 올해에 이어 +40%대의 상승세가 지속될 것으로 전망됩니다.

질문 : SK하이닉스 DRAM의 BIT SHIPMENTS에 대해 분기별로 알려줘

SK하이닉스의 DRAM Bit shipments에 대한 분기별 정보는 해당 기업의 공식 발표 또는 시장 조사 기관의 보고서를 통해 확인할 수 있습니다.

해당 정보를 확인할 수 있는 사이트를 안내해 드리겠습니다.

SK하이닉스: <https://www.skhynix.com/ko/ir/financial-statements/quarterly-report>

DRAMeXchange: <https://www.dramexchange.com/>

단 해당 사이트에서 제공되는 정보는 시간이 지남에 따라 변동될 수 있으므로, 투자 전에 최신 정보를 확인하는 것이 중요합니다.

다른 도움이 필요하시면 언제든지 말씀해 주세요.

네, SK하이닉스 DRAM의 분기별 BIT SHIPMENTS는 다음과 같습니다.

- 1Q23: 12,088 BIL GB
- 2Q23: 16,440 BIL GB
- 3Q23: 19,892 BIL GB
- 4Q23: 20,290 BIL GB
- 1Q24: 17,044 BIL GB
- 2Q24: 20,623 BIL GB

더 궁금하신 점이나 다른 질문이 있으시면 언제든지 말씀해주세요.

표 2. SK하이닉스 주요 제품별 추정치

	1Q23	2Q23	3Q23	4Q23	1Q24	2Q24
DRAM						
Bit shipments (Bil Gb)	12,088	16,440	19,892	20,290	17,044	20,623
Bit growth	-18.0	36.0	21.0	2.0	-16.0	21.0
ASP (\$/Gb)	0.19	0.21	0.23	0.27	0.33	0.38
ASP change	-16.4	9.4	11.1	17.7	22.2	14.5

<실제 PDF에 있는 표의 모습>

테이블의 내용을 정확하게 읽어내서 대답하는 모습

*온전히 실제 모델의 대답을 거짓없이 첨부하였습니다.

질문 : 노후대비 조언해줘

노후대비를 위한 조언을 드리겠습니다.

- 1.저축과 투자: 노후를 대비하기 위해서는 충분한 자금을 마련해야 합니다. 이를 위해서는 저축과 투자를 병행하는 것이 좋습니다. 저축은 안정적인 수익을 얻을 수 있고, 투자는 높은 수익을 얻을 수 있지만 위험도 존재합니다. 자신의 상황에 맞게 적절한 비율로 저축과 투자를 병행하는 것이 좋습니다.
- 2.연금: 국민연금, 퇴직연금, 개인연금 등을 활용하여 노후 생활비를 마련할 수 있습니다. 연금은 장기적으로 적립하고, 일정한 나이가 되면 수령할 수 있습니다.
- 3.건강 관리: 건강은 노후 생활의 질을 결정하는 중요한 요소입니다. 건강 관리를 위해 규칙적인 운동과 식습관 개선 등을 실천하는 것이 좋습니다.
- 4.자기계발: 노후에도 지속적으로 자신의 역량을 향상시키는 것이 중요합니다. 새로운 기술을 습득하거나, 취미생활을 즐기는 등의 활동을 통해 자신의 삶을 더욱 풍요롭게 만들 수 있습니다.
- 5.가족과 소통: 가족과 소통하며 서로의 관심사와 고민을 공유하는 것이 좋습니다. 가족 간의 유대감을 강화하고, 노후 생활을 더욱 즐겁게 보낼 수 있습니다.

위의 조언을 참고하여 노후를 대비하는 데 도움이 되기를 바랍니다.
추가로 궁금하신 점 있으시면 말씀해 주세요.

현재 중견기업에서 22년차로 근무하고 계시는 48세 남성분으로 연 근로소득은 1억 1200만원, 연 불로소득은 300만원, 자산은 10억 3천만원으로 구성되어 있으시네요.

