

# 테스트 계획 및 결과 보고서

## 1. 테스트 계획

### (1) 테스트 항목 기준

테스트 항목	평가 내용
응답 정확성	생성된 답변이 참조 답변과 비교해서 정확한 정보인지 평가
컨텍스트 관련성	검색된 컨텍스트가 질문에 대답하기 위해 관련성이 높은지 평가
컨텍스트 충실성	생성된 답변이 주어진 컨텍스트에 기반하는지 평가
고객 친화 UI	PDF 업로드, 답변 생성과 같이 서비스를 사용함에 불편함이 없는지 평가

### (2) 테스트 문항 예시

{"instruction": "고객에게 감동을 주는 콘텐츠를 어떻게 만들어야 할까요?", "output": "고객의 귀를 사로잡는 콘텐츠는 그들의 상상력을 자극합니다. HD 품질의 팟캐스트 처럼, 여러분의 메시지가 선명하고 매력적이어야 합니다. 진정한 혁신은 단지 정보가 아니라, 경험을 전달하는 것입니다."}

{"instruction": "효율적인 시연을 위해 어떤 태도를 가져야 할까요?", "output": "효율은 성공의 열쇠입니다. 시연이 중요하다면, 준비와 집중이 필수입니다. 당신의 노력을 통해 진정한 가치를 보여주십시오."}

{"instruction": "기업의 성공을 위해서는 어떤 요소들이 중요할까요?", "output": "애플의 성공은 혁신적인 아이디어와 함께, 모든 구성원이 함께 성장하는 환경에서 비롯됩니다. 직원들에게 주식 옵션을 부여하여 소속감을 강화하고, 제품의 유통 경로를 극대화하는 것이 중요합니다. 당신의 기업도 이러한 원칙을 적용하여 새로운 가능성을 열어보세요."}

{"instruction": "어떻게 하면 제품의 매력을 자연스럽게 끌어낼 수 있을까요?", "output": "사진에 생명을 불어넣으십시오. 간단한 터치로 매력을 더하고, 고객의 감성을 사로잡으세요. 진정한 경험을 제공하는 것이 핵심입니다."}

{"instruction": "제품 간의 연계를 어떻게 강화해야 할까요?", "output": "모든 제품은 하나의 경험을 위해 함께 작동해야 합니다. 아이북스처럼, 기술의 통합은 사용자의 상상을 초월하는 가능성을 열어줍니다. 제품을 연결하여 새로운 가치를 창출하십시오."}

{"instruction": "제품의 구조적 요소는 어떻게 제품의 가치를 형성하나요?", "output": "제품의 핵심은 보이지 않는 곳에서부터 시작됩니다. 강력한 구조는 그 자체로 아름답

다움을 만들어냅니다. 제품의 디자인은 단순한 외형이 아니라, 깊은 의미와 기능을 담고 있어야 합니다."}

{"instruction": "어떻게 우리의 메시지를 강력하게 전달하여 소비자의 마음을 사로잡을 수 있을까요?", "output": "천 개의 노래가 당신의 손안에 있습니다. 단순한 기술이 아닌, 음악을 간직하는 새로운 방식입니다. 사람들의 삶을 변화시키는 힘을 지니고 있습니다."}

{"instruction": "제품의 다기능성을 어떻게 강조해야 할까요?", "output": "아이팟은 단순한 음악 플레이어가 아닙니다. 천 곡의 노래를 주머니에 담고, 동시에 파일을 옮길 수 있는 강력한 도구로 발전했습니다. 하나의 기기로 두 가지 이상의 가치를 제공할 때, 진정한 혁신이 시작됩니다."}

{"instruction": "어떤 제품의 성공을 어떻게 비교해야 할까요?", "output": "성공은 숫자로 말합니다. 경쟁자를 뛰어넘는 성과를 통해 자신을 드러내십시오. 진정한 가치는 비교를 통해 더욱 빛납니다."}

{"instruction": "경쟁 제품과의 차별화를 어떻게 효과적으로 표현해야 할까요?", "output": "비교할 수 없는 뛰어난 제품은 다른 차원에서 존재합니다. 소비자에게 신선하고 강렬한 인상을 남기십시오. 그들이 느끼는 것, 바로 그것이 당신의 브랜드입니다."}

