

Implementation and Advancement of RAG Using LangChain

LangChain을 활용한 RAG 구현 및 고도화

LangChain 개요와 활용 사례

목차 LangChain 개요와 활용 사례

1. LangChain이란?
2. LangChain 패키지 구조
3. LangChain 활용 사례

1. LangChain이란?

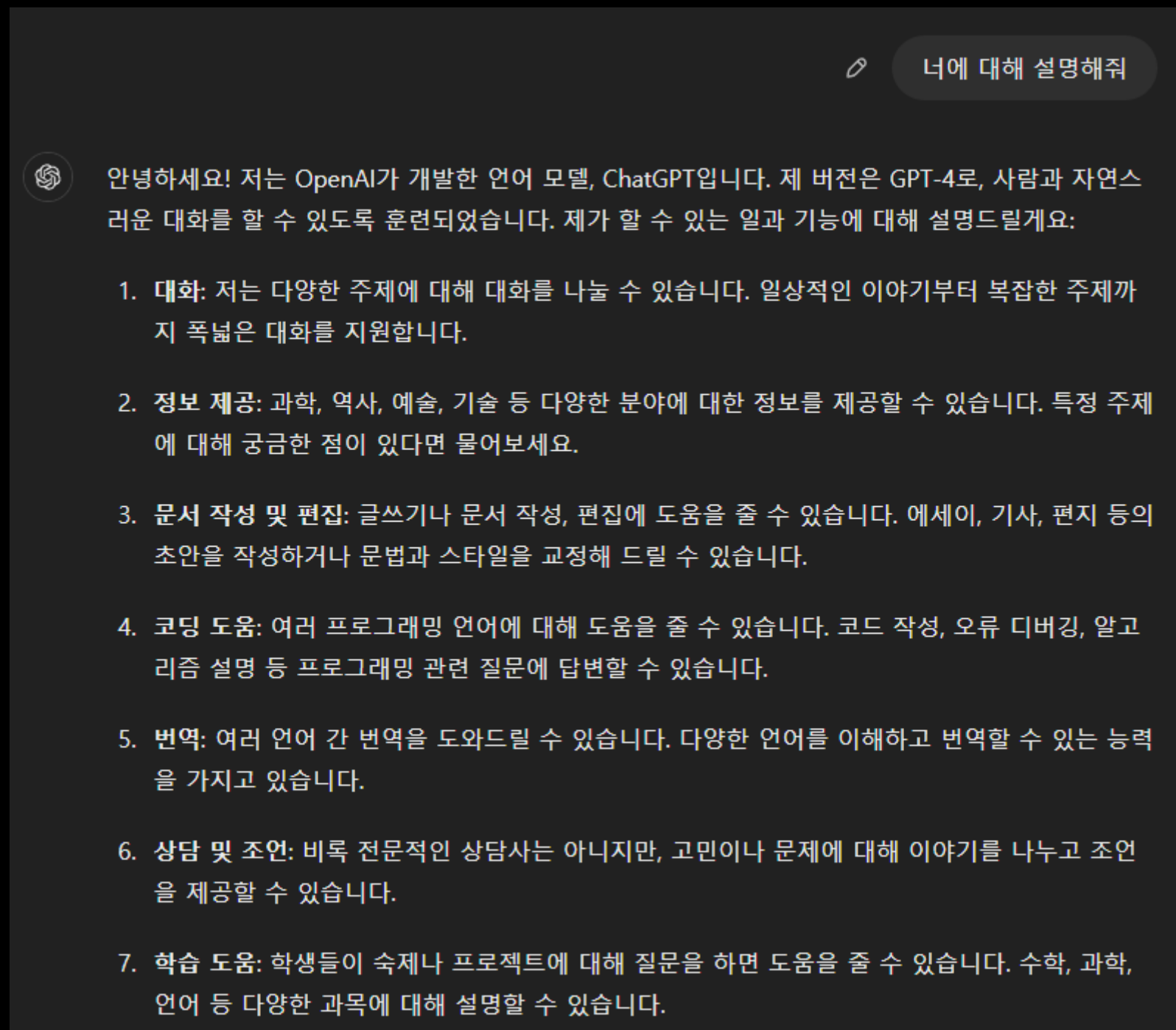


LLM (거대 언어 모델) 을 활용한 어플리케이션을 개발하기 위한 오픈소스 프레임워크



- 언어 모델: 입력된 자연어의 다음에 올 문자, 언어, 형태소 등을 예측하는 모델
- 거대 언어 모델: 언어 모델 중 그 크기 (파라미터 수) 가 매우 큰 경우
 - 많은 양의 언어 데이터를 학습하여 **사람과 자연어로 대화할 수 있고**, 코드 생성, 번역 등 다양한 목적에 맞게 활용할 수 있다.
 - 대표적으로 ChatGPT가 있다.

- 원하는 질문을 입력하면 자동으로 답변을 해주는 생성형 AI
- 2023년 출시 이후 정보 검색, 코드 자동 생성 등 전세계적으로 엄청난 영향을 끼치고 있다.