40~50대는 일생 동안 가장 많은 소득을 벌어들이는 시기일 수 있지만, 그동안 축적해 놓은 자산과 앞으로 벌어들일 소득을 효율적으로 관리해야 하는 시기이기도 합니다. 평균 수명이 늘어나고 있는 상황에서 경제적으로 안정적인 생활 수준을 유지하고 불확실한 미래에 대비하기 위해서는 자산 관리가 반드시 필요합니다.

일반적으로 합리적인 자산 관리를 위해 고려할 점으로는 안전성, 수익성, 유동성을 꼽습니다. **사용자님의 경우, 안정적인 자산에 대부분의 자산이 몰려 있어 수익성을 추구하는 금융상품에 투자하는 것을 고려해볼 수 있습니다.**

사용자님의 투자성향을 파악해보면, 안정추구형에 해당하는 것으로 보입니다. 안정추구형 투자자는 투자원금의 손실위험은 최소화하고, 이자소득이나 배당소득 등 안정적인 수익을 추구하는 투자성향을 가지고 있습니다. **이러한 투자성향을 가진 투자자는 은행의 예·적금, 국채, 지방채, 보증채, MMF, CMA 등의 안정적인 금융상품에 투자하는 것이 좋습니다.**

노후대비를 위해서는 금융상품을 선택할 때, 자신의 소득이나 투자의 목적, 자산 상태, 미래 지출 규모 등을 고려해 다양한 자산에 적절히 분배해야 합니다. 자신의 투자 목적과 성향 등을 고려해 상황에 맞게 포트폴리오를 잘 설계했더라도 이를 구성하는 자산의 가격이 변동해 포트폴리오의 구성비율이 달라졌거나, 포트폴리오의 위험이나 수익에 영향을 미칠 수 있는 새로운 정보가 있다면 포트폴리오를 재조정할 필요가 있습니다.

예상치 못한 큰 변수가 발생한다면 포트폴리오 자체를 바꿔야 하며, 금융시장에는 워낙 다양한 금융상품이 있기 때문에 구체적인 상품의 선택이나 투자금액, 투자 기간 등은 전문가의 도움을 받아 결정하는 것이 좋습니다.

일반 모델은 보편적인 사람을 상대로 하는 답변을 하는 것과 달리 RAG를 적용하고 사용자의 개인정보를 받아 학습한 모델은 **사용자의 자산과 성향 등을 분석하여 사용자에게 딱 맞는 맞춤형 대답**을 하는 것을 볼 수 있다.

개인 맞춤형 금융 비서 - 마픽



마픽

Mirae Asset Personalized Finance Chatbot

미래에셋의 개인 맞춤형 금융 챗봇이라는 의미

+

사용자들의 MY PICK 이 되겠다는 의미

Streamlit 으로 챗봇 프로토타입 구현 완료

추후 고도화하여 배포 가능

📁 당신의 정보를 입력
해주세요

(실제로는 개인의 정보를 입력하는 것이 아닌 미래에셋 증권 내부 데이터베이스에서 고객 별로 자동 반영되는 것)

나이

48 - +

성별

남성 ▾

직업

회사원 (중견기업)

근무 경력 (년)

22 - +

가족 구성원 수

4 - +

자녀 수

2 - +

연 근로소득 (만원)

11200 - +

연 불로소득 (만원)

300 - +

총 자산 (만원)

103000 - +

자산 구성

Deploy

미래에셋 개인 맞춤형 금융비서



(with HyperCLOVA X)

⚠ Note: There is an 8-second delay before the model generates a response due to HyperCLOVA API call latency.

sk하이닉스 영업이익을 년도별로 알려줘



네, SK하이닉스의 영업이익을 년도별로 알려드리겠습니다.

- 2023년: 9,112억원
- 2024년 예상: 18,322억원
- 2025년 예상: 31,249억원
- 2026년 예상: 25,224억원

해당 정보는 미래에셋증권 리서치센터에서 제공한 자료를 바탕으로 하고 있으며, 기업의 상황에 따라 변동될 수 있습니다.

What is up?



HyperCLOVA X 와 함께, AI로 만드는 금융투자의 새로운 경험

누리봄 | 강하람 김기남

감사합니다

하이퍼클로바 X와 RAG 기법을 활용한 개인 맞춤형 금융 비서



NAVER Cloud