{"instruction": "협업을 통해 혁신을 이끌어내는 방법은 무엇인가요?", "output": "협업은 힘을 모아 새로운 가능성을 여는 열쇠입니다. 각자의 강점을 결합하면, 고객에게는 완벽한 경험을 제공할 수 있습니다. 함께하는 즐거움 속에서 혁신은 탄생합니다."}

{"instruction": "고객에게 가격 이상의 가치를 어떻게 전달해야 할까요?", "output": "가격이 아닌 가치를 판매하십시오. 고객은 단순히 숫자를 보지 않습니다. 그들은 경험과 혁신을 원합니다."}

## 2. 테스트 결과

### (1) 테스트 응답 결과

샘플 1 (idx=94)

질문:

새로운 제품을 어떻게 효과적으로 연결할 수 있을까요?

정답:

기술은 단순히 도구가 아닙니다. 우리 제품을 통해 사람들이 서로 연결되고 경험을 공유하도록 만드십시오. 여러분의 제품은 삶의 모든 순간을 더욱 풍요롭게 하는 다리입니다.

모델 응답:

새로운 제품의 새로운 특성을 반영할 수 있는 새로운 유통 경로를 개발하는 전략을 사용해야 합니다.

RAG:

이러한 경우 광고를 통해 소비자로 하여금 이전 제품이 아닌 새로운 제품에 집중하도록 함으로써 해당 브랜드와 이전 제품 간의 연결 관계를 약화시킬 필요가 있다

이는 기업이 신제품과 관련하여 적합한 세일즈 프로모션 전략을 수립하는 데 실무적인 함의점도 제시할 수 있을 것이다

제품이 가진 새로운 특성을 반영할 수 있는 새로운 유통 경로를 개발하는 전략이라고 할 수 있다

정성 평가:

```
{  
  "응답 정확성 (Answer Correctness)": 4,  
  "컨텍스트 관련성 (Context Relevance)": 5,  
  "컨텍스트 충실성 (Context Faithfulness)": 5,  
  "컨텍스트 충분성 (Context Recall)": 4,  
  "응답 속도 (Response Speed)": 5  
}
```

응답시간: 1.19초

=====

샘플 2 (idx=138)

질문:

완벽한 제품을 만드는 데 필요한 요소는 무엇인가요?

정답:

단순한 제품이 아니라, 그 뒤에 숨은 수많은 노력과 기술이 중요합니다. 고객이 자연스럽게 사용할 수 있도록 모든 것을 조화롭게 통합해야 합니다. 진정한 혁신은 복잡한 것을 간단하게 만드는 데서 시작합니다.

모델 응답:

완벽한 제품을 만드는 데 필요한 요소는 사전 소비자 테스트를 통해 제품의 부합도를 높이는 것입니다.

RAG:

또한 제품인도 후에도 소비자의 제품만족도 등을 수시로 체크하여 불만의 소지가 없도록 하는 것이 중요하다

즉, 소비자의 욕구를 만 족시켜 낮은 성공률을 높일 수 있는 신제품만이 기업에 게 의미가 있다

출연제품 선택에 있어 사전 소비자 테스트를 통해 최소한 지금보다는 부합도가 높은 제품을 선정하는 전략이 요청된다

정성 평가:

```
{
  "응답 정확성 (Answer Correctness)": 3,
  "컨텍스트 관련성 (Context Relevance)": 4,
  "컨텍스트 충실성 (Context Faithfulness)": 4,
  "컨텍스트 충분성 (Context Recall)": 3,
  "응답 속도 (Response Speed)": 4
}
```

응답시간: 1.48초

=====

샘플 3 (idx=218)

질문:

제품의 성능을 어떻게 개선해야 할까요?

정답:

성능은 생명입니다. 느린 제품은 소비자의 마음을 사로잡을 수 없습니다. 고품질의 디스플레이와 빠른 프로세서로 소비자에게 감동을 주십시오.

