

# Solar Teacher

LLM 기반 백지 공부 피드백 및 맞춤 문제 생성 서비스

Recsys-04 | 업스테이지(Upstage)

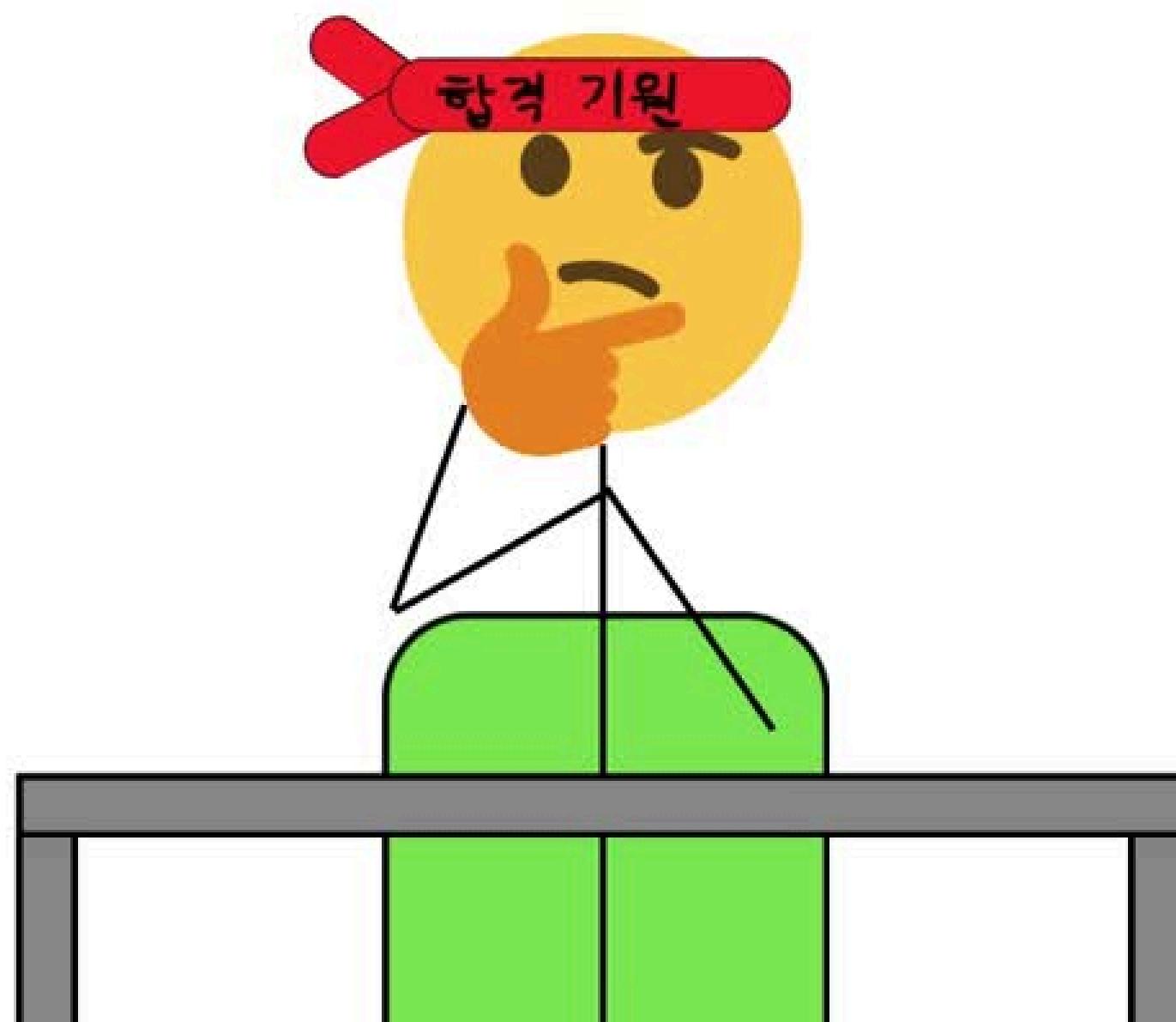
추천닭갈비 | 강현구, 서동준, 양시영, 이도걸, 이수미

[GitHub](#)

**boostcamp** ai tech

**Solar Teacher**  
SINCE 2025

# 이게 뭐였더라...??



# Contents

**01**

팀 설명

**02**

프로젝트 방향

**03**

프로젝트 소개

**04**

프로젝트 상세

**05**

최종 결과

**06**

자체 평가

**07**

appendix

---

01

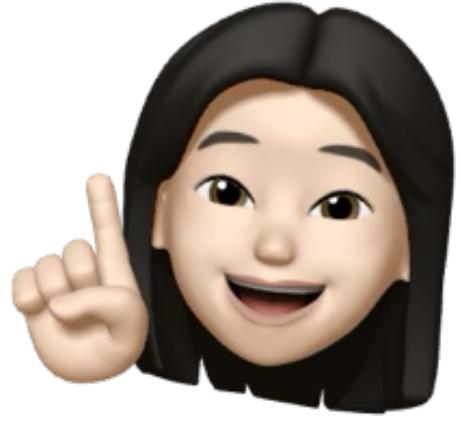
## 팀 설명

# 팀원 소개

컴퓨터공학과 통계학의 전문 지식, 그리고 실무 경험을 두루 갖춘 균형 잡힌 팀으로서,  
다각도의 접근을 통해 복잡한 문제를 효과적으로 해결할 수 있습니다