사용 모델: ChatGPT 4o



- ChatGPT는 본질적으로 다음에 올 단어/형태소 등을 예측하는 AI이기 때문에 학습하지 않은 정보에 대한 질문에는 정확히 답할 수 없다.

제 지식은 2023년 9월까지의 정보에 기반하고 있으며, 이후의 사건이나 변화에 대해서는 알지 못합니다. 또한 인터넷에 직접 접근할 수 없기 때문에 실시간 정보 제공에는 한계가 있습니다.

궁금한 점이 있거나 도움이 필요하시면 언제든지 말씀해 주세요!



- 가령, 2024년 7월에 구글이 공개한 오픈 소스 LLM 모델 Gemma 2에 대해 물어봤을 때, 정확하지 않은 정보를 답하는 것을 확인할 수 있다.
- 이처럼 정확하지 않은 정보를 출력하는 현상을 **할루시네이션(Hallucination)** 현상이라고 칭한다.

Gemma 2 Model에 대해 설명해줘



"Gemma 2 Model"에 대해 구체적인 정보가 없어서 정확하게 어떤 모델인지 판단하기가 어렵습니다. 하지만 일반적으로 "Gemma 2 Model"이라는 용어는 다양한 맥락에서 사용될 수 있습니다. 예를 들어:

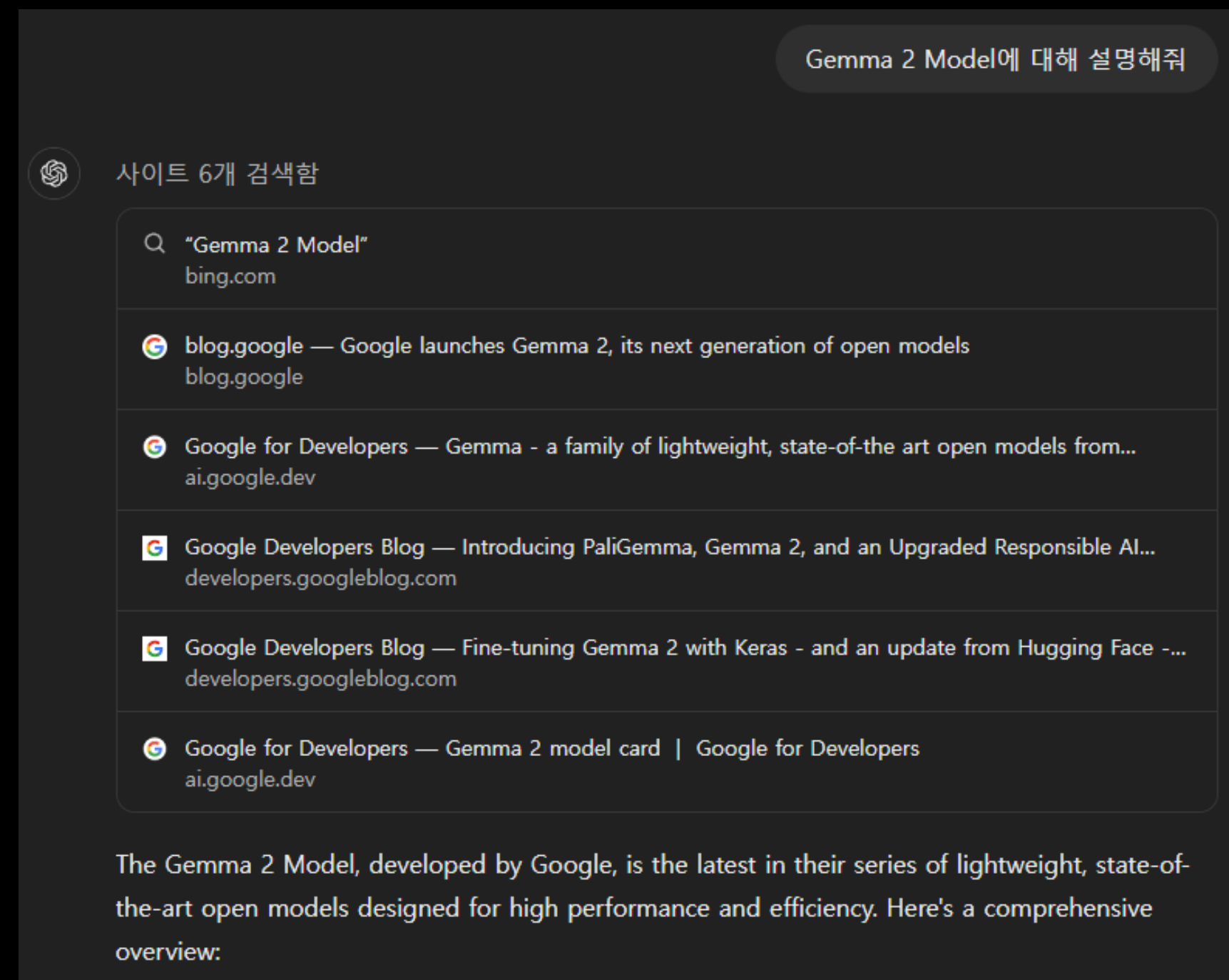
- Arduino Gemma:** Arduino의 작은 마이크로 컨트롤러 보드 중 하나인 Arduino Gemma가 있습니다. 이는 원래 Adafruit에서 만들어진 것으로, 작고 사용하기 쉬운 IoT 프로젝트에 적합합니다.
- Statistical Model:** "Gemma 2"라는 용어가 통계적 모델링에서 사용될 수 있습니다. 예를 들어, 로지스틱 회귀 모델(Gemma 2)과 같은 용어가 사용될 수 있습니다.
- 기타 모델:** 다른 기술 제품이나 시스템에서도 "Gemma 2 Model"이라는 명칭을 사용할 수 있습니다. 예를 들어, 머신 러닝 모델이나 제조업의 특정 장비 모델 등이 있을 수 있습니다.