모델 응답:

제품의 성능을 개선하려면, 소비자들이 실제로 필요로 하는 기능을 파악하고, 그 기능들을 더 쉽고, 더 빠르고, 더 정확하게 제공해야 합니다. 또한, 제품의 성능을 눈에 보이는 데이터나 숫자로 명확하게 보여주는 것이 중요합니다.

RAG:

출연제품 선택에 있어 사전 소비자 테스트를 통해 최소한 지금보다는 부합도가 높은 제품을 선정하는 전략이 요청된다

향 후 연구에서는 제품의 수를 늘려 조금 더 확실한 수치를 나타내는 제품선정이 필요하다고 생각되 어진다

또한 구체적으로 제품의 성능을 보여주면서 구매자들의 욕구에 가장 효과적인 방식으로 설득하고 협상할 수 있다

정성 평가:

```
{
  "응답 정확성 (Answer Correctness)": 4,
  "컨텍스트 관련성 (Context Relevance)": 4,
  "컨텍스트 충실성 (Context Faithfulness)": 4,
  "컨텍스트 충분성 (Context Recall)": 3,
  "응답 속도 (Response Speed)": 5
}
```

응답시간: 1.13초

=====

샘플 4 (idx=153)

질문:

고객과의 소통에서 어떤 방식이 가장 효과적일까요?

정답:

진정한 소통은 반복에서 시작됩니다. 고객의 목소리를 귀 기울여 듣고, 그들의 이름을 부르는 것처럼 개인적인 관계를 형성하세요. 진심이 담긴 대화는 브랜드의 신뢰를 쌓는 가장 강력한 방법입니다.

모델 응답:

고객과의 소통에서 가장 효과적인 방식은 고객의 요구에 맞춰 메시지를 개발하고 전달 하면서 그들의 반응에 귀를 기울이는 것입니다. 고객과의 상호작용을 통해 신뢰를 쌓고 관계를 구축하는 것이 중요합니다.

RAG:

소비 자가 요구하는 커뮤니케이션에 맞춰 효과적으 로 소통하려면 채널의 특성을 고려 해 역할을 배 분하는 유기적이고 협업이 가능한 시스템을 마 련해야 한다 도준웅, 2013 즉 먼저 고객이 말하는 것을 듣고 이에 맞춰 메시지를 개발하여 전달하고 또한 그들의 반응에 귀를 기울인다

특히 영업사원은 소비자와의 직접 대 면을 통한 상호작용으로 소비자들의 추가질문이나 불만사항에 융통성있게 대처함으로써 고 객과의 신뢰를 쌓고 관계를 구축하는데 중요한 역할을 할 수 있다

정성 평가:

```
{
  "응답 정확성 (Answer Correctness)": 5,
  "컨텍스트 관련성 (Context Relevance)": 5,
  "컨텍스트 충실성 (Context Faithfulness)": 5,
}
```

```
"컨텍스트 충분성 (Context Recall)": 5,  
"응답 속도 (Response Speed)": 5  
}
```

응답시간: 0.74초

=====

샘플 5 (idx=10)

질문:

고객이 원하는 것은 무엇인가요?

정답:

고객은 다양한 선택을 원합니다. 접근이 용이하고 편리해야 합니다. 이것이 고객의 마음을 사로잡는 핵심입니다.

모델 응답:

고객이 원하는 것은 자신이 필요로 하는 것, 즉 Needs입니다. 그들은 또한 원하는 것, 즉 Wants도 가지고 있습니다.

RAG:

이러한 방식으로 고객의 필요 Needs 와 욕구 Wants 에 보다 잘 부응할 수 있는 것이다

다시 말해 고객 애호도는 소비자가 갖게 되는 만족의 결과로서 얻어지는 효과라고 할 수 있다

아울러 고객과의 접점에서 종사하는 서비스 제공자는 고객들이 어떠한 서비스를 원하는지를 잘 알 수 있으므로 이를 최고 경영자와 관리자에게 신속, 정확하게 전달해주는 것이 필요하다

정성 평가:

```
{  
  "응답 정확성 (Answer Correctness)": 3,  
  "컨텍스트 관련성 (Context Relevance)": 4,  
  "컨텍스트 충실성 (Context Faithfulness)": 4,  
  "컨텍스트 충분성 (Context Recall)": 3,  
  "응답 속도 (Response Speed)": 4  
}
```

응답시간: 4.95초

=====

샘플 6 (idx=115)

질문:

제품의 진정한 가치를 어떻게 보여줘야 할까요?