강현구



서동준



양시영



이도걸



이수미

Feedback  
Backend

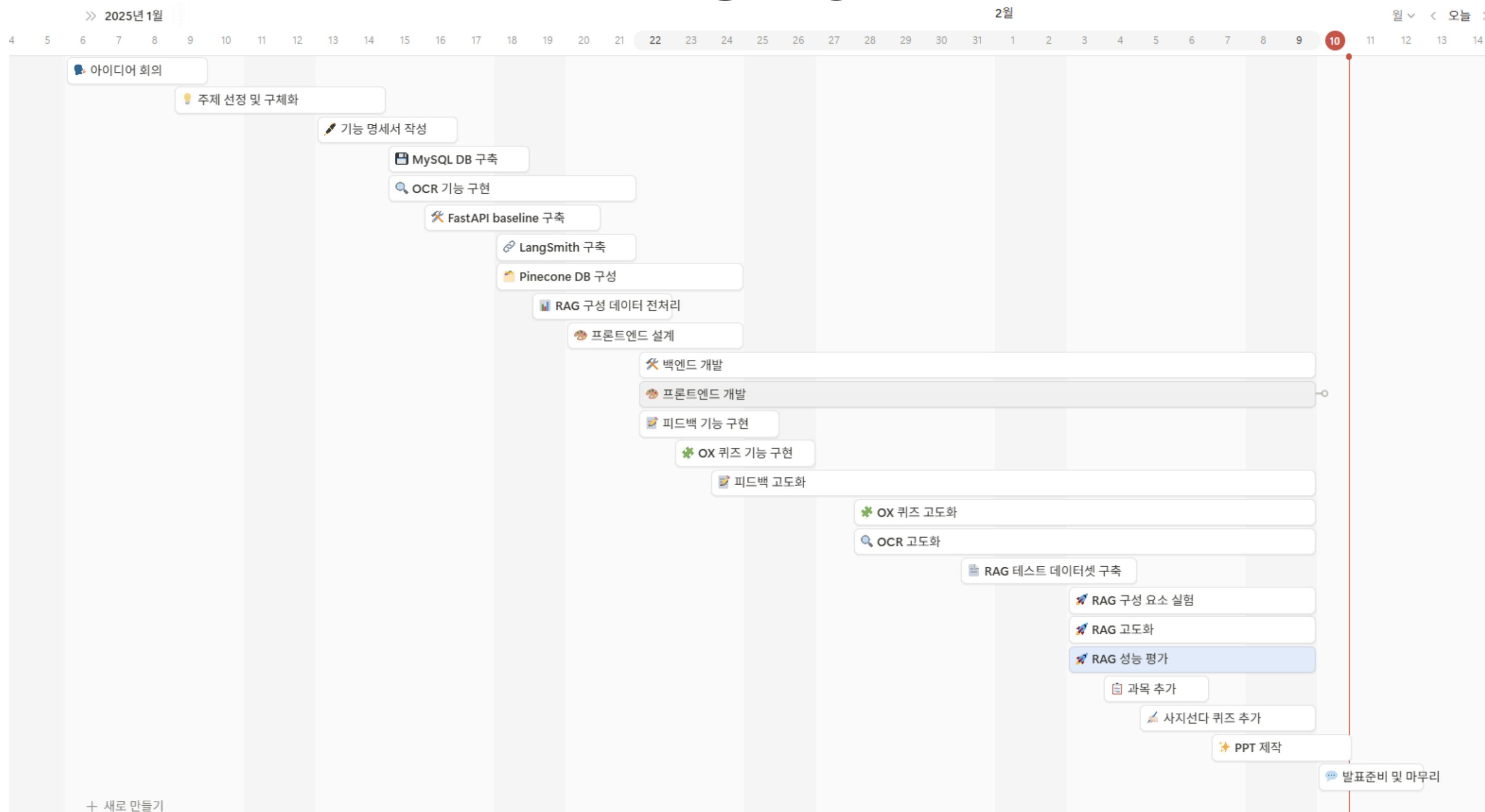
RAG  
Pinecone

RAG  
LangSmith

PM  
OCR

Quiz  
MySQL DB

# 진행 일정



---

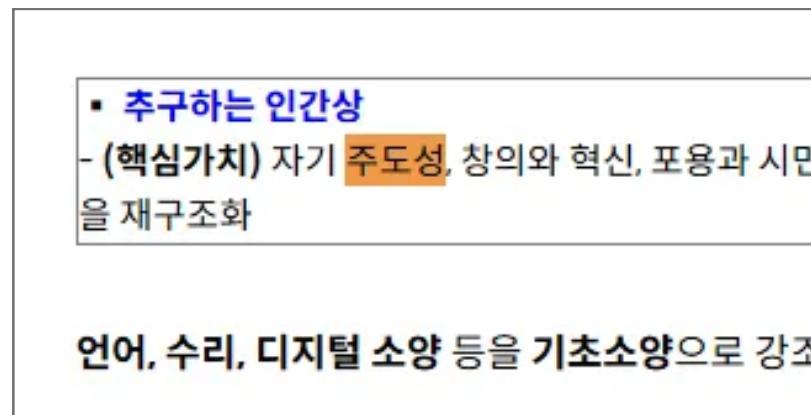
02

## 주제 선정 배경

# 시장성

## 01

### 자기주도 학습의 중요성 증가

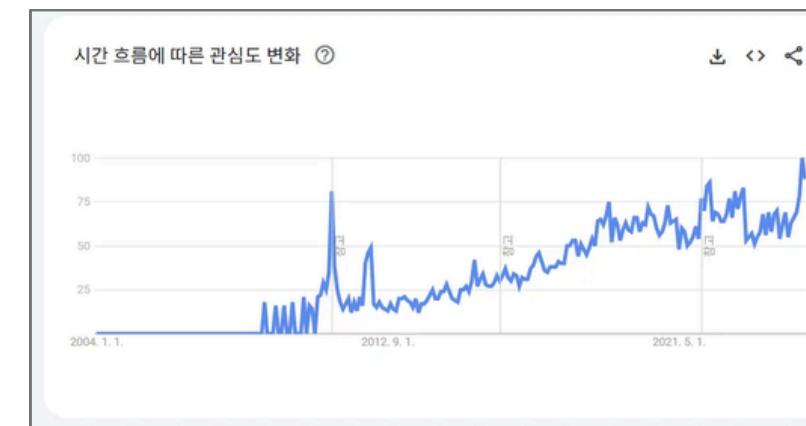


교육부의 학습자 중심 교육 강조

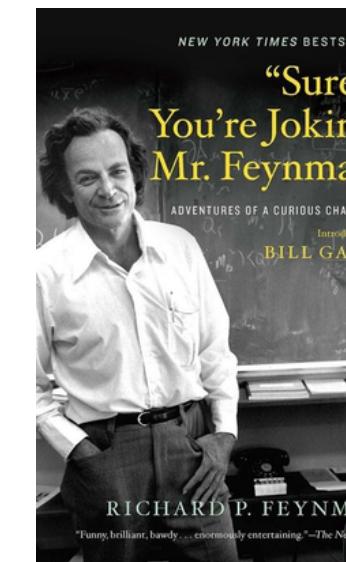


## 02

### 공부 방법에 대한 관심도 증가



‘공부법’에 대한 검색량 증가

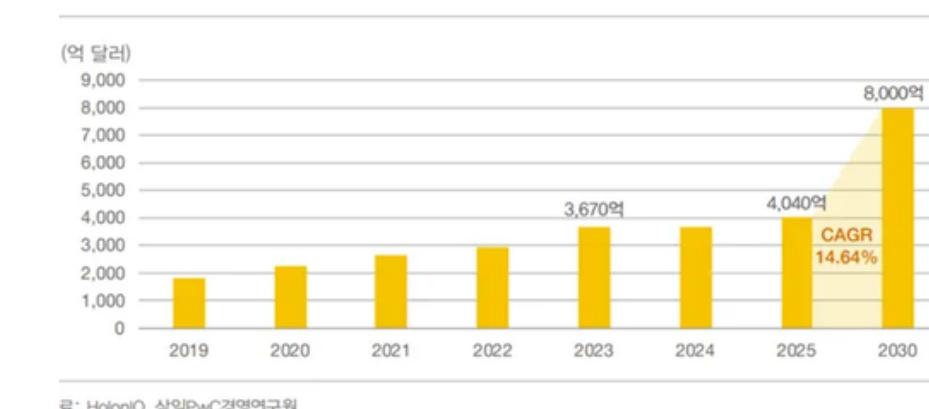


**Feynman Technique**  
어려운 개념을 쉽게  
설명하며 학습하는 방식  
→ 이미 검증된 효과적인  
자기주도 학습법  
직접 개념을 정리하고  
설명하는 과정이  
백지공부법과 유사

자기주도 학습 및 효율적인 학습 방법에 대한 관심 증가

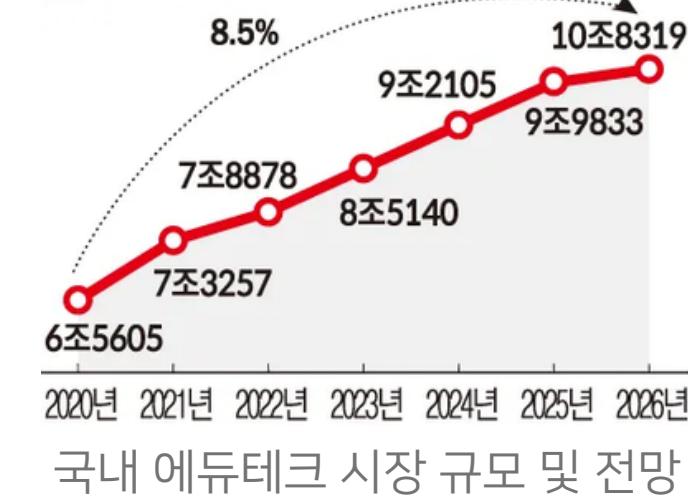
## 03

### 성장하는 에듀테크 시장



글로벌 에듀테크 시장 규모 및 전망

### 국내 에듀테크 시장 규모 및 전망



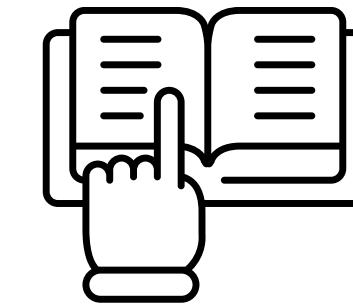
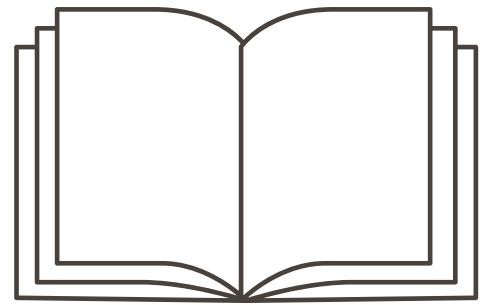
# 백지공부법을 아시나요?

# 백지 공부법

백지에 핵심 내용을 스스로 떠올려 정리하는 학습 방법



순서



범위와 목표를 설정한 후 공부

아무것도 쓰여지지 않은  
백지의 종이나 공책에  
공부한 내용을 써내려가기

공부 자료를 참고하여 작성내용을 확인하고 잘못된 부분을 체크하기

## 백지 공부법의 장점



- 메타인지 향상
  - 잘 기억하지 못해서 작성하지 못하거나 틀린 부분에 대해서 명확하게 인식 가능
- 효율적인 복습 시간
  - 처음에는 시간이 걸리지만 다음 복습에는 시간을 절약하고 효율적으로 학습 가능

- 백지 공부법의 장점을 간접적으로 나타내는 논문들

메타인지 촉진 수업이 대학생의 학업성취도, 메타인지, 내적동기 및 자아효능감에 미치는 효과

김혜진<sup>1)</sup>

The Effects of Meta Cognitive Strategy Instruction on Academic Achievement, Self-Regulation, and Metacognition in University Students

frontiers in  
PSYCHOLOGY

MINI REVIEW ARTICLE  
published: 04 April 2014  
doi: 10.3389/fpsyg.2014.00286

Retrieval practice enhances new learning: the forward effect of testing

Bernhard Pastötter\* and Karl-Heinz T. Bäuml

Department of Experimental Psychology, Regensburg University, Regensburg, Germany

Edited by:  
Christoph T. Weidemann, Swansea University, UK  
Reviewed by:  
Karl Srovnar, Harvard University

In the last couple of years, there has been a dramatic increase in laboratory research examining the benefits of recall testing on long-term learning and retention. This work was largely on the backward effect of testing, which shows that retrieval practice on previously studied information, compared to restudy of the same material, renders the

beyond this prominent  
there is also a forward  
died information can  
we provide a review

## The Power of Testing Memory

Basic Research and Implications for Educational Practice

Henry L. Roediger, III, and Jeffrey D. Karpicke

Washington University in St. Louis

## 백지 공부법의 장점



- 메타인지 향상
  - 잘 기억하지 못해서 작성하지 못하거나 틀린 부분에 대해서 명확하게 인식 가능
- 효율적인 복습 시간
  - 처음에는 시간이 걸리지만 다음 복습에는 시간을 절약하고 효율적으로 학습 가능

## 백지 공부법의 단점



- 잘못된 정보 학습 가능성
  - 스스로 기억을 떠올려서 쓰다 보면 잘못된 개념을 적을 수도
  - 이를 수정하지 않으면 오개념이 고착될 위험 존재
- 시간이 많이 걸릴 수 있음
  - 정리된 자료를 보는 것보다 오래 걸릴 가능성
  - 처음에는 비효율적으로 보일 수도



# 백지 공부법의 단점을 극복하고 효과를 극대화하는 학습 보조 도구 개발

AI가 개념 오류를 바로잡아 정확한 학습을 지원  
자동 퀴즈 제공으로 개념을 더욱 탄탄하게 정리

---

03

## 프로젝트 소개

부덕아 너 시험공부 다 했어??

## 주요 기능 요약

다음과 같은 주요 세 가지 기능을 가지고 있습니다.

**O1**

AI 피드백

**O2**

OX 퀴즈

**O3**

객관식 퀴즈

# 화이트보드&업로드

화이트보드로 자유롭게 필기

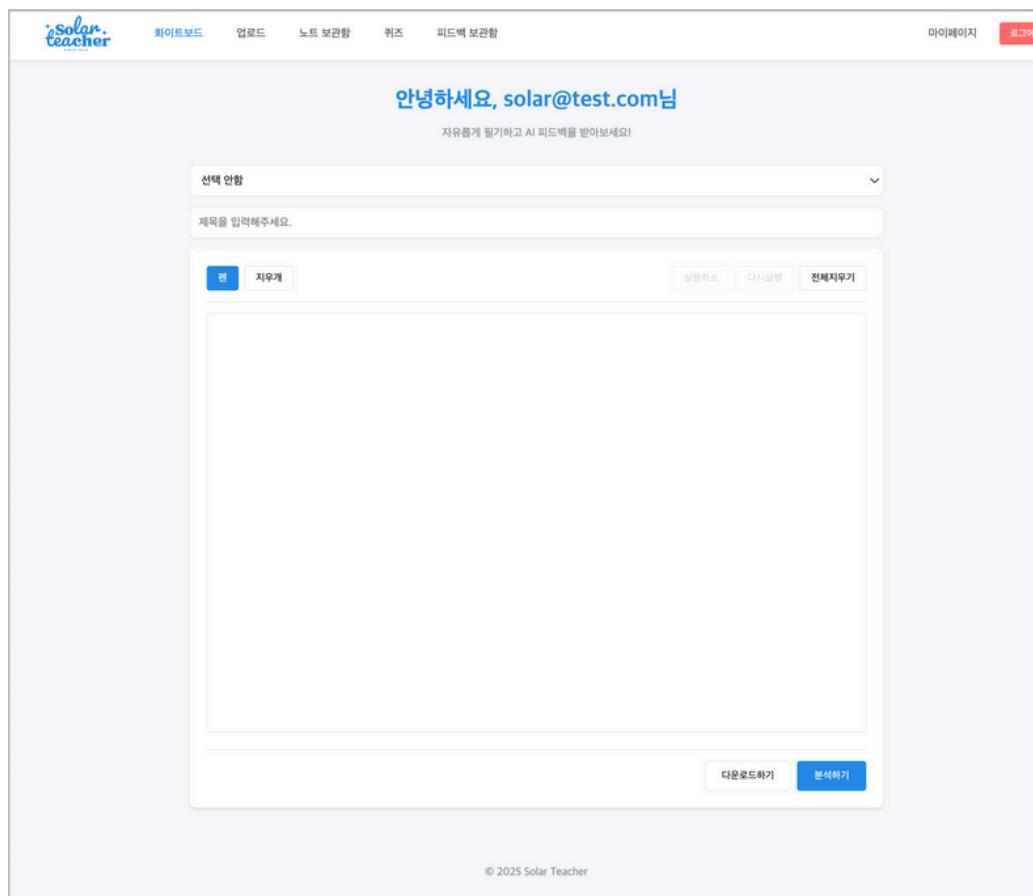
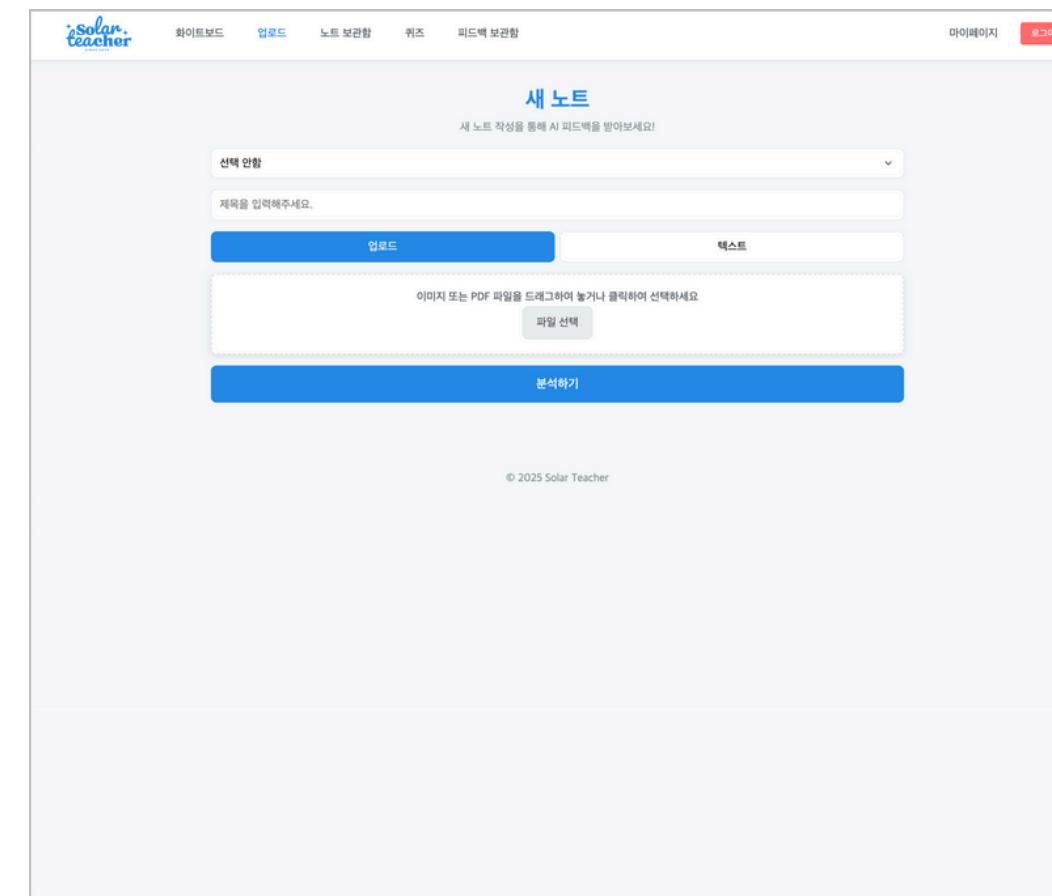
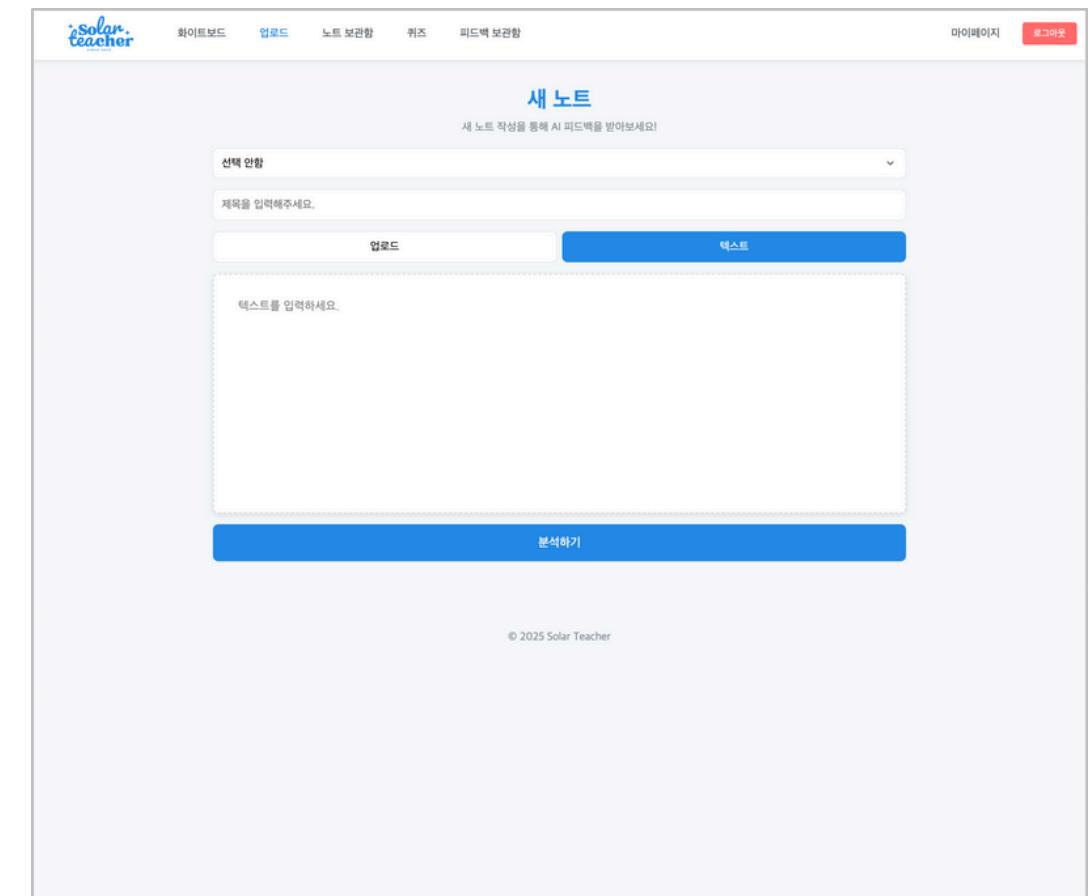


