**📄 수능 수학 문제 생성 시스템 명세서**

**1. 개요**

본 시스템은 한국 대학수학능력시험(수능) 스타일의 수학 문제를 자동 생성하고, 풀이와 해설까지 제공하는 **AI 문제 생성 파이프라인**이다.

* **문제 출제**: Gemini-2.5-Pro
* **풀이/증명**: Gemma-3
* **검색/참고자료 임베딩**: EmbeddingGemma

**2. 목표**

* 실제 수능과 유사한 형식의 문제를 생성 (선택형/단답형)
* 난이도, 주제, 배점 등을 설정할 수 있음
* 정답 + 풀이 + 해설까지 포함한 완결된 문제 제공
* 기출/교재 데이터를 기반으로 한 신뢰도 높은 문제 출제

**3. 시스템 아키텍처**

**3.1 아키텍처 개요**

사용자 요청

│

▼

[Retriever]

EmbeddingGemma + FAISS/HNSW

└─ 기출/교재 데이터 색인 후 유사 문항 검색

│

▼

[Generator]

Gemini-2.5-Pro

└─ 검색 문항 기반 신규 문제 출제

│

▼

[Solver]

Gemma-3 (12B/27B)

└─ 생성된 문제 풀이 및 해설 작성

│

▼

[Output Layer]

Streamlit/WebUI

└─ 문제 + 정답 + 해설 출력 및 PDF 저장

**4. 주요 모듈**

**4.1 데이터베이스 모듈**

* **입력 형식**: JSONL ({"id": "...", "text": "..."})
* **검색 방식**: EmbeddingGemma → 벡터화 → FAISS 인덱스 구축
* **지원 기능**:
  + Top-k 검색
  + 문단/청크 단위 색인
  + L2 또는 Cosine 유사도 기반

**4.2 문제 생성 모듈 (Gemini-2.5-Pro)**

* **입력**: 검색된 기출/참고자료 + 사용자 조건(주제, 난이도, 배점, 형식)
* **출력**:
  + 문제 지문
  + 보기(선택형인 경우)
  + 정답

**4.3 풀이/증명 모듈 (Gemma-3)**

* **입력**: 문제 텍스트 + 정답
* **출력**:
  + 단계별 풀이 과정
  + 증명 및 수학적 설명
  + 최종 해설

**4.4 UI 모듈 (Streamlit)**

* 기능:
  + 단일 문제 생성
  + 모의고사 세트 생성
  + 결과 PDF 다운로드 (LaTeX/ReportLab 활용)
  + 난이도, 유형, 주제 선택 인터페이스

**5. 데이터 흐름**

1. **사용자 입력**: 문제 생성 요청 (예: “미적분, 상 난이도, 선택형”)
2. **Retriever**: EmbeddingGemma로 쿼리와 코퍼스 문항 임베딩 → FAISS 검색
3. **Generator**: Gemini-2.5-Pro가 검색된 문항을 변형·조합해 신규 문제 출제
4. **Solver**: Gemma-3가 문제를 풀이하고 해설 생성
5. **Output**: Streamlit UI에 문제/정답/해설 표시 + PDF 다운로드

**6. 기술 스택**

* **임베딩**: EmbeddingGemma (HuggingFace Transformers, PyTorch)
* **생성 모델**: Gemini-2.5-Pro (Google Generative AI API)
* **풀이 모델**: Gemma-3 (HuggingFace Transformers, PyTorch)
* **DB/검색**: FAISS (CPU/GPU 버전 선택)
* **웹/UI**: Streamlit
* **PDF 출력**: ReportLab, LaTeX

**7. 성능 및 최적화**

* **임베딩 차원**: 768 (MRL 사용 시 128/256/512로 축소 가능)
* **검색 속도**: 약 100k 문항 색인 시 < 100ms (FAISS HNSW)
* **문제 생성 속도**: Gemini API 호출 → 평균 2~5초
* **풀이 생성 속도**: Gemma-3 12B 기준 GPU 환경에서 수 초~수십 초

**8. 확장 계획**

* **재랭킹 모듈 추가**: Gemini 또는 Gemma를 활용한 LLM-as-Reranker
* **폴리싱 모듈**: Claude/LLM 활용하여 지문을 더 자연스럽게 다듬기
* **Fine-tuning**: Gemma-3에 수능 전용 파인튜닝 적용 (문항 풀이 안정성 강화)
* **사용자 데이터 반영**: 학습 이력 기반 맞춤형 문제 추천