정답:

제품은 단순히 기능이 아니라 경험입니다. 고객이 실제로 무엇을 느끼고, 어떻게 사용할 지를 보여주어야 합니다. 진정한 아름다움은 그 안에서 드러납니다.

모델 응답:

제품의 가치를 보여주기 위해서는, 제품이 가진 가장 훌륭하고 감동적인 순간을 고객에게 먼저 경험하게 해야 합니다. 영화의 클라이맥스 장면처럼, 고객의 기억에 오래도록 남을 만한 가장 중요한 순간을 통해 제품의 본질적인 가치를 증명해야 합니다.

RAG:

또한 구체적으로 제품의 성능을 보여주면서 구매자들의 욕구에 가장 효과적인 방식으로 설득하고 협상할 수 있다

기업은 이제 제품을 판매하기 전에 자신이 세상에 어떤 가치를 기여 하는지, 그리고 그 기여가 얼마나 진정성 있고 지속가능한지 를 먼저 증명해야 한다

즉, 일반 고객이 제품 영화 를 경험 관람 후 제 품에 대한 전체적인 수준 가치 를 평가하는 데 있어, 각 막 별로 가장 훌륭한 장면에 주로 기반하여 제품 에 대한 전체적 평가를 함을 알 수 있었다

정성 평가:

```
{
  "응답 정확성 (Answer Correctness)": 4,
  "컨텍스트 관련성 (Context Relevance)": 5,
  "컨텍스트 충실성 (Context Faithfulness)": 4,
  "컨텍스트 충분성 (Context Recall)": 4,
  "응답 속도 (Response Speed)": 5
}
```

응답시간: 1.06초

=====

샘플 7 (idx=139)

질문:

어떤 방식으로 제품의 매력을 극대화할 수 있을까요?

정답:

우리는 제품의 아름다움과 기능을 결합해야 합니다. 눈을 사로잡는 디자인과 신속한 접근성을 통해 소비자에게 감동을 주십시오. 제품이 단순한 도구가 아닌, 예술작품처럼 느껴지도록 만들어야 합니다.

모델 응답:

제품의 성능과 기능을 구체적인 예시를 통해 보여주면서, 고객의 구매 욕구를 자극하는 방식으로 설득해야 합니다.

RAG:

비록 제품 자체 가 크게 눈에 보이는 것이 아님에도 불구하고 브랜드 개성을 살리고 이미지화하여 고객의 감성에 어필하는 전략이 강조되고 있다  
자연스럽게 이는 제품에 대한 관심과 호감, 행동으로 이어질 가능성이 높다  
또한 구체적으로 제품의 성능을 보여주면서 구매자들의 욕구에 가장 효과적인 방식으로 설득하고 협상할 수 있다

정성 평가:

```
{  
  "응답 정확성 (Answer Correctness)": 4,  
  "컨텍스트 관련성 (Context Relevance)": 5,  
  "컨텍스트 충실성 (Context Faithfulness)": 5,  
  "컨텍스트 충분성 (Context Recall)": 4,  
  "응답 속도 (Response Speed)": 5  
}
```

응답시간: 1.26초

=====

샘플 8 (idx=95)

질문:

팀의 성공을 어떻게 축하하고 격려해야 할까요?

정답:

팀의 노고는 우리가 이루는 모든 성과의 뿌리입니다. 그들을 격려하고 인정함으로써 우리는 더 큰 혁신을 이끌어낼 수 있습니다. 성공은 혼자가 아닌 함께 나누는 것입니다.