사진 또는 PDF 업로드



텍스트 입력



# 피드백 페이지

과목을 선택하지  
않았을 때  
자동 분류

사용자가  
입력한 노트

화이트보드 업로드 노트 보관함 퀴즈 피드백 보관함 마이페이지 로그아웃

과목: 생명과학 II

생명체의 구성 물질

작성 날짜: 2025년 02월 09일

**피드백**

- 잘못된 부분: 단당류(녹말, 글리코겐) 이당류(설탕), 다당류(포도당)  
 수정 사항: 단당류(포도당, 과당, 갈락토스 등), 이당류(엿당, 설탕, 젓당 등), 다당류(셀룰로스, 녹말, 글리코겐 등)
- 단당류, 이당류, 다당류의 예시가 잘못 기재되었습니다.  
 출처: 천재교육\_고등교과서\_생명과학II\_이준규(15개정)\_교과서 본문.pdf, 44쪽

**OX 퀴즈**

생명체의 구성 물질 중 가장 많은 비율을 차지하는 것은 물이다.

O X

노트에 대한 피드백

## 피드백

1

잘못된 부분: 단당류(녹말, 글리코겐) 이당류(설탕), 다당류(포도당)

수정 사항: 단당류(포도당, 과당, 갈락토스 등), 이당류(엿당, 설탕, 젓당 등), 다당류(셀룰로스, 녹말, 글리코겐 등)

단당류, 이당류, 다당류의 예시가 잘못 기재되었습니다.

출처: 천재교육\_고등교과서\_생명과학II\_이준규(15개정)\_교과서 본문.pdf, 44쪽

관련 내용 OX 퀴즈

## OX 퀴즈

생명체의 구성 물질 중 가장 많은 비율을 차지하는 것은 물이다.

O

X

### 3.4 객관식 퀴즈

# 객관식 퀴즈(사지선다)

## 1. 퀴즈 탭

작성한 내용을 바탕으로 퀴즈를 풀어보세요!

객관식 퀴즈 풀기

이전 퀴즈 기록

객관식 퀴즈 기록

OX 퀴즈 기록

© 2025 Solar Teacher

## 3. 문제 풀이 & 정답 체크 & 해설

다음 중 탄수화물의 종류가 아닌 것은?

단당류

이당류

다당류

인지질

1 / 5

다음 중 탄수화물의 종류가 아닌 것은?

단당류

이당류

다당류

인지질

틀렸습니다

인지질은 지질의 종류 중 하나이다.

다음 문제

1 / 5

## 2. 과목 선택

과목을 선택하세요

생명과학 I

지구과학 II

정보처리기사

정치외법

물리학 II

© 2025 Solar Teacher

# 이전 퀴즈 기록

## 사지선다 퀴즈 내역 확인

**퀴즈 기록**

전체 문제	정답률
6개	67%

형법이 보호하는 법익에 대한 설명으로 옳은 것은?

형법은 개인의 생명, 신체, 명예, 자유, 재산 등을 보호한다.

형법은 개인의 생명, 신체, 명예, 자유, 재산 등을 보호하며, 공공의 안전, 신용 등도 중요하게 다룬다.

정답입니다!

정답: 2번

형법은 개인의 생명, 신체, 명예, 자유, 재산 등을 보호하며, 공공의 안전, 신용 등도 중요하게 다룬다.

2025-02-09 05:04:03

다음 중 단수화물의 종류가 아닌 것은?

단당류

이당류

다당류

인지질

## OX 퀴즈 내역 확인

**OX 퀴즈 기록**

전체 문제	정답률
5개	20%

조석 현상은 지구의 물질에 작용하는 달과 태양의 인력에 의해 일어나는데, 조석을 일으키는 힘을 무엇이라고 하는가?

내가 체크한 답: ○

정답: ✕

조석 현상을 일으키는 힘은 기조력이라고 한다.

2025년 2월 9일 오전 04:53

한반도를 이루는 암석은 지난 30억 년 동안 여러 번의 지각 변동과 화성 작용을 겪었다.

내가 체크한 답: ✕

정답: ○

한반도를 이루는 암석은 지난 30억 년 동안 여러 번의 지각 변동과 화성 작용과 변성 작용을 겪었다.

2025년 2월 9일 오전 04:53

조석 현상의 영향으로 바닷가의 해수면은 주기적으로 상승과 하강을 반복한다.

내가 체크한 답: ○

정답: ○

조석 현상의 영향으로 바닷가의 해수면은 주기적으로 상승과 하강을 반복한다.

# 피드백 모아보기

**화이트보드** 업로드 노트 보관함 퀴즈 피드백 보관함 마이페이지 로그아웃

**피드백 보관함**  
지금까지 받은 피드백을 모아보세요!

**암석과 광물의 형성과정** 지구과학 II  
2025년 02월 09일

1 잘못된 부분: 느린 냉각·큰 결정 ( 뭐더라?)  
수정 사항: 느린 냉각·큰 결정 ( 심성암)  
💡 마그마가 천천히 냉각되면 큰 결정의 심성암이 형성됩니다.

**형법** 정치와법  
2025년 02월 09일

1 잘못된 부분: 귀형 법정극의  
수정 사항: 죄형 법정주의  
💡 죄형 법정주의의 올바른 표현입니다.

2 잘못된 부분: 공권력 ; 국가나 공공단체가 우월한 의사의 주체로서  
수정 사항: 공권력 ; 국가나 공공단체가 우월한 의사의 주체로서 국민에 대하여 명령하거나 강제하는 권리  
💡 공권력의 올바른 정의입니다.

3 잘못된 부분: 국가가 범리를 정하고 형벌을 부과하는 효과  
수정 사항: 국가가 범죄를 정하고 형벌을 부과하는 효과  
💡 범죄를 정하고 형벌을 부과하는 것이 올바른 표현입니다.

4 잘못된 부분: 응보로  
수정 사항: 응보론  
💡 응보론은 응보의 원칙을 의미합니다.

**과목 필터링 및 정렬**

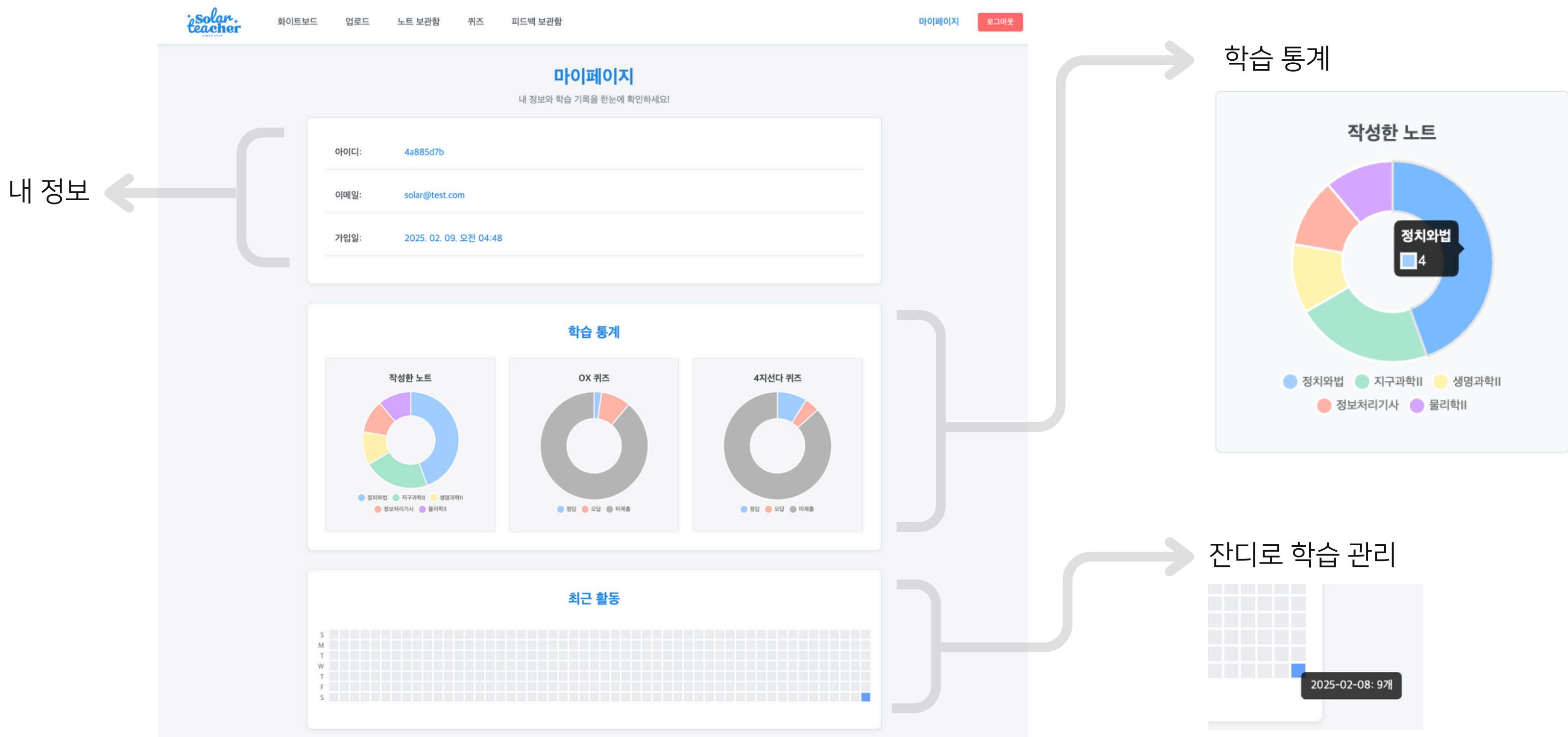
**피드백 보관함**  
지금까지 받은 피드백을 모아보세요!

