

모델 구현 보고서

제출일자 : 2025.02.12. (수)

작성 팀원 : 장정호

▶ 프로젝트 개요

- 프로젝트 주제 : 카메라 사용자 매뉴얼 검색 시스템

- 프로젝트 목표

- 1) 고객상담 업무 효율성을 극대화하기 위해 구매고객을 대상으로 서비스되고 있는 문서의 검색 시스템을 질의 응답 형식으로 구축
- 2) 카메라 사용자의 카메라에 대한 정보 접근성 확대 및 초기 카메라 문제 발생 시 간편 해결 방안 제시로 편의성 증대

▶ LLM 활용 소프트웨어

1. Query Analysis

- **역할:** 사용자의 질문을 분석하고, 질문 의도(브랜드 정보, 모델 정보, 카메라 관련 질문 여부, 매뉴얼(설정/사용법) 관련 여부)를 파악
- **LLM 활용:**
 - 문장 및 키워드 파싱: 사용자의 질문에서 브랜드, 모델, 주요 키워드를 추출
 - 질문 유효성 검토: 질문이 카메라와 관련이 있는지 여부를 판단

2. Refine Question

- **역할:** 사용자의 질문이 불명확하거나 정보가 부족할 경우, 역질문을 하고 그 답변을 다시 분석
- **LLM 활용:**
 - 불충분한 질의 맥락을 보완하기 위해 추가 정보를 요청
 - 추가로 받은 답변을 재분석하여 완전한 질의 형태로 정제

▶ LLM 활용 소프트웨어

3. Setting Generate

- **역할:** 카메라에 대한 질문이지만 매뉴얼과 관련된 명확한 질문이 아니라면, 매뉴얼과 관련된 질문으로 재작성
내부적으로 활용하기 쉬운 형태로 질문을 재구성(세부 매개변수·설정값 등)
- **LLM 활용:**
 - 자연어로 된 질의를 매뉴얼 기반의 설정 중심 질문으로 변환
 - “음식을 맛있게 촬영하는 방법” → “음식 사진 촬영 시 추천되는 화이트 밸런스, 조리개 값, 조명 설정 방법을 알려줘”

4. SONY, CANON, FUJI RAG

1. Query Expansion (SONY, CANON, FUJI)

- **역할:** 사용자가 입력한 질문을 기반으로 여러 다른 형태의 질문을 생성하여, 벡터 데이터베이스에서 더 관련성 높은 문서를 검색할 수 있도록 돕는다.
- **LLM 활용:**
 - 사용자의 원래 질문을 다양한 방식으로 변환 (동의어, 유사 개념, 질문 구조 변경 등)
 - 핵심 개념을 유지하면서도 다른 표현 방식으로 질문을 재구성
 - 검색 엔진의 거리 기반 유사성 검색의 한계를 극복할 수 있도록 의미적 다양성을 고려

2. Query Decompose (FUJI)

- **역할:** 복합적 질의를 더 작은 단위로 분해하여 처리
- **LLM 활용:**
 - “후지 카메라 색감 설정과 노이즈 억제 방법” → “1) 색감 설정, 2) 노이즈 억제” 식으로 나누어 각각 검색

3. Generate (SONY, CANON, FUJI)

- **역할:** 최종적으로 필터·정렬된 문서를 참고해 사용자에게 제시할 답변을 생성
- **LLM 활용:**
 - 요약, 구체적 매뉴얼 안내, 단계별 설정 가이드 등
 - 예) “메뉴 → [카메라 설정] → [AF 모드]에서 ‘연속 AF(AF-C)’로 변경하십시오” 등 구체적 표현

▶ LLM 활용 소프트웨어

5. Relevance Check

- **역할:** `langchain_upstage` 라이브러리 내 `UpstageGroundednessCheck`를 활용해 생성된 답변이 **질문·맥락과 합치되는지** 검증
- **LLM 활용:**
 - 답변이 엉뚱한 내용은 아닌지, 브랜드나 기능 측면에서 정확히 대응하는지 체크
 - 만약 불일치가 심하면 **Rewrite Query** 단계로 이동

6. Rewrite Query

- **역할:** Relevance Check에서 부적합 판정이 내려진 질문을 RAG 파이프라인이 보다 정확한 정보를 검색할 수 있도록 구체적이고 맥락이 풍부한 형태로 재작성
- **LLM 활용:**
 - 질문의 주요 키워드를 유지하면서도 세부 정보를 추가하여 명확성을 향상
 - 특정 모델, 오류 코드, 기능 옵션 등을 포함하여 질문을 보다 구체적으로 변환
 - 원래 질문이 지나치게 일반적이거나 맥락이 부족한 경우, 관련 사례 또는 조건을 반영하여 보완
 - RAG 파이프라인이 보다 정확한 결과를 반환할 수 있도록 검색 최적화된 형태로 변형

