

LLM 기반 검색·분석 교육 콘텐츠 기획서

작성자: 김병현

제출일: 2024. 12. 12

1. 교육 목적 (Learning Objectives)

본 교육 과정은 기업 내 방대한 문서를 효율적으로 활용하기 위한 LLM 기반 검색(Retrieval) 및 답변 생성(Generation) 기술을 습득하는 것을 목표로 합니다. 학습자는 단순한 개념 이해를 넘어, 실제 데이터셋을 활용해 RAG(Retrieval-Augmented Generation) 파이프라인을 구축하고 검색 성능을 정량적으로 평가 및 개선하는 실무 역량을 배양합니다.

2. 대상 학습자 (Target Audience)

- 사내 문서를 다루는 실무자: 기획, 운영, 개발 등 직군에 상관없이 LLM을 업무에 도입하고자 하는 분
- Python 초중급 사용자: Pandas 기초 문법 이해도가 있으며, 코드를 통해 데이터 처리가 가능한 분
- RAG 입문자: LLM에 관심은 많으나 실제 파이프라인 구축 및 성능 평가 경험이 부족한 분

3. 강의 구성 (Curriculum)

이론과 실습을 병행하여 단계별로 RAG 시스템을 고도화하는 3단계 구성입니다.

단계	주제	주요 학습 내용
Part 1	RAG 핵심 개념 및 구축	<ul style="list-style-type: none">키워드 검색(BM25) vs 의미 검색(Vector) vs 하이브리드 검색 비교RAG 파이프라인 이해 (수집 → 청킹 → 임베딩 → 검색 → 답변)
Part 2	검색 품질 평가 (Evaluation)	<ul style="list-style-type: none">정량적 평가 지표 학습 (Hit@k, MRR 등)검색 기법별 4단계 성능 비교 및 시각화
Part 3	고급 기법 및 최적화	<ul style="list-style-type: none">Query Expansion: LLM을 활용한 질의 확장을 통해 검색 범위 확대Generation: 검색 결과를 Context로 활용하여 할루시네이션 최소화

4. 실습 산출물 (Key Deliverables)

수강생은 강의를 통해 아래와 같은 구체적인 결과물을 직접 제작하게 됩니다.

- 실무형 데이터셋 활용: AGORA (글로벌 AI 거버넌스 문서) 1,500+개 처리 경험 (Kaggle)
- 검색 성능 리포트: 4가지 검색 기법(BM25, Vector, Hybrid, Hybrid+QE)에 대한 Hit@1, MRR 비교 분석 결과
- 완성된 RAG 파이프라인: 사용자 질문에 대해 근거 문서(Reference)와 함께 정확한 답변을 생성하는 Python 노트북 코드

5. 필요 환경 (Prerequisites)

원활한 실습 진행을 위해 다음 환경을 권장합니다.

- 실습 환경: Google Colab (T4 GPU 이상 권장)
- 언어 및 도구: Python 3.x, Jupyter Notebook
- 필수 라이브러리: openai, sentence-transformers, rank-bm25, faiss-cpu, pandas
- API Key: Upstage Solar API Key, Kaggle API token (설정 필요)