✓ 전체 과목  
생명과학 II  
지구과학 II  
정보처리기사  
정치와법  
물리학 II

최신순

지금까지 받았던 피드백 모아보기

# 마이페이지



---

# 04

## 프로젝트 상세

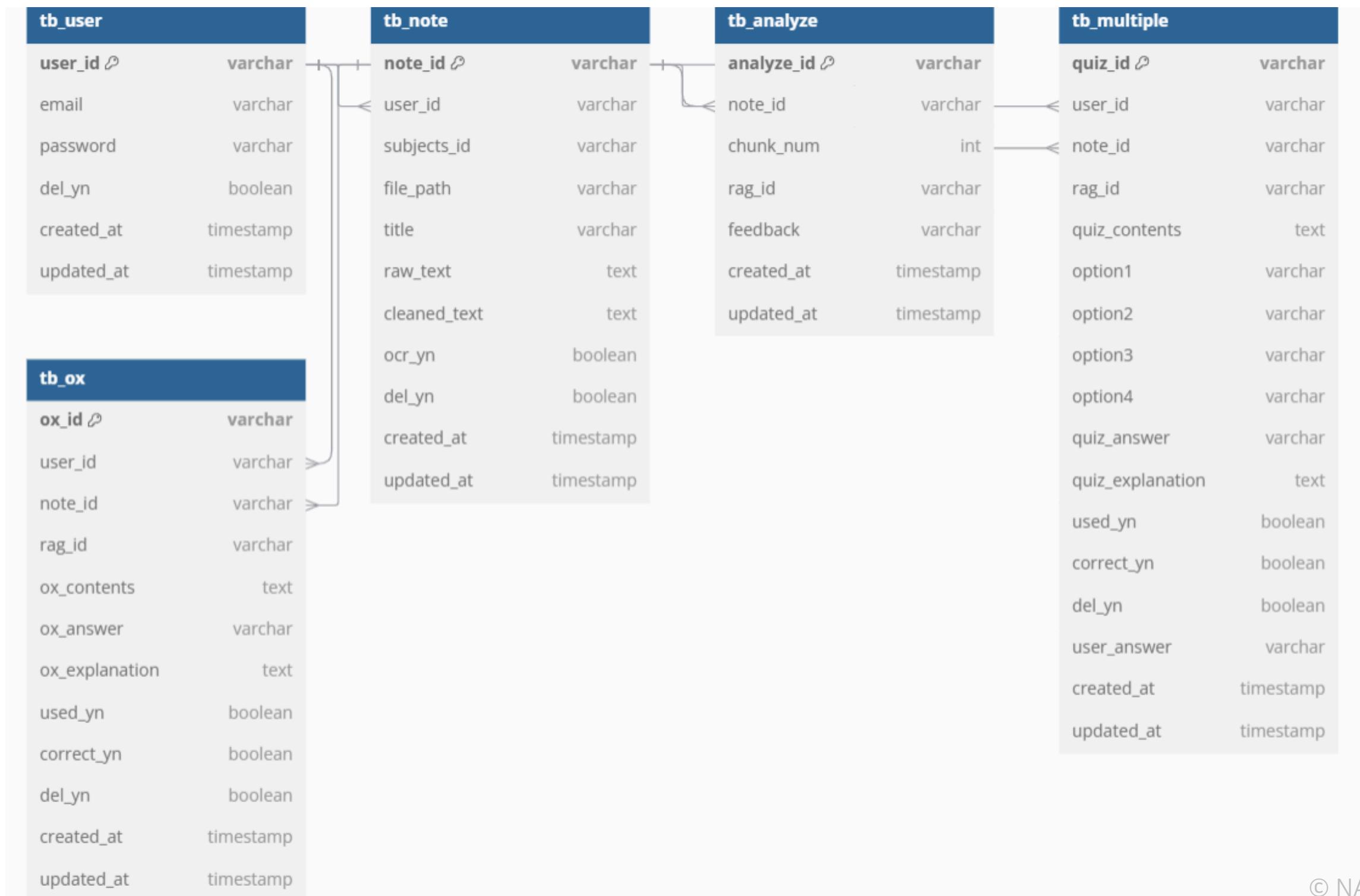
# 데이터



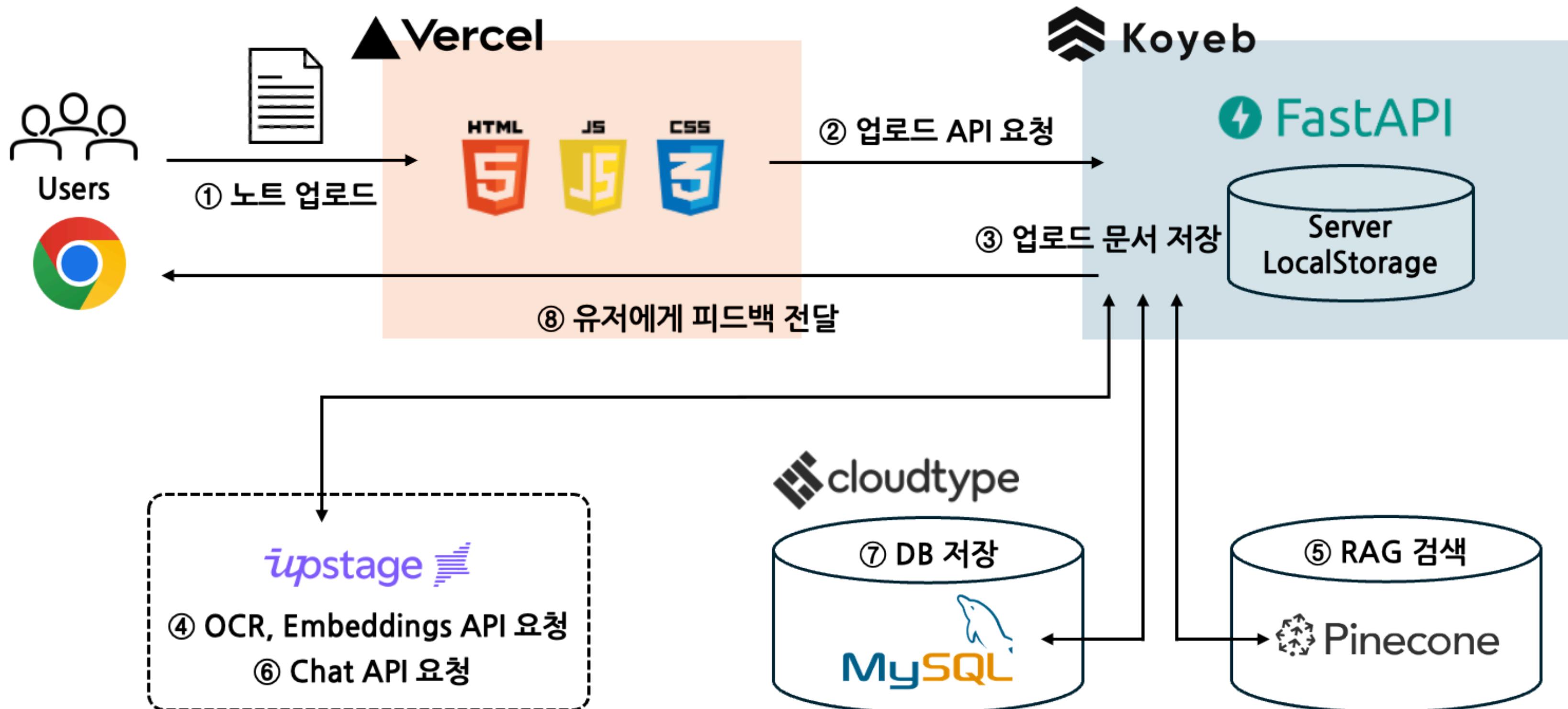
- 15개 과목 PDF 파일  
(지구과학2, 물리학2, 사회 문화, 정보처리기사 필기 등)
  - 교과서, 수능 완성, 개념 정리

# ERD

## MySQL로 데이터베이스 구축

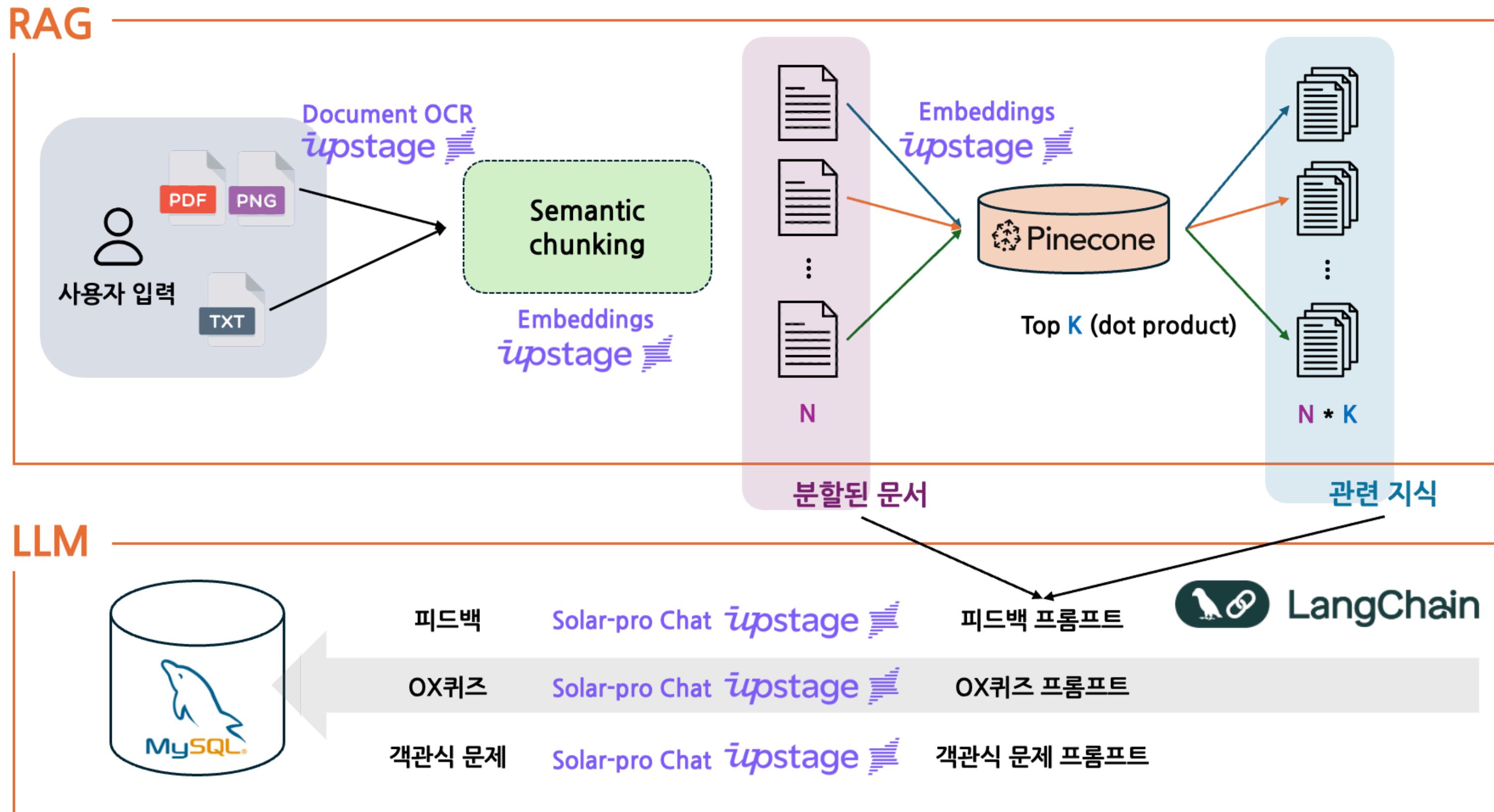


# 서비스 아키텍처



# 프로젝트 구조도

\* Monitoring  LangSmith



# OCR

## 입력 된 파일에 대해 전처리 후 성능 테스트 진행

Table 1: Accuracy and processing time comparison of different image and PDF models.

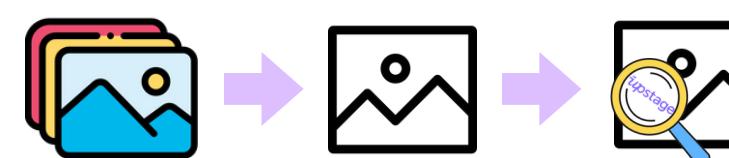
Method	Image		PDF	
	Accuracy (%)	Time (ms)	Accuracy (%)	Time (ms)
Base	99.06	1847.21	97.56	1413.22
Gray	99.06	1620.86	95.16	1307.12
Quality	98.74	1626.69	86.91	1333.27
Gray + Quality	<b>99.07</b>	<b>1554.02</b>	86.49	<b>1295.27</b>
LLM	98.28	9616.40	<b>98.25</b>	8690.97

### 전처리 목록

- Base: 기존 상태
- Gray: Gray Scale 변환
- Quality: 압축
- Gray+Quality: Gray Scale  
변환 후 압축
- LLM: OCR 이후 LLM 후처리