더 구체적인 맥락이나 정보를 제공하시면, 더욱 정확한 설명을 드릴 수 있을 것입니다.

사용 모델: ChatGPT 3.5



- 할루시네이션 현상은 이번 과정에서 학습할 **RAG (Retrieval-Augmented Generation, 검색 증강 생성)** 기법을 통해 완화할 수 있다.



사용 모델: ChatGPT 4o + 웹 검색 결과



- ChatGPT를 비롯한 LLM은 다음에 올 가능성이 가장 높은 단어/형태소 등을 예측하는 확률 모델이기 때문에, 같은 질문을 해도 원하는 답변이 다르게 나오는 경우가 많다.

사용 모델: ChatGPT 4o

프로그래밍에 입문할 때 가장 먼저 배우면 좋은 언어를 알려줘



프로그래밍에 입문할 때 가장 먼저 배우면 좋은 언어는 여러 가지 요인에 따라 다를 수 있지만, 다음은 일반적으로 추천되는 몇 가지 언어입니다:

1. **Python:** 문법이 간단하고 읽기 쉬우며, 다양한 분야(웹 개발, 데이터 과학, 인공지능 등)에서 활용될 수 있어 입문자에게 가장 많이 추천됩니다.
2. **JavaScript:** 웹 개발을 배우고 싶다면 필수적인 언어입니다. HTML, CSS와 함께 사용되며, 브라우저에서 직접 실행할 수 있어 배우기 쉽습니다.
3. **Scratch:** 어린이나 프로그래밍 완전 초보자를 위한 비주얼 프로그래밍 언어로, 블록을 쌓는 방식으로 프로그래밍 개념을 배울 수 있습니다.

프로그래밍에 입문할 때 가장 먼저 배우면 좋은 언어를 알려줘



프로그래밍에 입문할 때 가장 먼저 배우면 좋은 언어로는 다음 몇 가지를 추천합니다:

1. Python:

- 장점: 문법이 간단하고 읽기 쉬워 초보자에게 적합합니다. 다양한 라이브러리와 커뮤니티 지원이 잘 되어 있어 실습과 학습에 용이합니다.
- 응용 분야: 데이터 분석, 웹 개발, 인공지능, 자동화 등 다방면에서 사용됩니다.

2. JavaScript:

- 장점: 웹 개발의 기본 언어로, 즉시 웹 브라우저에서 실행해 볼 수 있어 피드백이 빠릅니다.
- 응용 분야: 프론트엔드 및 백엔드 웹 개발, 모바일 앱 개발 등.

3. Ruby:

- 장점: 문법이 간결하고 자연어와 비슷하여 배우기 쉽습니다. Rails 프레임워크를 통해 빠르게 웹 애플리케이션을 개발할 수 있습니다.
- 응용 분야: 웹 개발.



- ChatGPT를 비롯한 LLM을 더 잘 사용하기 위한 방법론
- 목적에 맞게 LLM에 입력할 프롬프트(Prompt)를 조정해서 더 정확하고 목적에 맞는 답변을 생성할 수 있다.

REST API를 사용하는 방법을 알려줘

< 3/3 >



REST API(Representational State Transfer API)를 사용하는 방법에 대해 단계별로 설명드리겠습니다. REST API는 HTTP 프로토콜을 사용하여 서로 다른 시스템 간의 데이터를 교환하는 데 사용됩니다. REST API를 사용하기 위해서는 주로 HTTP 요청을 생성하고, 서버로부터 응답을 받는 작업을 합니다.

다음은 REST API를 사용하는 기본적인 단계입니다:

1. REST API 이해하기

주요 HTTP 메서드:

- **GET:** 서버로부터 리소스를 가져옵니다.
- **POST:** 서버에 새로운 리소스를 생성합니다.
- **PUT:** 서버에 존재하는 리소스를 업데이트합니다.
- **DELETE:** 서버에 존재하는 리소스를 삭제합니다.

사용 모델: ChatGPT 4o



REST API를 파이썬 프로그램에서 사용하는 방법을 알려줘



REST API를 파이썬 프로그램에서 사용하는 방법은 몇 가지 단계로 나뉩니다. 여기에서는 ``requests`` 라이브러리를 사용한 예제를 통해 설명하겠습니다. ``requests`` 라이브러리는 REST API 요청을 보내고 응답을 처리하는 데 매우 유용한 라이브러리입니다.