7. Extract Keyword

- **역할:** 최종 답변 내 핵심 키워드 추출
- **LLM 활용:**
 - 자연어 답변에서 주요 개념(“AF 모드”, “렌즈 오류”, “펌웨어 업데이트”)을 추출

8. Suggest Question

- **역할:** 사용자가 카메라 사용자 매뉴얼을 기반으로 한 질문에 대해 답변을 얻은 후, 추가로 궁금해할 수 있는 관련 질문을 추천한다.
- **LLM 활용:**
 - 사용자의 기존 질문과 답변을 분석하여 논리적 후속 질문을 생성
 - 현재 답변이 다루지 않은 관련된 기능 또는 추가 설정을 탐색하는 질문 포함

▶ LLM 활용 소프트웨어

9. Keyword Explain

- **역할:** 사용자가 입력한 카메라 관련 키워드를 검색하여, 그 용어에 대한 쉽고 간결한 설명을 제공한다. 검색된 정보를 기반으로 정확성과 관련성을 유지하면서도 이해하기 쉬운 방식으로 정리한다.
- **LLM 활용:**
 - Tavily API를 이용해 키워드 관련 정보를 검색
 - 검색 결과에서 사진 촬영 및 카메라 기술과 관련된 정보만 추출
 - LLM을 활용하여 복잡한 기술 용어를 간결하고 직관적으로 설명
 - 필요할 경우 관련 개념을 덧붙여 사용자 이해도를 향상

▶ 학습된 인공지능 모델

RAG system의 전체 파이프라인

라우팅(Routing) 기법을 사용해 브랜드별 RAG로 분기되는 **RAGsystem**의 전체 파이프라인에 대한 개요입니다. 각 단계는 LLM과의 결합을 통해 사용자 질의에 대한 분석, 검색, 답변 생성을 수행합니다.

A. Query Processing

1. **Query Analysis**
 - 사용자 질문이 유효한지(카메라 관련 여부, 형식 등) 판별
 - 질문 의도를 파악해 **브랜드 라우팅** 혹은 추가 정보 요청 등을 결정
2. **Decide Next Step**
 - Query Analysis 결과에 따라 Sony/Canon/Fuji RAG 또는 Not For Camera 노드로 분기하며, 쿼리가 불명확하면 Refine Question 노드로 분기한다.
3. **Refine Question**
 - 질의가 불분명하면 역질문을 하여 사용자 응답을 받아 **질문을 보강**
 - 이후 보강된 질문을 다시 Query Analysis 수행
4. **Setting Generate**
 - 카메라에 대한 질문이지만 매뉴얼과 관련된 **명확한 질문이 아닐 경우**, 해당 질문을 매뉴얼과 관련된 형태로 **재작성**
5. **Not For Camera**
 - 카메라 분야와 무관한 질문이면, 안내 메시지를 출력하고 프로세스 종료

▶ 학습된 인공지능 모델

B. 브랜드별 RAG

브랜드별 RAG는 공통적으로 **검색(Retrieval)**과 **답변 생성(Generate)** 과정이 포함되어 있으나, 일부 세부 단계가 다릅니다.

1. SONY RAG

- **Query Expansion**: 사용자가 입력한 질문을 기반으로 여러 다른 형태의 질문을 생성
- **Ensemble Retrieval**: VectorDB(의미 기반) + BM25(키워드 기반) 결과를 결합하여 최적의 문서 목록 생성
- **Reranker**: 검색된 문서 순위를 재조정해 **가장 적절한 자료**를 상단에 배치
- **Generate**: LLM이 최종 답변(매뉴얼 안내, 문제 해결 가이드 등)을 자연어로 작성

2. CANON RAG

- **Query Expansion**: 사용자가 입력한 질문을 기반으로 여러 다른 형태의 질문을 생성
- **Ensemble Retrieval**: 의미+키워드 기반 검색 결과를 취합
- **Filter**: Canon 관련성이 떨어지거나 불필요한 정보를 제거
- **Reranker**: 필터 후 남은 문서들을 **우선순위 재정렬**
- **Generate**: 적절한 문서 내용을 종합하여 **최종 답변**을 생성