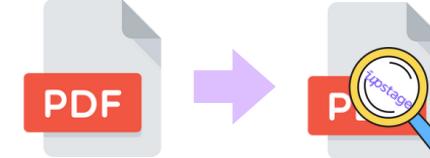
### <Image>

Gray Scale 변환,  
50% 압축 한 것이  
성능, 처리 속도에서 우세



### <PDF>

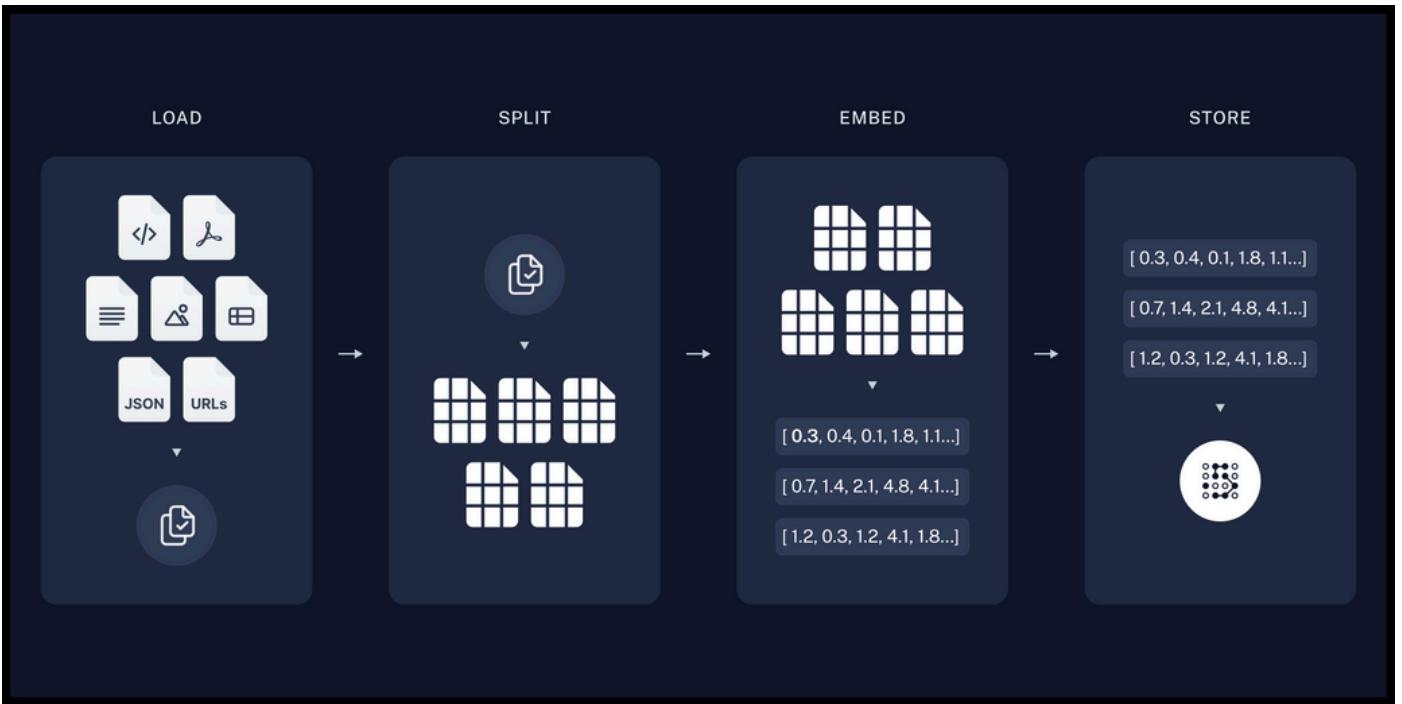
입력 된 파일 그대로  
OCR API에 전달  
하는 것이 가장 적합



# RAG 구성 요소

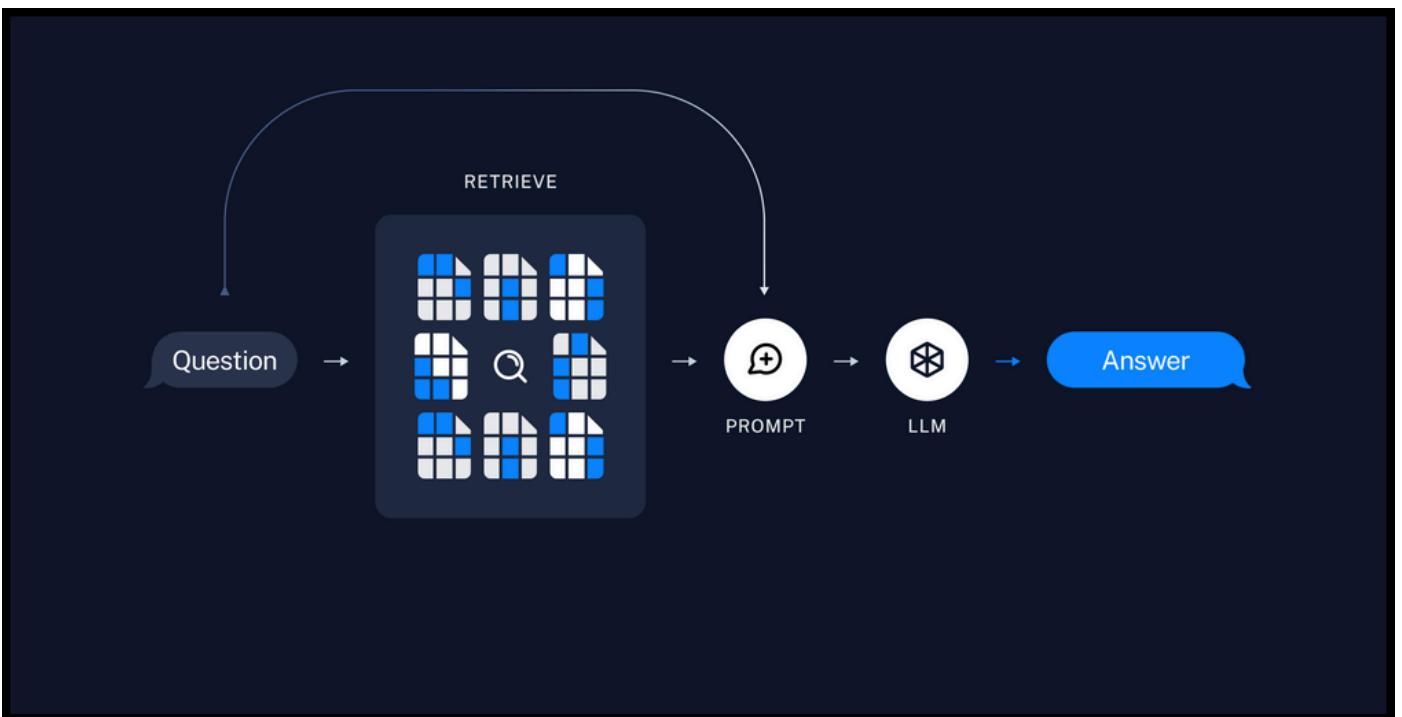
## O1 문서 로드 및 벡터 DB 저장

1. 문서 로드(Load): PyMuPDFLoader
2. 분할(Split): RecursiveCharacterTextSplitter
3. 임베딩(Embedding): Upstage Embeddings
4. 벡터 DB(VectorStore): Pinecone



## O2 문서 검색 및 결과 도출

1. 검색(Retrieval): 벡터스토어 기반 검색기(VectorStore-backed Retriever)
  - 검색한 chunk의 메타 데이터를 통해 과목 자동 분류
2. 프롬프트(Prompt)
3. 모델(LLM): Upstage Chat(Solar-pro)
4. 결과(Output): StrOutputParser



# RAG 사용 효과

## O1 피드백/퀴즈 문구 비교

- RAG를 활용하면 피드백의 정확성, 신뢰도, 교과서적 표현 반영이 향상
- 부적절한 정보 생성 방지: 검증된 소스만을 사용하여 허위정보 최소화

RAG 사용 X (LLM 단독 피드백)	RAG 사용 O
피드백이 상대적으로 일반적인 설명에 그침	피드백이 정교하고 정확함
개념을 수정하지만, 근거가 부족하거나 애매한 경우가 있음	개념을 정확하게 지적하며 고차원적인 표현을 반영함
표현이 정확하지 않고 부정확한 피드백이 생성될 수도 있음	체계적인 절차와 정답 내용이 포함됨

## O2 기존 LLM 대비 성능

- 검증된 데이터 기반의 응답 생성
- 기존 LLM에 학습되지 않은 개념 보충

Table 1: Accuracy and inference time comparison of different models on various datasets. (chunk size: 250, k: 1)

Dataset	LLM		LLM with RAG		LLM with RAG+Pre		Improvement with RAG-pre (%p)
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)	
물리학II	74.72	241.54	84.27	247.18	<b>90.45</b>	216.34	+15.73
화학II	77.55	258.86	80.00	262.30	<b>83.59</b>	250.02	+6.04
생명과학II	82.07	189.64	<b>90.28</b>	195.83	88.28	173.81	+6.21
지구과학II	79.31	227.92	84.57	231.00	<b>85.71</b>	208.53	+6.40
정치와 법	89.47	148.98	84.96	160.06	<b>92.04</b>	138.72	+2.57
정보처리기사	73.31	303.89	<b>78.72</b>	321.80	78.30	281.49	+4.99

# 벡터 DB



- 고성능과 확장성을 갖춘 클라우드 기반 벡터 데이터베이스
- AI 및 머신러닝 애플리케이션을 위한 효율적인 벡터 저장 및 검색 솔루션

Pinecone 화면 캡쳐

데이터(교과서, 수능완성, 개념정리) 전처리

- 페이지 필터링
- 불용어 제거
- 청크 분할
- 메타 데이터(과목, 페이지, 파일 이름 등) 추가



- Metric: dotproduct
- Dimension: 4,096
- 총 개수: 5,080

# LangSmith



# LangSmith

- LLM 애플리케이션 개발, 모니터링 및 테스트를 위한 플랫폼
- LLM의 실행 과정을 시각화하고 추적하며, 각 단계의 입출력을 상세히 보여주어 디버깅 용이

Personal > Tracing projects > dev-02

**dev-02**

Runs Threads Monitor Setup

1 filter Last 7 days Root Runs LLM Calls All Run

Name Input

RunnableSequence 헌법의 의의와 기본원리 요약

RunnableSequence 헌법의 의의와 기본원리 요약

RunnableSequence 헌법의 의의와 기본원리 요약

VectorStoreRetriever 6. 시민 참여형 헌법·아이...

VectorStoreRetriever 헌법의 의의와 기본원리 요...

RunnableSequence 애자일 개발 4가지 핵심가...

RunnableSequence 애자일 개발 4가지 핵심가...

VectorStoreRetriever 계획을 따르기 보다는 변화...

VectorStoreRetriever 애자일 개발 4가지 핵심가...

RunnableSequence 다 17-1

RunnableSequence 다 17-1

**TRACE** Waterfall Most relevant

**RunnableSequence** 19.92s 3,228

**ChatUpstage** solar-pro 19.92s

Some runs have been hidden. Show 2 hidden runs

**Input**

**Raw Text**

헌법의 의의와 기본원리 요약  
헌법의 의의 규정함.  
헌법은 국가의 최고법이며 기본법으로 국민의 기본권과 국가조직 및 구성을 관한 내용을.  
법체계에서 가장 하위에 위치하며 일반 법률보다 추상적이고 일반적인 내용을 다룬.  
근대 이후 헌법은 국민의 기본권을 보장하고 국가 권력을 제한하는 역할을 수행.  
현대 민주 국가에서는 사회적 기본권을 보장하여 모든 국민이 인간다운 생활을 할 수 있도록 함.  
2. 헌법의 기능  
정치적 기능: 국가 창설 및 정치 생활 주도 권력 행사와 한계를 규정  
법적 기능: 국가 법질서의 최고 규범으로 적용 법률이나 정책의 근거가 됨.  
사회적 기능: 사회 통합 실현 다원화된 이익 조정 및 갈등 해결,  
3. 각국 헌법 제1조 비교  
대한민국: 사회 공화국이며, 주권은 통치자에게 있음.  
프랑스: 비종교적·민주적·사회적 공화국으로, 평등 보장.  
미국: 입법권이 의회에 속함.  
독일: 천황이 국민 통합의 상징임.  
일본: 인간의 존엄성 보호가 국가의 의무임.  
4. 법치주의와 헌법  
법치주의는 자의적·폭력적 지배를 배제하고 국민 의사에 따른 법에 의한 통치를 의미.  
형식적 법치주의 (법률로 규정)뿐만 아니라 실질적 법치주의 (헌법 이념 부합)도 중요.  
5. '87년 헌법'과 민주주의

**START TIME**  
02/08/2025, 02:38:31 PM

**END TIME**  
02/08/2025, 02:38:51 PM

**TIME TO FIRST TOKEN**  
N/A

**STATUS**  
Success

**TOTAL TOKENS**  
3,228 tokens

**LATENCY**  
19.92s

**TYPE**  
Sequence

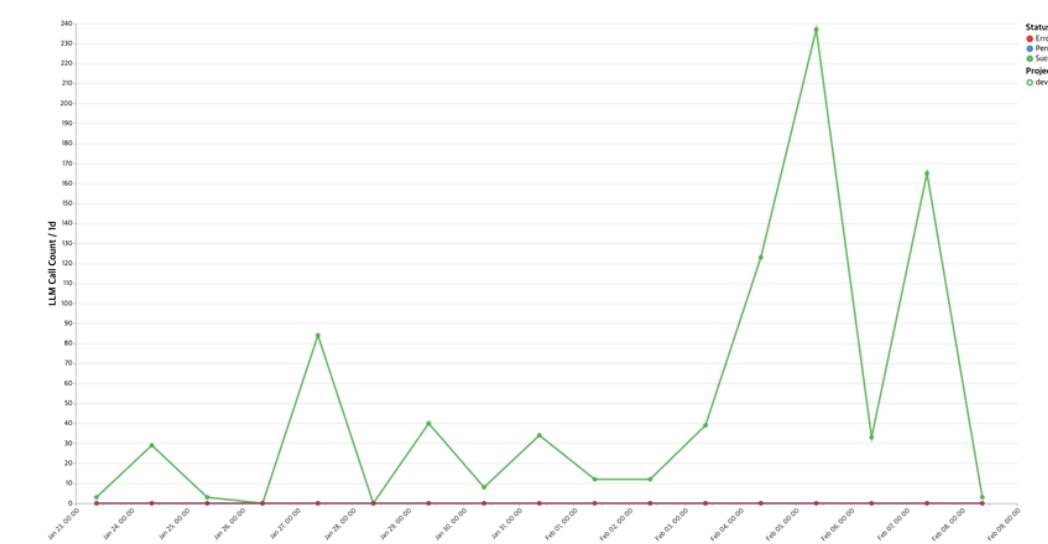
LangSmith 화면 캡처

**Observability**

Tracing Projects 9 Dashboards 0

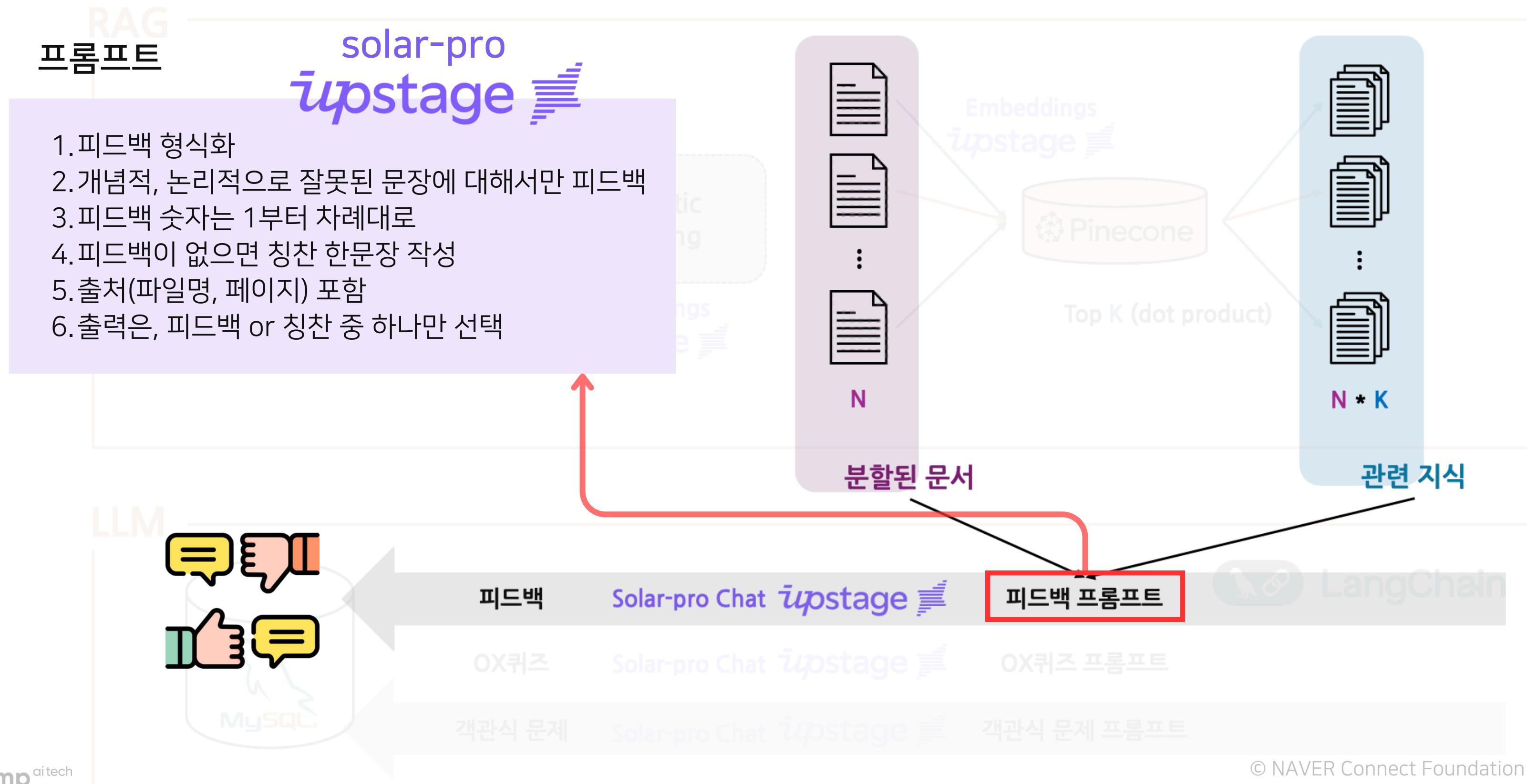
Name	Feedback (7D)	Run Count (7D)
dev-02	931	
Structured-Output-Chain	4	
SQL	0	