1. ``requests`` 라이브러리 설치

먼저 ``requests`` 라이브러리가 설치되어 있는지 확인하고, 설치되어 있지 않다면 설치합니다.

bash

코드 복사

```
pip install requests
```



- LLM의 기능을 가져와 JS, Python 코드를 통해 자유자재로 활용할 수 있도록 하는 **프레임워크**
- 프롬프트 엔지니어링, RAG 등 LLM의 여러 단점을 보완할 기법을 쉽게 적용할 수 있는 **도구**
- **Language + Chain** 에서 알 수 있듯, 여러 기능, LLM 등을 체인 처럼 엮어 활용할 수 있다.
- 또한, LangChain의 확장 프레임워크인 **LangGraph**를 활용해 더 높은 자유도의 챗봇을 구성할 수 있다.

2. LangChain 패키지 구조



LangChain 패키지 구조

LangChain을 활용한
RAG 구현 및 고도화

LangChain 개요와 활
용 사례

2. LangChain 패키지
구조

LangChain

Langchain의
다양한 컴포넌트를
정의

LangChain
Core

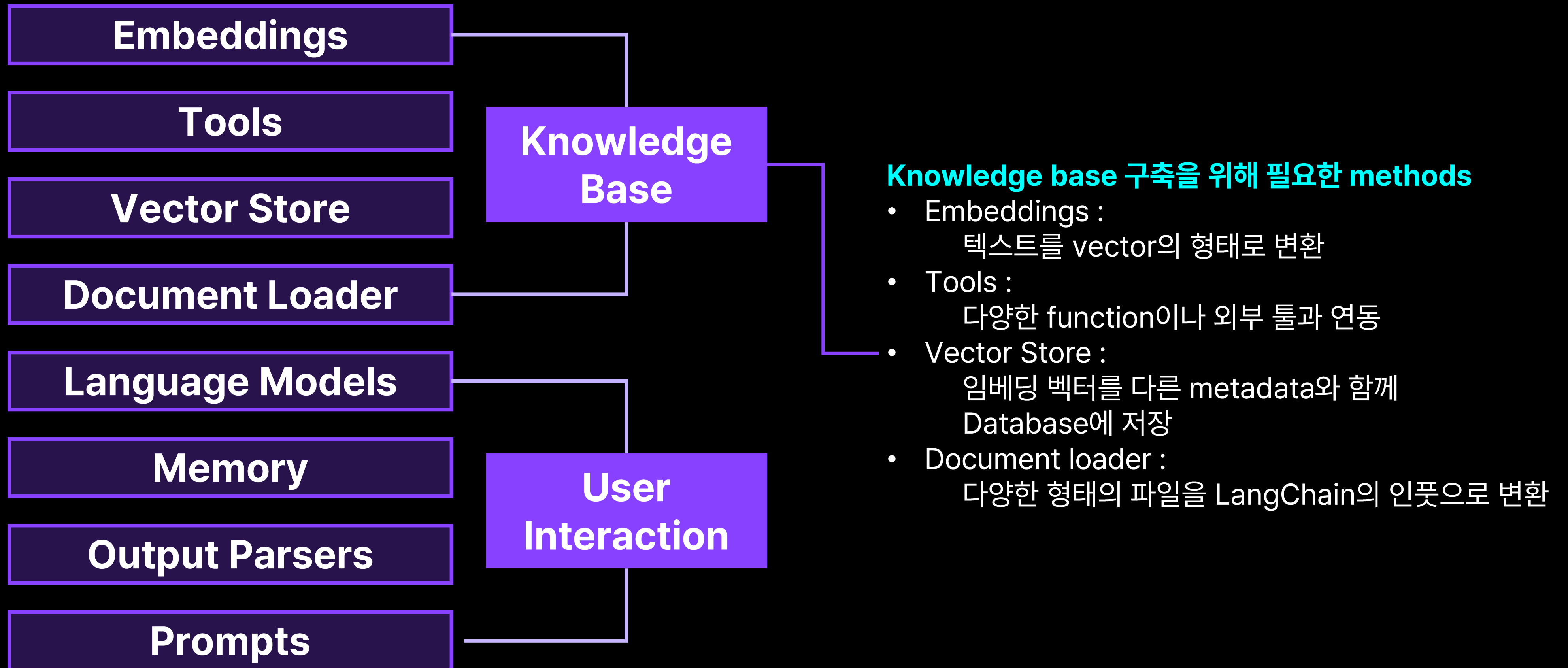
다양한 컴포넌트
구성을 위한 base
package

LangChain
Community

외부 패키지와의
연동을 통해
고도화된 llm 시스템
구축 가능

Experimental

초기 개발 단계인
만큼 다양한
실험적인
function들을
추가함



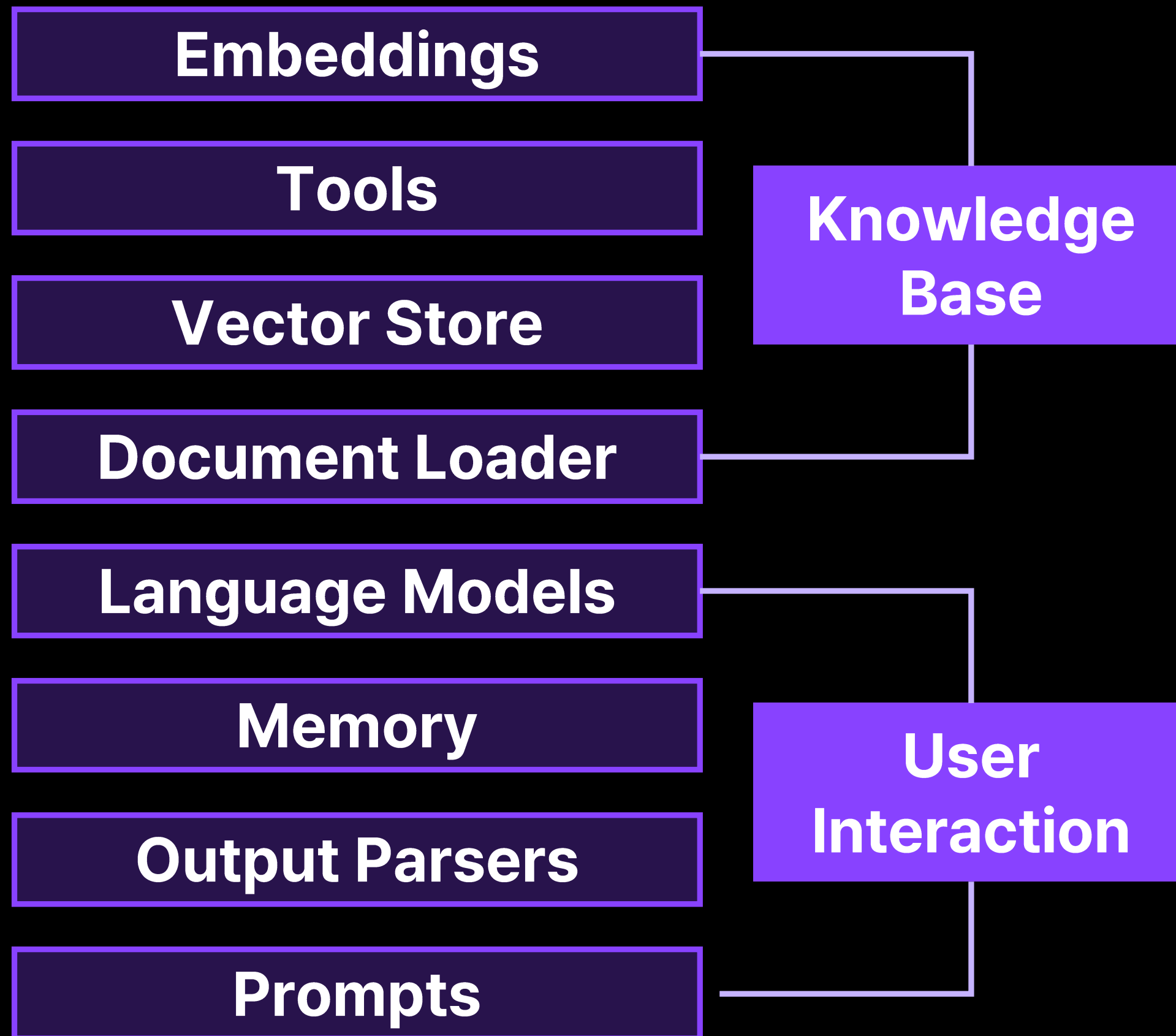


LangChain 패키지 구조

LangChain을 활용한
RAG 구현 및 고도화

LangChain 개요와 활
용 사례

2. LangChain 패키지
구조



사용자 **interaction**을 위해 필요한 **methods**

- Language Models:
GPT 등 가장 핵심이 되는 LLM
- Memory :
사용자와 LLM이 나눈 “대화”의 history를 저장
- Output parsers :
LLM의 아웃풋을 다른 시스템의 input으로
활용 할 수 있도록 특정 형태로 강제함
- Prompts :
LLM의 input