3. FUJI RAG

- **Query Decompose**: 복합적 질의(다중 질문)를 작은 단위로 나누어 처리
- **Query Expansion**: Decompose된 질의 각각에 대해서 여러 다른 형태의 질문을 생성
- **Hybrid Retrieval**: VectorDB + BM25를 결합
- **Reranker**: 가장 타당한 문서를 상위에 위치
- **Generate**: 각 소질문 결과를 통합해 답변을 생성

C. Post-processing

1. Relevance Check

- 생성된 답변이 **문맥과 합치**되는지 자동 검증(**UpstageGroundednessCheck** 등 활용)
- 불일치가 하면 다음 단계(Rewrite Query)로 이동

2. Rewrite Query

- 답변이 충분히 정확하지 않거나 맥락에 어긋날 경우, **질문을 재작성**하여 다시 RAG에 투입

3. Extract Keyword

- 최종 답변 내 **핵심 키워드**를 추출(예: 초점, ISO, 펌웨어 등)

4. Suggest Question

- 현재 답변과 연관된 **후속 질문**을 자동 생성
- 사용자에게 추가적인 정보를 탐색할 기회를 제공

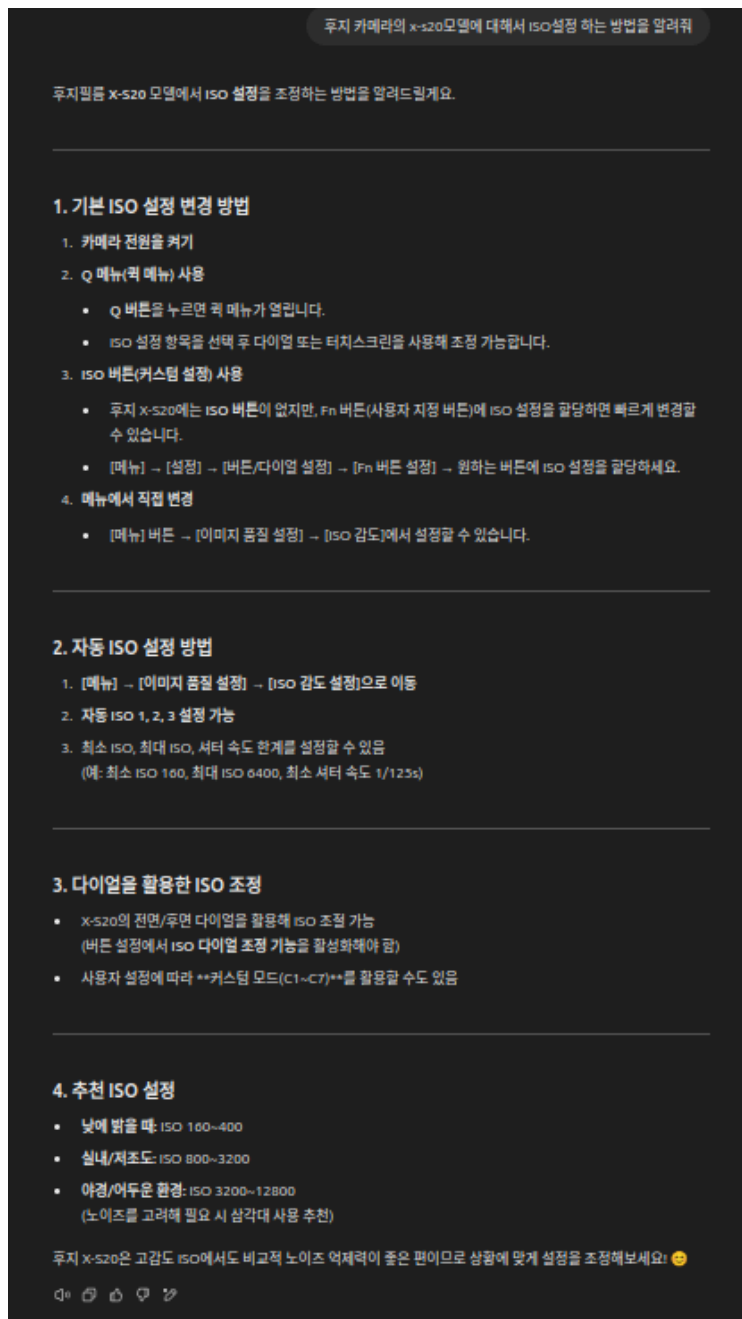
▶ 상황 예시 및 답변 비교

1. Fuji

1.1 사용자 질문 (예시)