Showing 5 most active projects from the past 7 days. Tab shows number of total projects.



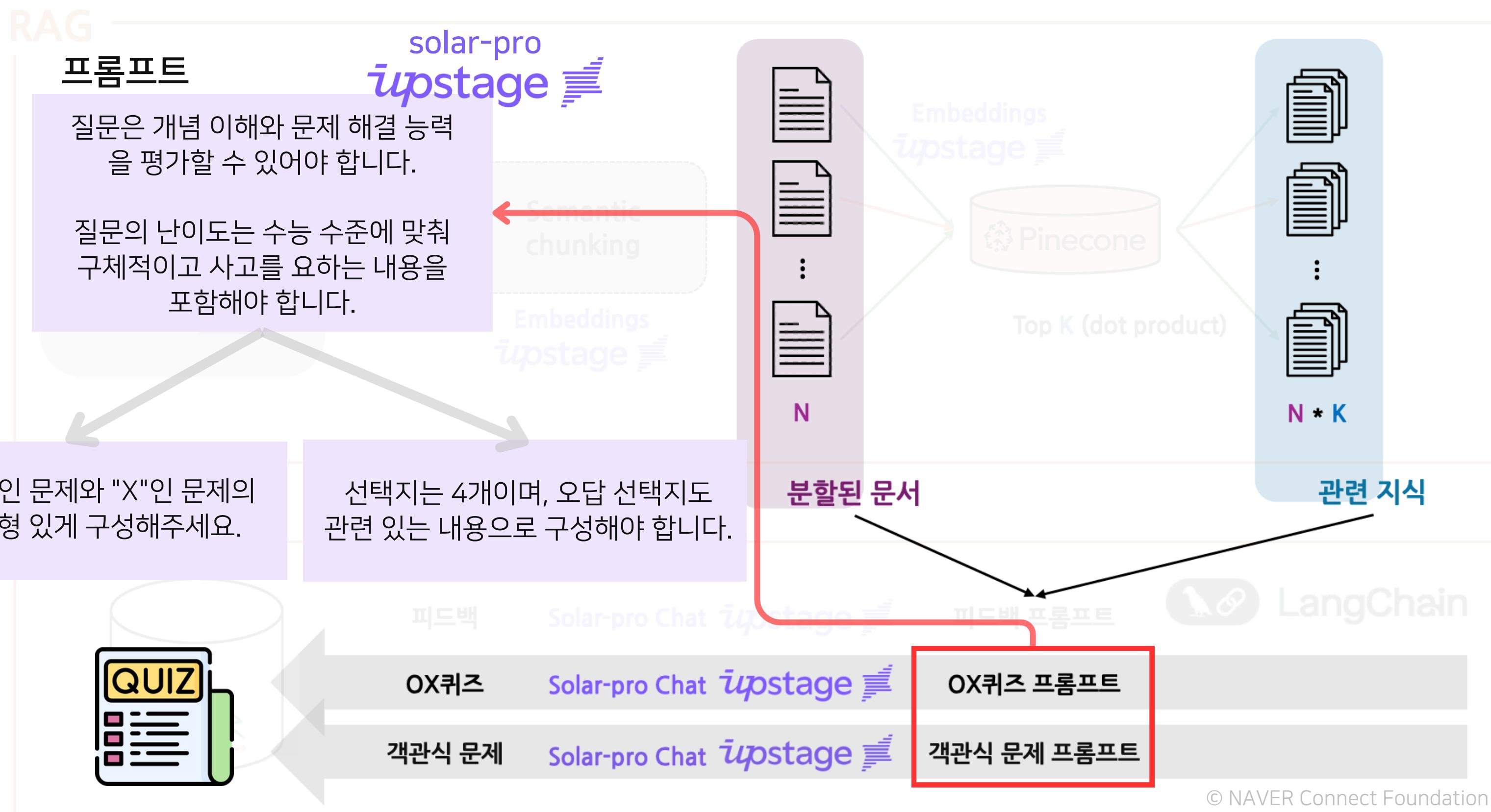
# 피드백 생성

\* Monitoring  LangSmith



# 퀴즈 생성

\* Monitoring  LangSmith

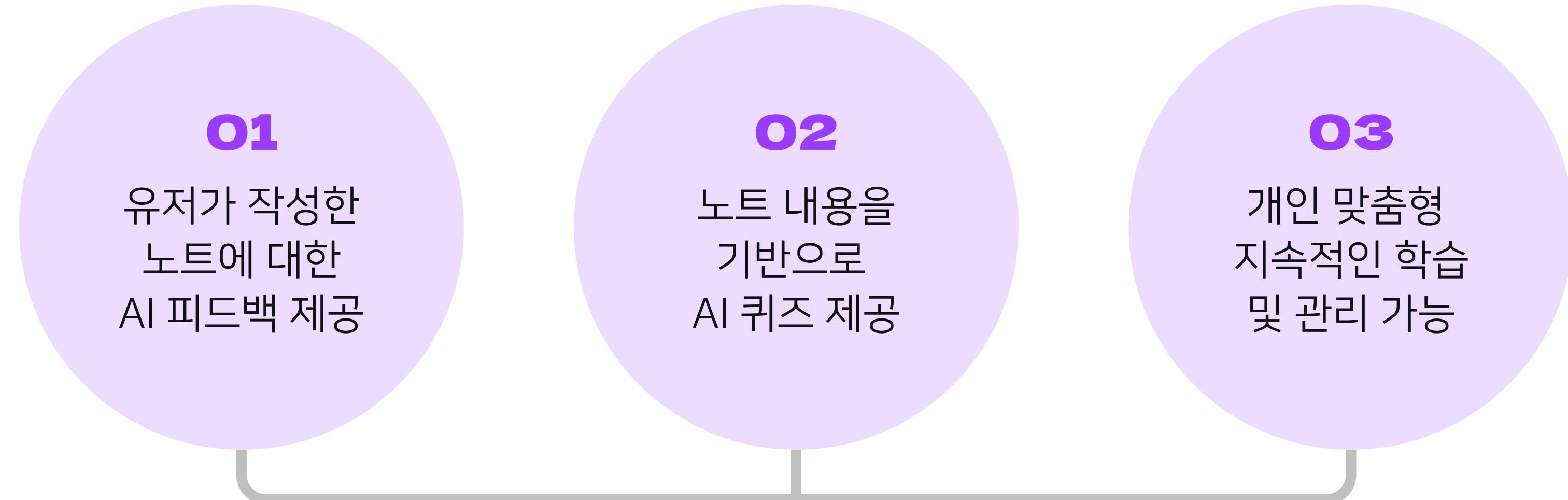


---

05

## 최종 결과

## 핵심 기능 요약



**Solar  
teacher**  
SINCE 2025

## 타겟&기대효과



예상 서비스 이용자

- 개념 복습을 확실히 하고 싶은 수험생/자격증 준비생
- 학습한 내용에 관련된 퀴즈를 풀고 싶은 학생
- 학생 관리를 하고 싶은 학부모/선생님

기대효과

- 학습 효율성 향상
- 자기주도적 학습 능력 강화
- 학습 내용의 장기 기억력 향상
- 취약점 파악 및 보완
- 부족한 부분을 즉각적으로 인지하고 보완

---

06

## 자체 평가

# 아쉬운 점

01



## OCR 관련 기술적 한계

- 한글 손글씨 인식률 저하 문제
- 특수 문자와 수식 기호 오인식 현상

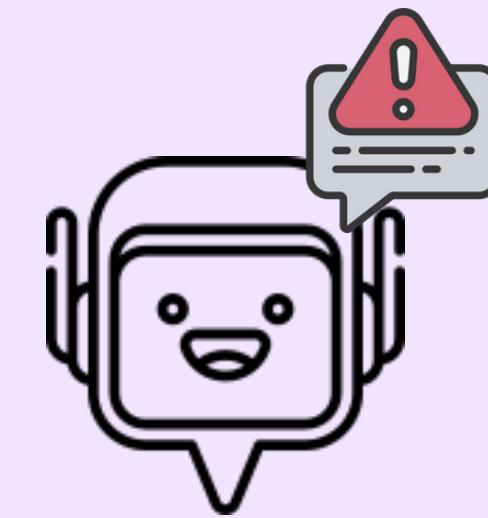
02



## 성능 및 처리 속도 관련 이슈

- 평균 응답 시간 60초로 사용자 경험을 저해
- 리소스 사용량 대비 최적화 부족

03



## 불안정한 LLM 출력

- 출력 결과와 실행 속도가 일관적이지 않음

# 발전 방향

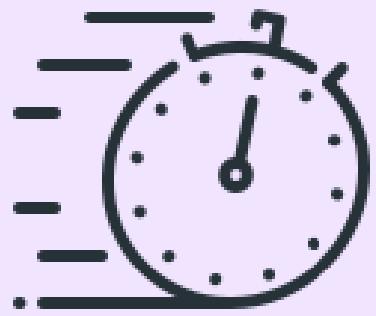
01



## OCR 성능 개선 방안

수식/특수문자 인식을 위한 전처리 로직 강화  
다양한 필기체 데이터셋 구축 및 학습

02



## 서비스 속도 최적화 전략

비동기 처리를 통한 응답 시간 단축  
효율적 관리를 위한 로드밸런싱 도입

03



## LLM 출력 안정화 방안

JSON Parser 도입을 통한 출력 포맷 표준화  
Upstage groundness Check 시스템을 도입

# 확장 가능성

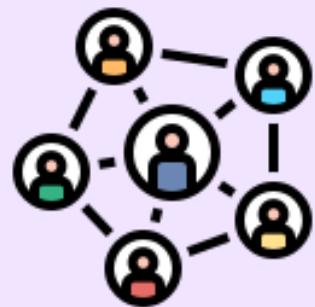
01



## 출판사와의 제휴

교과서 및 문제집을 이용한 추가 컨텐츠 제공  
각종 자격증 및 시험으로 확장 가능

02



## 소셜 서비스 추가

팔로우/팔로워 등 커뮤니티 기능 추가  
보상 시스템으로 게이미피케이션 기능 추가

03



## 다국어 서비스 추가

어학 시험 등 범위 확장  
다국어 기능 추가

# 느낀점

## O1 • 기획의 중요성

- 아이디어를 현실로 만들기 위해선 명확한 기획이 필수적이었다. 초반에 기능을 계속 추가하려다 보니 핵심 가치를 흐릴 뻔했지만, 핵심 목표를 정리하며 방향을 잡을 수 있었다.

## O2 • 실제 사용자 관점에서의 고민

- “이 기능이 정말 유용할까?”라는 질문을 계속 던지며 개발했다. 사용자 입장에서 필요한 기능을 고민하고, 직관적인 UX/UI 설계를 위해 노력했다.

## O3 • 협업과 의사소통

- 팀원들과의 원활한 협업이 프로젝트 성공의 핵심이었다. 각자의 역할을 명확히 분배하고, 효율적으로 의견을 조율하는 과정이 큰 도움이 되었다.

## O4 • 빠르게 만들고, 빠르게 개선하기

- 완벽한 서비스를 만들기보다, 빠르게 프로토타입을 만들어 피드백을 반영하며 개선하는 과정이 중요하다는 것을 깨달았다.