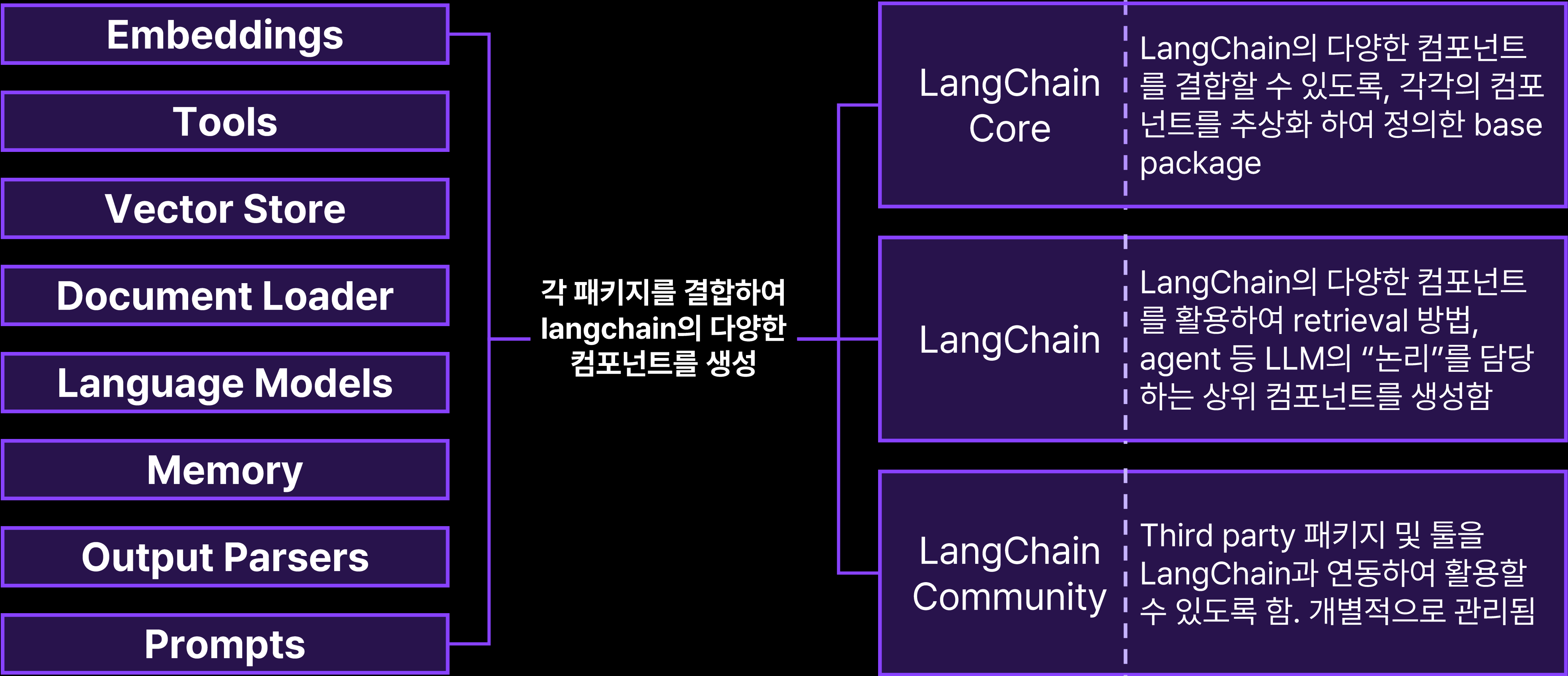


LangChain 패키지 구조

LangChain을 활용한
RAG 구현 및 고도화

LangChain 개요와 활
용 사례

2. LangChain 패키지
구조



3. LangChain 활용 사례



아래 기능을 가진 Excel 챗봇을 구현할 수 있다.

- (1) 개발자가 사전에 등록한 Excel 데이터를 바탕으로 한 시스템 프롬프트를 사용한다.
- (2) 사용자의 질문을 받아 그 질문에 답할 데이터를 추출하는 파이썬 코드를 생성하고 실행한다.
- (3) 추출한 데이터를 바탕으로 답하는 챗봇을 만들 수 있다.

```
In [5]: system_message = "당신은 주어진 데이터를 분석하는 데이터 분석가입니다. \n"
system_message += f"주어진 DataFrame에서 데이터를 출력하여 주어진 질문에 답할 수 있는 파이썬 코드를 작성하세요. {df_name} DataFrame에는 액,
system_message += f"`{df_name}` DataFrame에는 다음과 같은 열이 있습니다: {df_columns}\n"
system_message += "데이터는 이미 로드되어 있으므로 데이터 로드 코드를 생략해야 합니다."

message_with_data_info = [
    ("system", system_message),
    ("human", "{question}"),
]
```



아래 기능을 가진 Excel 챗봇을 구현할 수 있다.

- (1) 개발자가 사전에 등록한 Excel 데이터를 바탕으로 한 시스템 프롬프트를 사용한다.
- (2) 사용자의 질문을 받아 그 질문에 답할 데이터를 추출하는 파이썬 코드를 생성하고 실행한다.
- (3) 추출한 데이터를 바탕으로 답하는 챗봇을 만들 수 있다.

```
In [7]: prompt_with_data_info = ChatPromptTemplate.from_messages(message_with_data_info)

# 체인 구성
code_gen_chain = (
    {"question": RunnablePassthrough()}
    | prompt_with_data_info
    | llm
    | StrOutputParser()
)
```

```
In [8]: print(code_gen_chain.invoke("Velocity가 가장 큰 데이터를 찾아줘"))
```

주어진 `df_inkjet` DataFrame에서 Velocity의 최댓값을 찾는 코드입니다.

```
```python
max_velocity = df_inkjet['Velocity'].max()
print(f"The maximum velocity is: {max_velocity}")
```
```




아래 기능을 가진 Excel 챗봇을 구현할 수 있다.