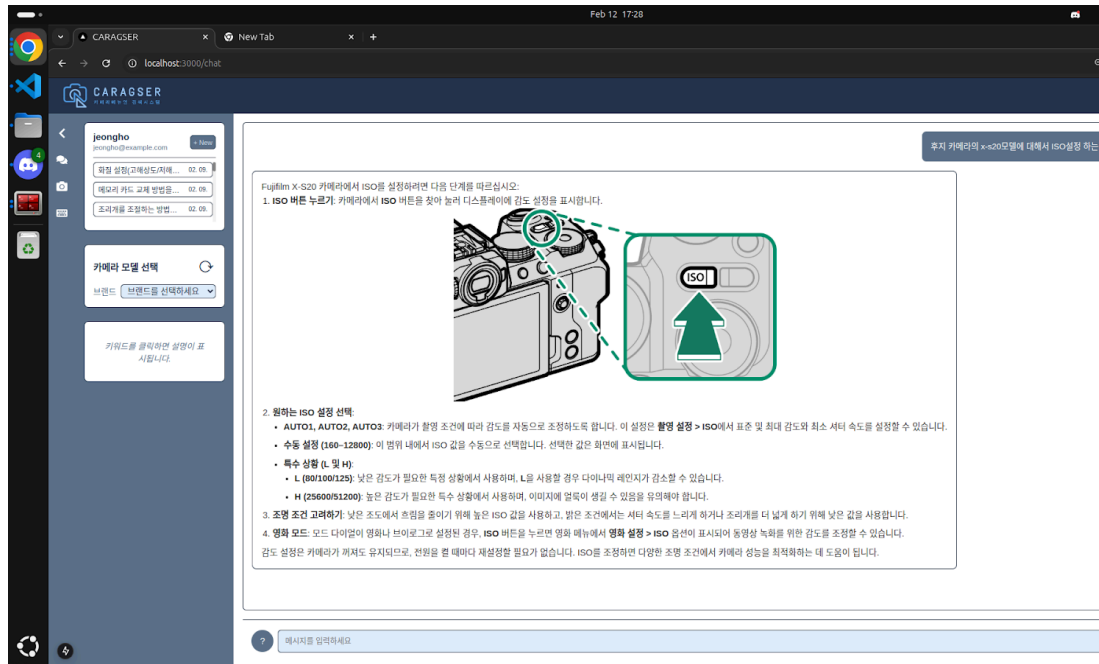
- 질문: 후지 카메라의 x-s20모델에 대해서 ISO설정 하는 방법을 알려줘