---

Thank You

---

# 07

## Appendix

# 백지공부법 관련 논문 자료

## The Power of Testing Memory Basic Research and Implications for Educational Practice

Henry L. Roediger, III, and Jeffrey D. Karpicke

Washington University in St. Louis

이 논문은 테스트 효과(Testing Effect)에 대한 내용으로 시험을 통해 기억력을 향상시키고 학습을 강화하는 방법에 대해서 이야기 하고 있다

백지 공부법 = 자기 테스트(Self-Testing)

- 논문에서 강조한 테스트 효과(The Testing Effect)를 활용한 학습법
- 단순히 읽거나 필기하는 것보다 직접 기억을 떠올리는 과정이 기억 유지에 효과적

왜 효과적인가? (논문과의 연관성)

1. 적극적인 회상(Active Recall) → 기억 강화
  - 백지에 스스로 떠올려 쓰는 과정이 자유 회상 테스트(Free Recall Test)와 동일
  - 연구에 따르면, 단순 복습보다 기억 검색이 학습 효과를 높임
2. 테스트 후 피드백 제공 → 학습 최적화
  - 작성한 내용을 교재와 비교하며 부족한 부분 보완
  - 연구에 따르면, 피드백을 제공했을 때 장기 기억 효과 증가
3. 반복 테스트(Repeated Testing) → 장기 기억 전환
  - 논문에서는 반복적인 테스트가 장기 기억 형성에 필수적이라고 강조
  - 백지 공부법도 주기적으로 반복하면 학습 효과 극대화

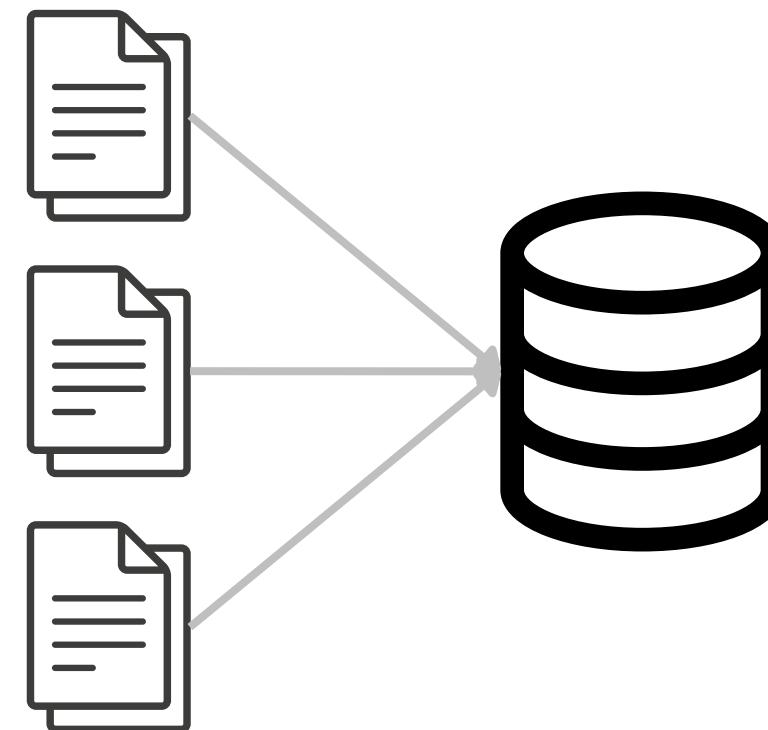
# RAG DB 구성 - 데이터 전처리

## 데이터 전처리 방법

- 페이지 필터링: 목차, 부록 등 불필요한 페이지 제외
- 불용어 제거: 임의의 불용어가 포함된 페이지 제외
  - 불용어: “창의·융합”, “창의·융복합”, “마무리”, “자료 해석”, “부록”, “정답과 해설”, “융복합적 사고”
- 청크 분할: RecursiveCharacterTextSplitter
- 메타 데이터 추가: 과목, 페이지, 파일 이름 추가

## 전처리 결과

- 6,408개의 데이터 --> 5,080개의 데이터



1	ID 0041ebd9-9a29-4cca-8026-88b550ce37f7
SCORE	FIELDS
1.0008	<b>page_num:</b> 168
	<b>source:</b> "천재교육_고등교과서_생명과학 II_이준규(15개정)_교과서 본문.pdf"
	<b>subject:</b> "생명과학 II"
	<b>text:</b> "진핵생물이 다세포 진핵생물로 진화하였다. 원시\n원핵세포\n산소 호흡\n세
	<b>total_pages:</b> 248

# 테스트 데이터 생성

## 테스트 데이터 생성 프롬프트

'''Given the following text, please:

- 1) Create 2 statements using the core concepts
  - one correct and one incorrect
- 2) Mark each statement as O (correct) or X (incorrect)

All statements must:

- Include core concepts from the original text
- Be different from the original text format
- Must answer in Korean.

Original text:

{text}

Please respond in the following format:

{format\_instructions}

'''

OX 퀴즈를 생성하여, 모델의 정확도를 평가.

- 모델
  - gpt-4o
- 데이터
  - 물리학2: EBS 수능완성 개념 페이지 (Splitting: chunk\_size: 500, overlap: 150)
  - 화학2: EBS 수능완성 개념 페이지 (Splitting: chunk\_size: 500, overlap: 150)
  - 생명과학2: EBS 수능완성 개념 페이지 (Splitting: chunk\_size: 500, overlap: 150)
  - 지구과학2: EBS 수능완성 개념 페이지 (Splitting: chunk\_size: 500, overlap: 150)
  - 정치와 법: EBS 수능완성 개념 페이지 (Splitting: chunk\_size: 500, overlap: 150)
  - 정보처리기사: 시나공 핵심 요약 (Splitting: chunk\_size: 2000, overlap: 150)
- OX 퀴즈 생성 데이터와 RAG DB 구성 데이터 분리.
- 한국어 버전과, 영어 한국어 혼합 버전 2가지로 구성.
- 총 문제 수
- 영어, 한글 혼합 데이터
  - 물리학2: 178, 생명과학2: 145, 지구과학2: 175, 화학2: 196
  - 정치와 법: 114, 정보처리기사: 236
- 한글 데이터
  - 물리학2: 140, 생명과학2: 69, 지구과학2: 135, 화학2: 155
  - 정치와 법: 96, 정보처리기사: 160

# 과목 선정

## 국어, 영어를 제외한 이유

- 문법과 문맥을 틀리게 작성하여도 맞고 틀림의 기준이 애매함.

원문

Different creative pursuits require varying degrees of unconscious flexible thinking, in combination with varying degrees of the conscious ability to adjust it and shape it through analytical thinking. In music, for example, at one end of the creative spectrum are improvisational artists, such as jazz musicians.



Different creative pursuits require varying degree of unconscious flexible think, in combine with vary degrees of the conscious ability to adjust it and shape it through analytical thinking. In music, However, at one start of the creative spectrum are improvisational artists, such as jazz musicians.

피드백

✓ 좋은 글입니다!

## 수학을 제외한 이유

- OCR이 수식을 제대로 인식하지 못함

$$f(x) = \frac{\frac{x}{2} - \sqrt{5} + 2}{\sqrt{1 - x^2}}$$

$$f(x) = +$$

$$\int (\sin x + \frac{x}{2}) dx = \int \frac{1 + \sin x}{2} dx$$



x	
- V5 + 2	
2	
f(x)	
v1 - x	
f(x)	
1+ sinx	
(sinx + - dx dx	
2 2	

# 실험 과정 - OCR

## 01. 손글씨 데이터 생성

1. 줄 글로 작성
2. 한글과 영어만 사용
3. window에서 입력 가능한 특수 기호만 사용
4. 한 단락으로 구성
5. 틀린 점을 포함하여 작성

## 02. 테스트 코드 작성

1. 파일을 전처리 없이 OCR로 전달
2. 파일을 흑백 처리한 후 OCR로 전달
3. 파일을 압축한 후 OCR로 전달
4. 파일을 흑백+압축한 후 OCR로 전달

## 03. 정확도 측정

1. 손글씨 데이터에 대한 원본 txt 파일 준비
2. OCR 결과, 원본 데이터를 단어 단위로 분할
3. 원본 데이터 배열에서 OCR 결과 배열의 합집합을 삭제
4. 원본 단어 수와 합집합을 삭제한 후 단어 수를 비교해 정확도를 측정

## 04. 시간 측정

1. 손글씨 데이터를 전처리 -> OCR 하는 함수 코드 작성
2. 같은 데이터를 10번 반복해서 처리한 뒤 처리 시간 평균을 측정

[GitHub issue #39](#)