- (1) 개발자가 사전에 등록한 Excel 데이터를 바탕으로 한 시스템 프롬프트를 사용한다.
- (2) 사용자의 질문을 받아 그 질문에 답할 데이터를 추출하는 파이썬 코드를 생성하고 실행한다.
- (3) 추출한 데이터를 바탕으로 답하는 챗봇을 만들 수 있다.

```
In [14]: code_execute_chain = (  
        code_gen_chain_with_parser |  
        run_code  
    )
```

아까랑 같은 질문을 해봅시다.

```
In [15]: print(code_execute_chain.invoke("Velocity가 가장 큰 데이터를 찾아줘"))
```

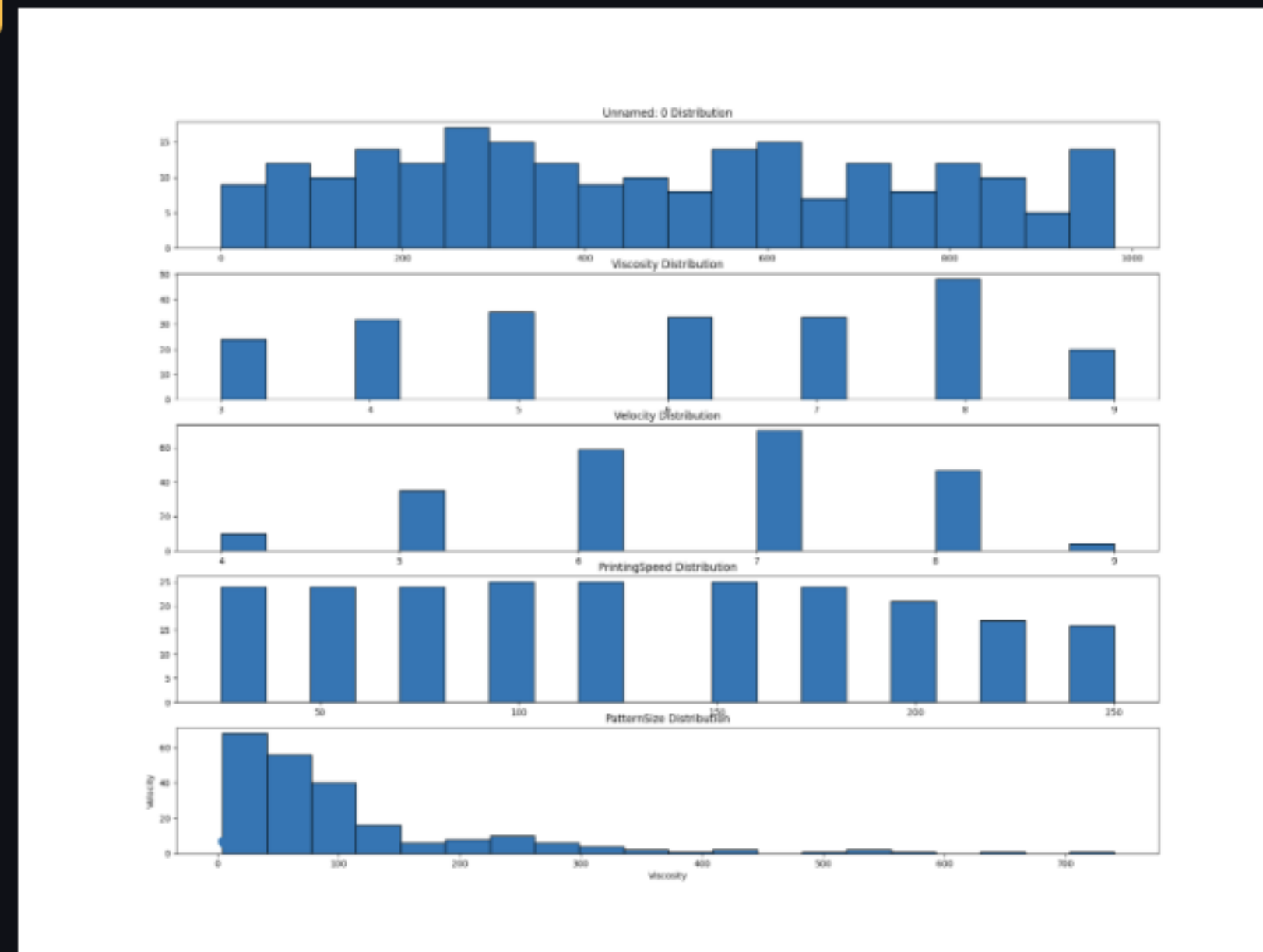
```
Viscosity      8  
Velocity       9  
PrintingSpeed 250  
PatternSize    14  
Name: 125, dtype: int64
```



또한, 데이터를 분석하고 그래프를 그려 출력하는 것도 가능하다.



잉크젯 데이터의 분포를 그래프로 그려줘





RAG (Retrieval-Augmented Generation, 검색 증강 생성) 기법을 활용해서 문서를 근거로 하는 QA 챗봇을 개발할 수 있다.



올해 키오스크 시장의 전망을 알려줘



주제: 국내 키오스크(무인정보단말기) 이용 실태조사 및 개선 방안에 대한 분석

소개: 최근 시장조사국 시장감시팀의 조사 결과, 전반적으로 '외식업'에서는 많은 피해가 발생하고 있다고 나왔습니다. 이러한 상황은 모든 연령대가 외식업 키오스크를 가장 많이 이용하기 때문으로 추정됩니다.