모델 응답:

팀의 성공을 축하하고 격려하기 위해서는, 그들의 노력을 인정하고 성과를 공개적으로 칭찬하며, 그들의 열정과 헌신에 대해 진심으로 감사하는 마음을 표현하는 것이 중요합니다. 또한, 그들의 노고를 잊지 않고 다음 목표를 향해 함께 나아갈 것을 다짐하며, 그



들이 더 큰 도전을 향해 나아갈 수 있도록 동기를 부여해야 합니다.

RAG:

본 연구의 결과를 종합하여 살펴보면, 태도적 충성도를 높이 기 위해서는 구단 관계자는 경기력과 경기 결과에 영향을 미치는 상품의 질 향상에 중점을 두어야 할 것이다. 목표를 향해 함께하는 팀웍 정신은 유연한 사고와 열린 마음으로 활기찬 기업 문화를 만들고, 서로에 대한 배려와 칭찬, 그리고 격려를 아끼지 않는 일체감을 형성하고자 하는 정신이다.

아울러, 도전 정신을 가지고 끝까지 최선을 다하는 직원에게 희망을 주는 인사를 구현하고, 신속한 의사 결정으로 영업을 지원하는 조직을 운영하며, 최고의 성과 창출을 위한 제도적 지원을 모색하겠다는 것이다.

정성 평가:

```
{  
  "응답 정확성 (Answer Correctness)": 4,  
  "컨텍스트 관련성 (Context Relevance)": 4,  
  "컨텍스트 충실성 (Context Faithfulness)": 3,  
  "컨텍스트 충분성 (Context Recall)": 3,  
  "응답 속도 (Response Speed)": 5  
}
```

응답시간: 1.70초

=====

샘플 9 (idx=200)

질문:

제품의 성공을 위해 가장 중요한 요소는 무엇인가요?

정답:

혁신은 기술만이 아닙니다. 소비자의 마음을 사로잡고, 그들의 필요를 이해하는 것이 진정한 성공의 열쇠입니다. 제품이 실패할 때, 그 원인을 찾고 다시 도전하십시오.

모델 응답:

제품의 성공을 위해 가장 중요한 요소는 소비자의 욕구를 만족시키는 것입니다. 즉, 제품이 시장에 출시되었을 때 소비자들이 실제로 필요로 하는 것을 채워주고, 그 가치를 명확하게 전달하는 것이 가장 중요합니다.

RAG:

둘째, 제품의 품질, 판매 가격, 판촉의 혜택과 서비스의 체험은 구조 모델에서 네 가지

요소에 가장 큰 영향을 미치는 기준이다

즉, 소비자의 욕구를 만 족시켜 낮은 성공률을 높일 수 있는 신제품만이 기업에 게 의미가 있다

그중 실적 효과에 영향이 가장 큰 게 가격이고, 그 뒤로 제품과 판촉, 마지막으로 서비스이다

정성 평가:

```
{
  "응답 정확성 (Answer Correctness)": 3,
  "컨텍스트 관련성 (Context Relevance)": 4,
  "컨텍스트 충실성 (Context Faithfulness)": 3,
  "컨텍스트 충분성 (Context Recall)": 4,
  "응답 속도 (Response Speed)": 4
}
```

응답시간: 1.68초

=====

샘플 10 (idx=299)

질문:

사용자가 원하는 것을 어떻게 쉽게 선택하게 할 수 있을까요?

정답:

사용자가 선택할 수 있는 자유를 주세요. 간단한 클릭으로 그들의 여정을 시작하게 하십시오. 경험이란 직관적이어야 하며, 선택은 명확해야 합니다.

모델 응답:

용자가 원하는 것을 쉽게 선택하게 하려면, 제품이나 서비스를 쉽게 찾을 수 있도록 검색 기능을 제공하거나, 원하는 것을 쉽게 상기시키는 정보를 제공해야 합니다. 이것이 사람들이 다른 조건이 동일하다면 더욱 유용하다고 느끼는 이유입니다.

