

# RAG의 응답 정밀도를 비약적으로 높이는 최신 데이터 개선 테크닉

By Jaewoo, Kim

RAG(검색증강생성)이라는 기술은 AI를 한층 강력하게 해주는 주목의 기술 중하나입니다. 그러나, RAG의 장점을 최대한 발휘시키기 위해서는 단순히 데이터를 AI에 넣는건만으로는 충분하지 않습니다.

- ◇ 데이터 품질
- ◇ 데이터를 정리하는 방법
- ◇ 프롬프트(AI에 대한 지시) 최적화
- ◇ 검색기법

등의 다양한 요소를 세밀하게 관리해야 합니다. 특히 할루시네이션이라는 문제가 발생할 수 있으므로, 정밀도 향상을 고민해야 합니다.

이번에는 최근에 적용했던 RAG 응답 정밀도를 비약적으로 높이기 위한 개선 기법을 쉽게 설명합니다.

## 1. 원래 RAG란?

우선 RAG에 대해 설명부터 시작합니다. 한마디로 정의하면 '생성형AI와 정보검색을 결합한 기술'이라고 할 수 있습니다. 기존의 생성형 AI는 트레이닝 데이터로부터 학습한 지식에 근거해 응답을 생성하는 것만 실시하고 있었습니다. 그런 기존의 생성형 AI는 교육 데이터의 지식은 제한되어 있으므로 새로운 정보와 상세 데이터에 대한 부정확한 답변을 생성합니다. 이것때문에 등장한 것이 RAG라는 구조입니다.

RAG는 AI가 사전에 학습한 지식 외에도 인터넷 및 데이터베이스에서 실시간으로 정보를 검색하고 그 정보를 바탕으로 고정밀 답변을 생성합니다. 이를 통해 AI는 방대한 정보에서 최적의 답변을 얻을 수 있어 기존 AI보다 훨씬 정확하게 결과를 제공할 수 있습니다. AI에 추가학습시키는 정보는 사내 정보도 가능하고 기업특화의 생성형 AI를 작성하거나 사용할 수 있습니다. 그리고 RAG를 활용시 중요한 것은,

- ✓ 알맞은 데이터 관리,
- ✓ 프롬프트 디자인
- ✓ 검색기법 최적화

입니다.

이것들을 구사하는 것으로 AI정밀도를 비약적으로 향상시키는 것이 가능합니다.

## 2. RAG의 정확성 향상을 위한 기술

RAG응답 정밀도가 보다 정확하고 신뢰성이 높은 것으로 하기 위한 구체적인 테크닉을 소개합니다.

- ◇ 데이터 서비스
- ◇ 프롬프트 엔지니어링
- ◇ 검색기법 최적화
- ◇ 청크분할 및 최적화
- ◇ 피드백 및 개선

### ① 데이터 서비스: 신뢰할 수 있는 출처정보

RAG의 정확성을 극대화하기 위해 가장 중요한 것은 데이터서비스입니다. AI가 적절한 정보를 얻으려면 검색하는 데이터가 정확하고 신뢰할 수 있는 것이 필수적입니다. 데이터 준비는 다음 2지점에 집중해야 합니다.

- 오래된 정보와 새로운 정보 정리
- 메타데이터 활용

#### - 오래된 정보와 새로운 정보 정리

낡은 정보를 그대로 사용하면 AI가 잘못 응답하는 원인이 되므로 AI의 응답 정밀도를 높이기 위해서는 데이터가 항상 최신이어야 합니다. 예를 들어, 회사의 규칙이나 제품정보가 변경되어도 업데이트되지 않은 이전 자료를 참조하면 AI가 잘못된 정보를 제공하게 됩니다.

즉! 최신 지식을 사용하여 답변을 생성하기 위해 정기적으로 정보를 검토하고 최신 데이터셋으로 유지하십시오.

## - 메타데이터 활용

메타데이터는 데이터에 대한 추가 정보입니다.

예를 들어, 문서에는

- ✓ 작성일
- ✓ 작성자명
- ✓ 카테고리

등의 정보를 추가할 수 있습니다. 이렇게 하면 AI가 데이터의 우선순위를 정확하게 결정할 수 있습니다.

메타데이터를 적절히 활용함으로써 AI는 가장 관련성이 높은 정보를 우선적으로 참조하고 정밀한 답변을 생성합니다. 특히 여러 출처에서 선택할 때 이 우선순위가 중요합니다. 즉! 가장 관련성이 높은 정보를 참조하기 위해 메타데이터 활용이 중요합니다.

## ② 프롬프트 엔지니어링: 질문 최적화

프롬프트 엔지니어링은 RAG를 효과적으로 활용하는 핵심기술로 AI에 적절한 지침을 제공하여 보다 정확한 답변을 얻을 수 있습니다. 프롬프트를 최적화하는 몇가지 방법을 소개합니다.

- |   |
|---|
| <ul style="list-style-type: none"><li>◇ 역할 프롬프트 활용</li><li>◇ 작업 명확화</li></ul> |
|---|

## - 역할 프롬프트 활용

“당신은 의료 전문가입니다.”

“당신은 엔지니어입니다.”

이와 같은 상황에 AI에 특정한 역할을 부여함으로써 보다 전문적인 답변을 얻을 수 있습니다. 예를 들어, 기술적인 질문에 대해

“당신은 데이터과학자입니다”

라고 지정함으로써 AI는 전문적인 관점에서 보다 깊은 이해를 나타낼 수 있게 됩니다.

즉! 전문적인 관점에서 더 깊은 이해를 보여주기 위해 AI에 특정 역할을 부여합니다.