주요 내용:

- '키오스크 기능 표준화' 필요성과 연령대별 불편사례를 보여줍니다.
 - [표7-1-1] '키오스크 기능 표준화' 필요성을 나타내는 테이블을 제시하며, 20대부터 60대까지 모든 연령대에게 도움이 될 것이라고 답변하는 명의 비율이 가장 높습니다.
 - [표7-1-2] 연령대별 키오스크 불편 사례 1~3순위를 나타내는 테이블도 제시하며, '기기 오류', '조작 어려움' 등의 불편함이 20대부터 60대까지 모든 연령대에게 있다고 나타냅니다.
- 디지털 약자층을 포함한 모든 소비자의 키오스크 접근성과 편의성 확보를 위해서는 점진적으로 개선이 필요하다고 설명합니다.

키오스크(무인정보단말기) 이용 실태조사

VII 문제점 및 개선방안

1. 키오스크의 기능 표준화를 통한 이용 편의성 제고

- (문제점) 키오스크 관련 소비자 불만·피해 발생 원인 분석 결과 키오스크 '이용 미숙(18건/18.8%)', '이용법 안내 부족(13건/13.5%)'으로 인한 사례가 많았음.
- 소비자 설문조사 결과, 전 연령대가 키오스크를 통한 주문 순서 등 기능을 표준화하면 키오스크 사용에 도움이 될 것이라고 응답함.

[표7-1-1] '키오스크 기능 표준화' 필요성(N=500)

(단위: 명, %)

| 키오스크 기능 표준화 | 20대 | 30대 | 40대 | 50대 | 60대~ |
|---------------|-----|-----|-----|-----|------|
| 도움이 될 것이다 | 79 | 87 | 77 | 87 | 94 |
| 보통이다 | 14 | 12 | 17 | 11 | 5 |
| 도움이 되지 않을 것이다 | 7 | 1 | 6 | 2 | 1 |

- 또한 전 연령대가 키오스크의 조작이 어렵다고 응답하였는데, 이 역시 업종별 또는 브랜드별로 키오스크의 화면 구성 및 조작 방법 등이 모두 다르기 때문으로 판단됨.

[표7-1-2] 연령대별 키오스크 불편 사례 1~3순위 (중복응답, N=233)

(단위: %)

| 구분 | 20대 | 30대 | 40대 | 50대 | 60대~ |
|-----|------------------|------------------|------------------|------------------|------------------|
| 1순위 | 기기 오류
(70.8) | 뒷사람 눈치
(52.9) | 뒷사람 눈치
(66.7) | 뒷사람 눈치
(61.5) | 조작 어려움
(53.6) |
| 2순위 | 뒷사람 눈치
(47.9) | 조작 어려움
(51.0) | 검색 어려움
(46.2) | 조작 어려움
(51.3) | 검색 어려움
(42.9) |
| 3순위 | 조작 어려움
(35.4) | 기기 오류
(45.1) | 조작 어려움
(41.0) | 할인·적립
(38.5) | 뒷사람 눈치
(41.1) |



RAG (Retrieval-Augmented Generation, 검색 증강 생성) 기법을 활용해서 문서를 근거로 하는 QA 챗봇을 개발할 수 있다.

또한, **LangGraph**를 활용하여 한 챗봇에서 사용자의 질문의 종류를 분석하고, 그에 맞춘 기능을 사용해서 답변하게 만들 수 있다.

3. '무인정보단말기 접근성 지침'에 화면 구성, 조작 순서 등의 구체적인 가이드를 추가하여 키오스크 기능을 표준화함으로써 개선할 수 있다고 제안합니다.

- 올해 키오스크 시장의 전망에 대해서는 일시적인 영향을 받은 듯한 COVID-19로 인한 변경으로 인하여 예측이 어려울 수 있다고 설명합니다.

결론: 최근 시장조사국 시장감시팀의 조사결과에 따른 분석을 통해, '키오스크 기능 표준화'가 필요함으로 설명하고 있습니다. 이를 통해 디지털 약자층을 포함한 모든 소비자의 키오스크 접근성과 편의성 확보를 위해서는 점진적으로 개선이 필요하다고 나타내며, '무인정보단말기 접근성 지침'을 통한 표준화함으로써 개선할 수 있다고 제안합니다.



잉크젯 데이터의 각 컬럼의 평균값을 알려줘



잉크젯 데이터의 각 컬럼의 평균값은 다음과 같습니다.

- Unnamed: 0: 471.986667 (약 472)
- Viscosity: 6.080000
- Velocity: 6.537778
- PrintingSpeed: 129.777778
- PatternSize: 107.284444 (약 107.3)

키오스크 관련 질문에는 RAG 기법 적용

잉크젯 질문에는 Excel 데이터 분석 기능 적용



이외에도 웹 검색 기반 RAG 챗봇, 유튜브 영상 요약 기능 등 LangChain을 통해 LLM을 활용하여 다양한 기능의 챗봇을 구현할 수 있다.