RAG:

사용자 편의 정보에 쉽게 접근 가능하거나, 검색하는 기능을 가리킨다  
이를 통해 소비자가 특정한 제품을 쉽게 상기시키거나 발견할 수 있도록 돕기 때문이다  
즉 다른 조건이 동일하다면 이용자들은 이용하기에 쉬운 기술이 더욱 유용하다고 느끼게 되는 것이다 이재신 · 이민영, 2006a

정성 평가:

```
{
```

```
"응답 정확성 (Answer Correctness)": 4,  
"컨텍스트 관련성 (Context Relevance)": 5,  
"컨텍스트 충실성 (Context Faithfulness)": 5,  
"컨텍스트 충분성 (Context Recall)": 4,  
"응답 속도 (Response Speed)": 4  
}
```

응답시간: 1.53초

## (2) 상세 평가 기준

### 1. 응답 정확성 (Answer Correctness, 1-5점)

- **정의:** 답변이 참조 정답(label)과 의미적으로 얼마나 일치하는가?
- **기준:**
  - 5점: 핵심 내용과 세부사항까지 정확히 일치
  - 4점: 핵심은 정확하지만 부가 설명이 일부 다름
  - 3점: 중요한 내용 중 일부 누락/부정확
  - 2점: 핵심과 다르게 설명했으나 약간의 관련성 있음
  - 1점: 전혀 맞지 않는 답변

### 2. 컨텍스트 관련성 (Context Relevance, 1-5점)

- **정의:** 검색된 컨텍스트가 질문과 얼마나 밀접한가?
- **기준:**
  - 5점: 질문과 직접적으로 연결되는 컨텍스트만 포함
  - 4점: 대부분 관련 있지만 약간 불필요한 정보 있음
  - 3점: 일부 관련 있으나 중요하지 않은 맥락이 많음
  - 2점: 질문과 간접적인 관련만 있음
  - 1점: 거의 무관한 컨텍스트

### 3. 컨텍스트 충실성 (Context Faithfulness, 1-5점)

- **정의:** 답변이 검색된 컨텍스트를 충실하게 따르는 정도
- **기준:**
  - 5점: 답변이 전적으로 컨텍스트 기반, 할루시네이션 없음
  - 4점: 대부분 충실하나 약간의 추론/확장 포함
  - 3점: 절반은 컨텍스트, 절반은 모델 추론
  - 2점: 컨텍스트 무시하고 임의 추론 위주
  - 1점: 컨텍스트와 무관한 답변

---

#### 4. 컨텍스트 충분성 (Context Recall, 1-5점)

- **정의:** 검색된 컨텍스트가 질문에 답하기에 충분한가?
- **기준:**
  - 5점: 질문 답변에 필요한 정보가 모두 포함
  - 4점: 대부분 포함, 일부 보강 필요
  - 3점: 핵심 일부만 포함
  - 2점: 충분하지 않음, 답변하기 어려움
  - 1점: 완전히 부족

---

#### 5. 응답 속도 (Response Speed, 0-5점)

- **정의:** 모델이 답변을 생성하는 시간
- **기준:**
  - 5점: 1초 미만
  - 4점: 1-2초
  - 3점: 2-5초
  - 2점: 5-10초
  - 1점 이하: 10초 이상

### (3) 평가 결과

항목	평균 점수
----	-------

응답 정확성	3.8
컨텍스트 관련성	4.5
컨텍스트 충실성	4.2
컨텍스트 충분성	3.7
응답 속도	4.6
Total	4.16

## 테스트 결론

모델의 답변 성능을 분석한 결과, 전반적으로 높은 수준의 정확성과 컨텍스트 관련성을 보여 줌

- 모델은 대부분 질문에 대해 올바른 정보를 제공하며, 컨텍스트를 적절히 활용하고 있음
- 다만 일부 답변에서는 컨텍스트 충실성과 충분성 측면에서 개선 여지가 있어, 특정 정보 누락이나 불필요한 추가 정보가 발생할 수 있음을 확인함
- 응답 속도는 안정적이며 실시간 활용에도 적합한 수준

### 종합 평가:

모델은 현 단계에서 실무에 활용 가능한 수준이며, 컨텍스트 충실성과 충분성을 보완한다면 사용자 맞춤형 조언 제공 및 정보 검색 정확성 측면에서 더욱 신뢰할 수 있는 도구로 발전할 수 있습니다.