## - 작업 명확화

AI에 요구하는 작업을 구체적으로 지시하는 것도 중요합니다.

예를 들어,

“이 문장을 요약하십시오.”

“다음 질문에 간결하게 대답하십시오.”

태스크를 명확하게 지정함으로써 AI는 그 지시에 따라 보다 정확하고 간결한 답변을 생성할 수 있습니다.

즉! AI는 가능한 구체적인 질문 및 작업을 제공합니다.

### ③ 검색기법 최적화: 효율적인 정보 추출

RAG의 정확성을 높이려면 정보검색 기술을 최적화하는 것이 필수적입니다. AI가 방대한 데이터에서 최적의 정보를 얻으려면 적절한 검색방법을 선택하는 것이 매우 중요하며 주로 다음과 같은 유형이 있습니다.

- ✧ 전체 텍스트 검색
- ✧ 시맨틱 검색
- ✧ 벡터 검색
- ✧ 하이브리드 검색

이러한 검색방법을 이해하고 상황에 따라 최적으로 활용하면 AI가 보다 정확한 응답을 생성할 수 있습니다.

## - 전체 텍스트 검색

사용자가 입력한 키워드를 기반으로 데이터베이스나 문서 내에서 일치하는 단어가 포함된 정보를 찾는 방법입니다.

매우 간단하고 직관적인 검색 기술이며 키워드가 그대로 검색조건이므로 특히 특정 단어와 구문을 기반으로 정보가 필요한 경우에 유용합니다. 그러나 전체텍스트 검색은 키워드 일치만으로 결과를 추출하므로 의미가 다른 동의어나 컨텍스트를 이해할 수 없습니다. 따라서 복잡한 질문과 의미를 포함하는 정보를 검색하는데에는 한계가 있으며, 단순한 일치를 초과하는 정확도를 얻으려면 다른 검색방법이 필요합니다.

## - 시맨틱 검색

시맨틱 검색은 사용자 입력 내용의 맥락과 의미를 이해하고 검색하는 방법입니다. 이 기술은 키워드뿐만 아니라, 입력된 쿼리 뒤에 있는 의미를 파싱하고 보다 적절한 정보를 추출합니다. 즉, 단어의 의미와 그 관계를 이해하고 검색할 정보를 찾을 수 있으므로 단어가 일치하지 않아도 관련 정보를 찾을 수 있습니다.

특히 질문이 모호하거나 복잡한 땀등 문맥에 근거한 정보추출이 요구되는 경우에 효과를 발휘합니다.

## - 벡터 검색

벡터 검색은 텍스트 데이터를 숫자형 벡터(다차원 공간의 포인트)로 변환하고 유사도에 따라 검색하는 기술입니다. 이 기술을 사용하면 어휘의 차이나 단어 순서에 관계없이 의미적으로 유사한 정보를 검색할 수 있습니다.

AI가 대량의 정보 중에서 관련성이 높은 정보를 순간적으로 찾아내고, 정밀한 대답을 생성하기 위해 특히 대규모의 데이터베이스나 방대한 텍스트 정보를 취급하는 케이스에서 편리합니다.

## - 하이브리드 검색

하이브리드 검색은 여러 검색 알고리즘을 결합하여 검색하는 기술입니다.

구체적으로는,

- ✓ 전체텍스트 검색 및 의미 검색
- ✓ 벡터검색 및 시맨틱 검색

와 같은 검색 방법을 동시에 사용함으로써 각 방법의 장점을 최대한 활용하여 정밀도를 향상시키려는 방법입니다. 예를 들어, 다음과 같은 방법을 사용하면 응답 정밀도를 더욱 높일 수 있습니다.

- ✓ 전체 텍스트 검색에서 특정 키워드 추출: 결과에 대한 의미 검색
- ✓ 벡터 검색을 사용하여 대량의 데이터에서 관련성이 높은 정보 추출: 결과를 의미적으로 조사

하이브리드 검색의 가장 큰 장점은 서로 다른 검색 알고리즘의 장점을 결합하여 특정 문제에 대한 최적의 정보를 추출할 수 있다는 것입니다. 이렇게 하면 사용자의 질문이 복

잡하거나 의미 있는 정보가 엄청난 경우에는 정확한 응답을 신속하게 생성할 수 있습니다.

### 3. 청크분할 및 최적화 (데이터의 적절한 처리)

RAG의 정확성을 높이려면 데이터를 적절한 청크로 분할하는 것도 중요합니다.

텍스트 데이터를 적절히 분할하고 AI가 효율적인 데이터를 처리할 수 있도록 하여 보다 정확한 답변을 기대할 수 있습니다.

데이터

- ✓ 어떤 단위로 나눌 것인가?
- ✓ 인접한 청크와의 중복도를 조정하는 방법

정확도향상을 크게 영향을 마칩니다. 예를 들어 컨텍스트를 유지하면서 적절한 크기로 청크를 분리하면 더 정확한 정보 추출이 가능합니다.

### 4. 피드백 및 개선: 지속적으로 최적화

마지막으로 RAG정확성을 유지하기 위해 지속적인 피드백이 필수적입니다. 사용자의 피드백을 바탕으로 데이터 프롬프트를 개선하고 AI의 정확성을 지속적으로 향상시켜 봅시다.

AI의 정확도를 평가하기 위해,

- ✓ 충실도
- ✓ 관련성
- ✓ 재현성

같은 지표르 사용합니다.

이미지 지표를 바탕으로 AI의 성능을 측정하고 필요에 따라 개선합니다.

### 5. 정리

이런 개선기술을 실천함으로써 RAG를 구사한 AI 정밀도를 비약적으로 향상시킬 수 있는 것 기대할 수 있습니다.