1.2 Chat GPT 응답 스크린샷



▶ 상황 예시 및 답변 비교

1.3 RAG system 응답 스크린샷



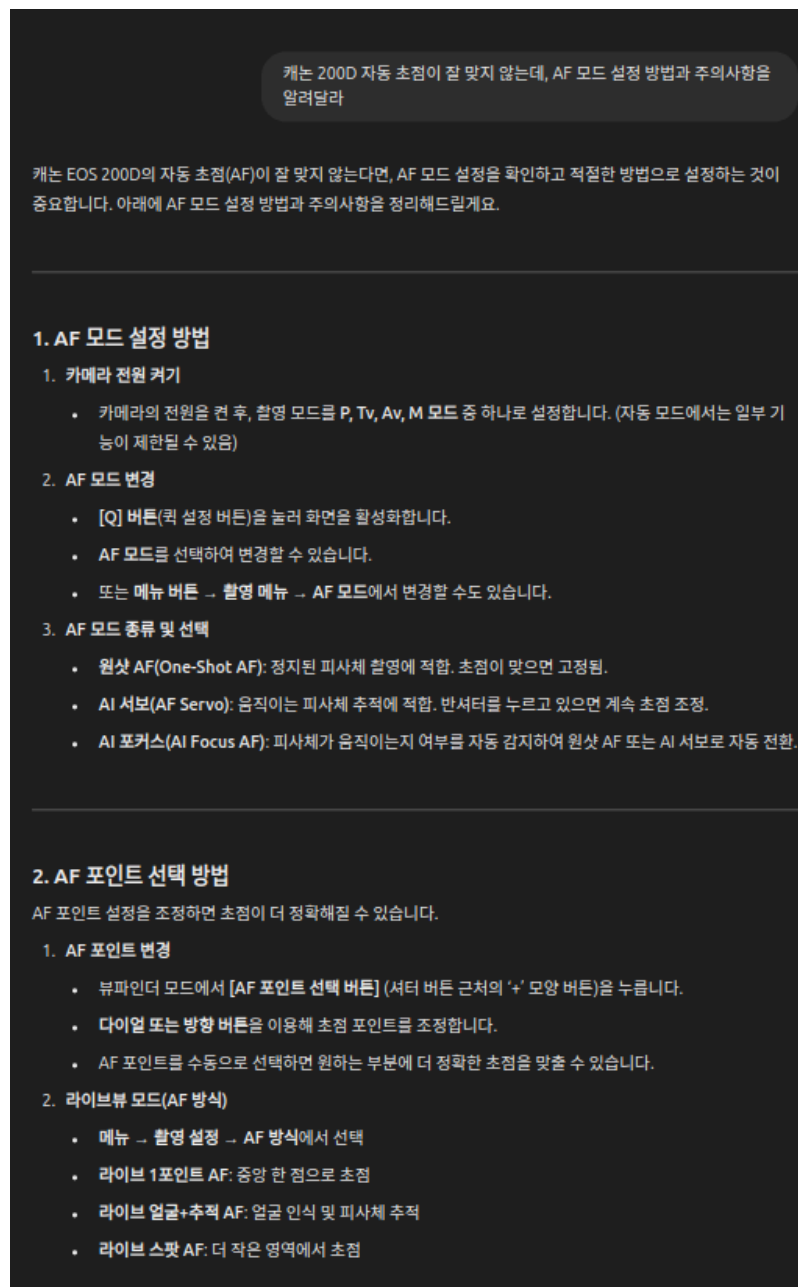
▶ 상황 예시 및 답변 비교

2. Canon

2.1 사용자 질문 (예시)

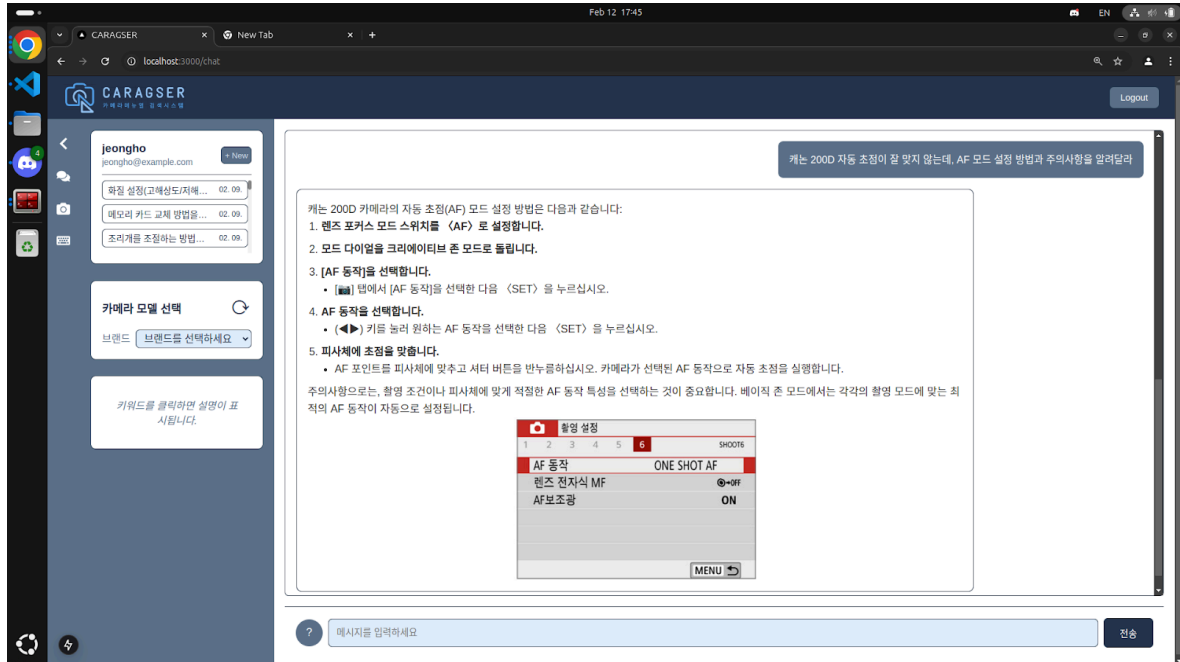
- 질문: "캐논 카메라에서 자동 초점이 잘 맞지 않는데, AF 모드 설정 방법과 주의사항이 궁금해요."

2.2 Chat GPT 응답 스크린샷



▶ 상황 예시 및 답변 비교

2.3 RAG system 응답 스크린샷



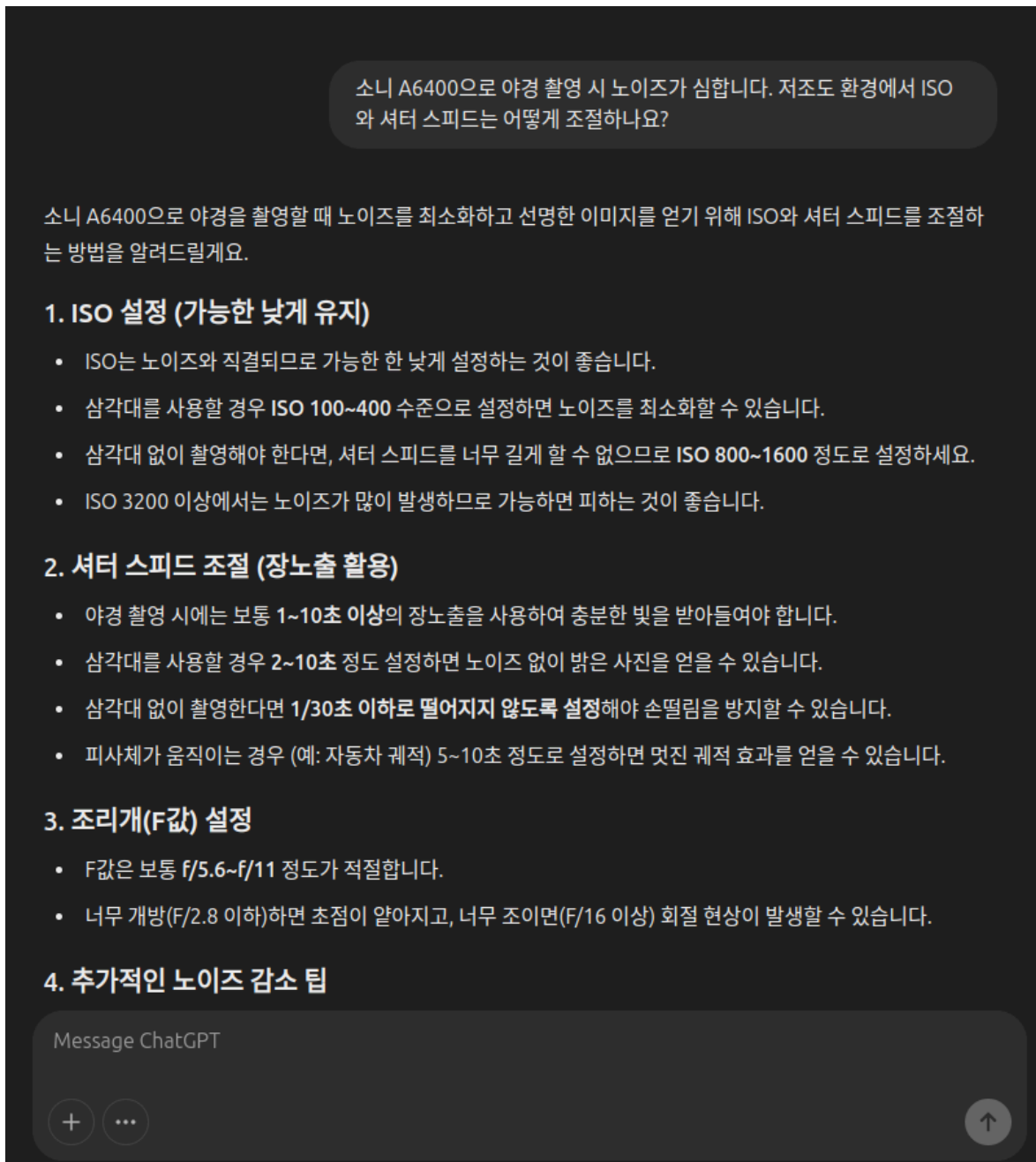
▶ 상황 예시 및 답변 비교

3. Sony

3.1 사용자 질문 (예시)

- 질문: "소니 A7 시리즈로 야경 촬영 시 노이즈가 심합니다. 저조도 환경에서 ISO와 셔터 스피드는 어떻게 조절하나요?"

3.2 Chat GPT 응답 스크린샷



▶ 상황 예시 및 답변 비교

3.3 RAG system 응답 스크린샷