# 실험 과정 - RAG 성능 비교

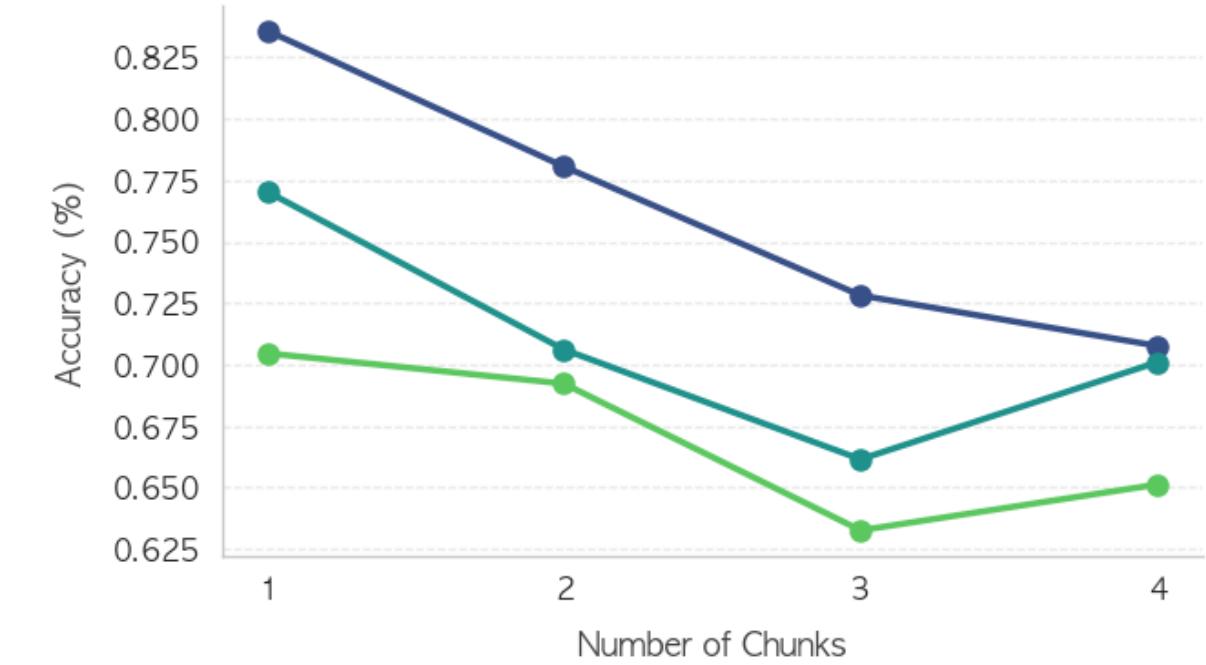
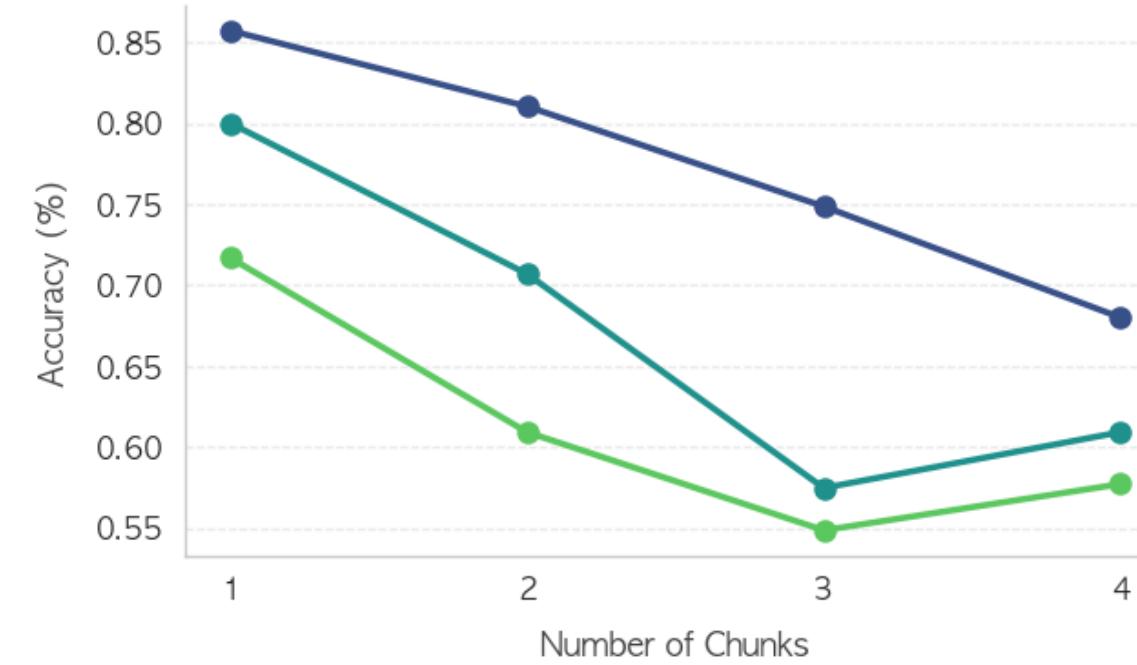
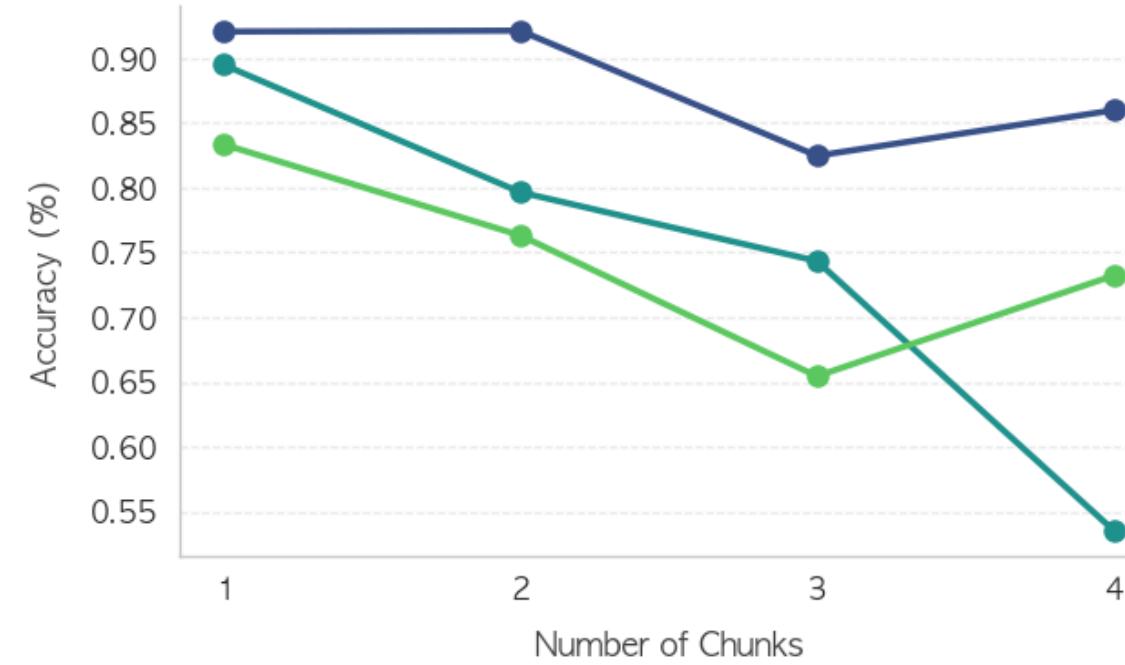
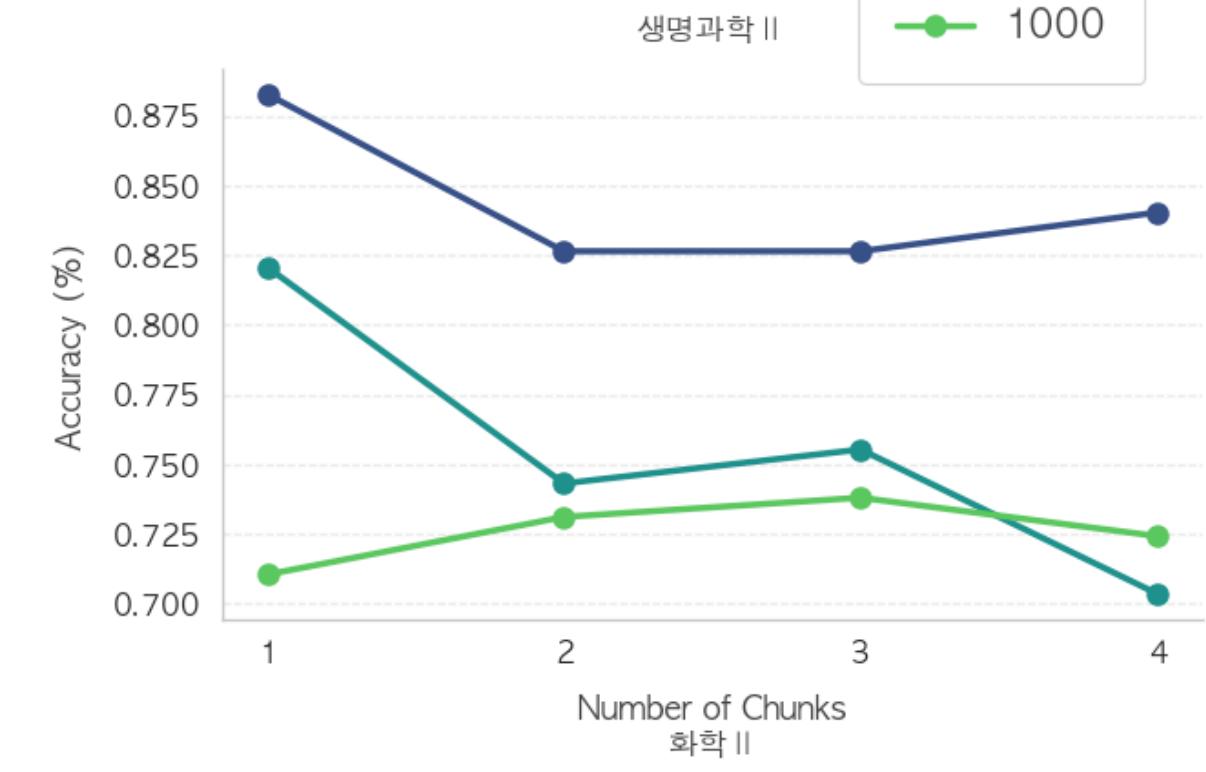
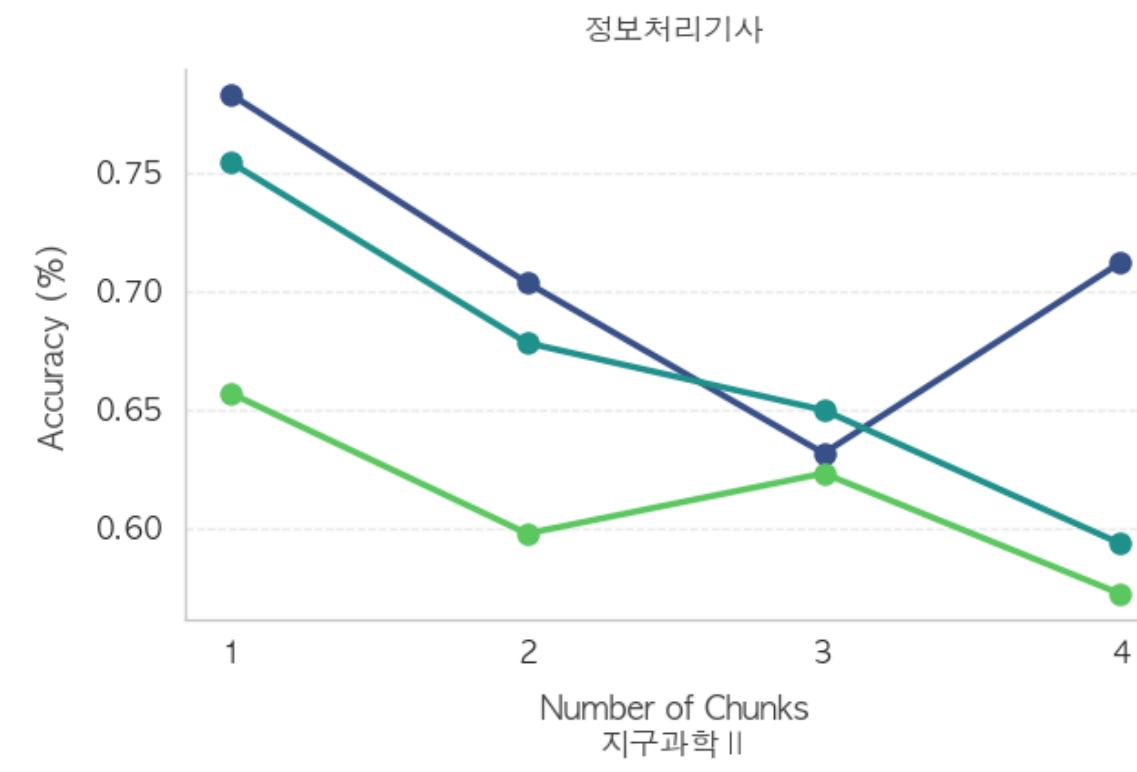
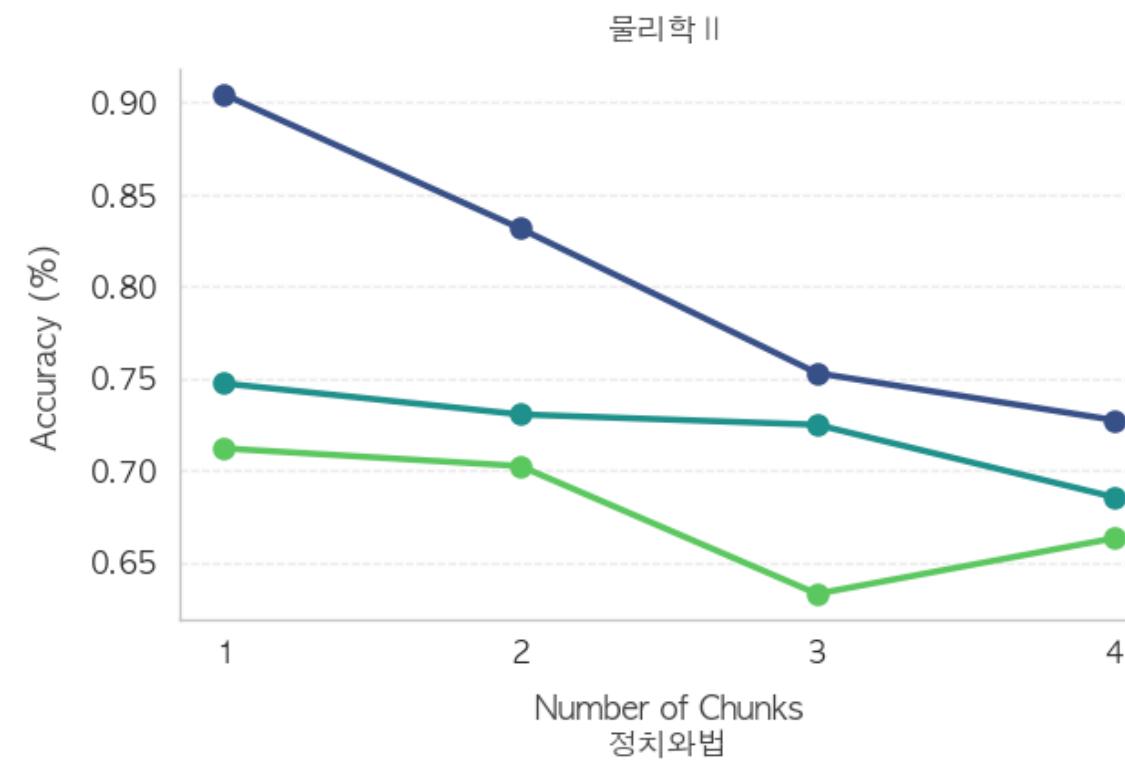
- “틀린 문장을 얼마나 잘 찾는가?”를 OX로 라벨링한 참/거짓 문장 데이터셋으로 평가.
- Variables
  - 참고하는 청크의 수(1, 2, 3, 4)
  - 분할된 청크의 크기(250, 500, 1000)
  - RAG DB 구성 시, 전처리 유무
- 각 조합별로 6개 과목에서 Accuracy, Precision, Recall, 실행 시간을 평가.
- Results
  - [RAG Result](#)

Table 1: RAG Experiment Results for '물리학II' and '화학II' in Mixed English & Korean OX Questions							
Dataset	Model	Chunks	Chunk Size	Metric			Time(s)
				Accuracy	Precision	Recall	
	LLM	-	-	0.7472	0.8519	0.7603	241.54
물리학II	LLM with RAG	1	250	0.8427	0.9115	0.8512	247.18
		2	250	0.7865	0.8745	0.8017	248.34
		3	250	0.7345	0.8284	0.8027	250.17
		4	250	0.7027	0.8087	0.7711	251.02
		1	500	0.7022	0.8017	0.6904	275.34
		2	500	0.7528	0.8348	0.7931	343.18
		3	500	0.6839	0.7857	0.7395	561.27
		4	500	0.618	0.7624	0.6364	287.51
		1	1000	0.7159	0.8017	0.775	273.4
		2	1000	0.6934	0.7816	0.5763	312.76
화학II	LLM with RAG-pre	3	1000	0.6322	0.7368	0.7119	444.3
		4	1000	0.5898	0.6916	0.6667	467.35
		1	250	<b>0.9045</b>	0.9067	0.9174	216.34
		2	250	0.8315	0.876	0.876	214.28
		3	250	0.7528	0.8468	0.7769	212.61
		4	250	0.7278	0.7951	0.8083	227.74
		1	500	0.7472	0.8585	0.7521	275.19
		2	500	0.7302	0.8067	0.7934	267.37
		3	500	0.7247	0.7951	0.8017	231.97
		4	500	0.6854	0.8351	0.6694	283.7
물리학II	LLM with RAG	1	1000	0.7119	0.7949	0.775	243.17
		2	1000	0.7047	0.7982	0.7521	255.72
		3	1000	0.6328	0.7273	0.731	311.04
		4	1000	0.6629	0.7099	0.719	416.06
		1	250	0.7756	0.8571	0.7907	258.86
		2	250	0.7268	0.7953	0.7801	269.83
		3	250	0.6923	0.7674	0.7674	319.04
		4	250	0.6276	0.7545	0.6434	349.06
		1	500	0.7077	0.8198	0.7109	241.22
		2	500	0.7011	0.7934	0.7442	646.61
화학II	LLM with RAG	3	500	0.6719	0.7119	0.7009	308.72
		4	500	0.6434	0.6434	0.6434	349.06
		1	1000	0.7077	0.8198	0.7109	241.22
		2	1000	0.7011	0.7934	0.7442	646.61
		3	1000	0.6719	0.7119	0.7009	308.72
		4	1000	0.6434	0.6434	0.6434	349.06

# Model Accuracy with Increasing Chunk Numbers

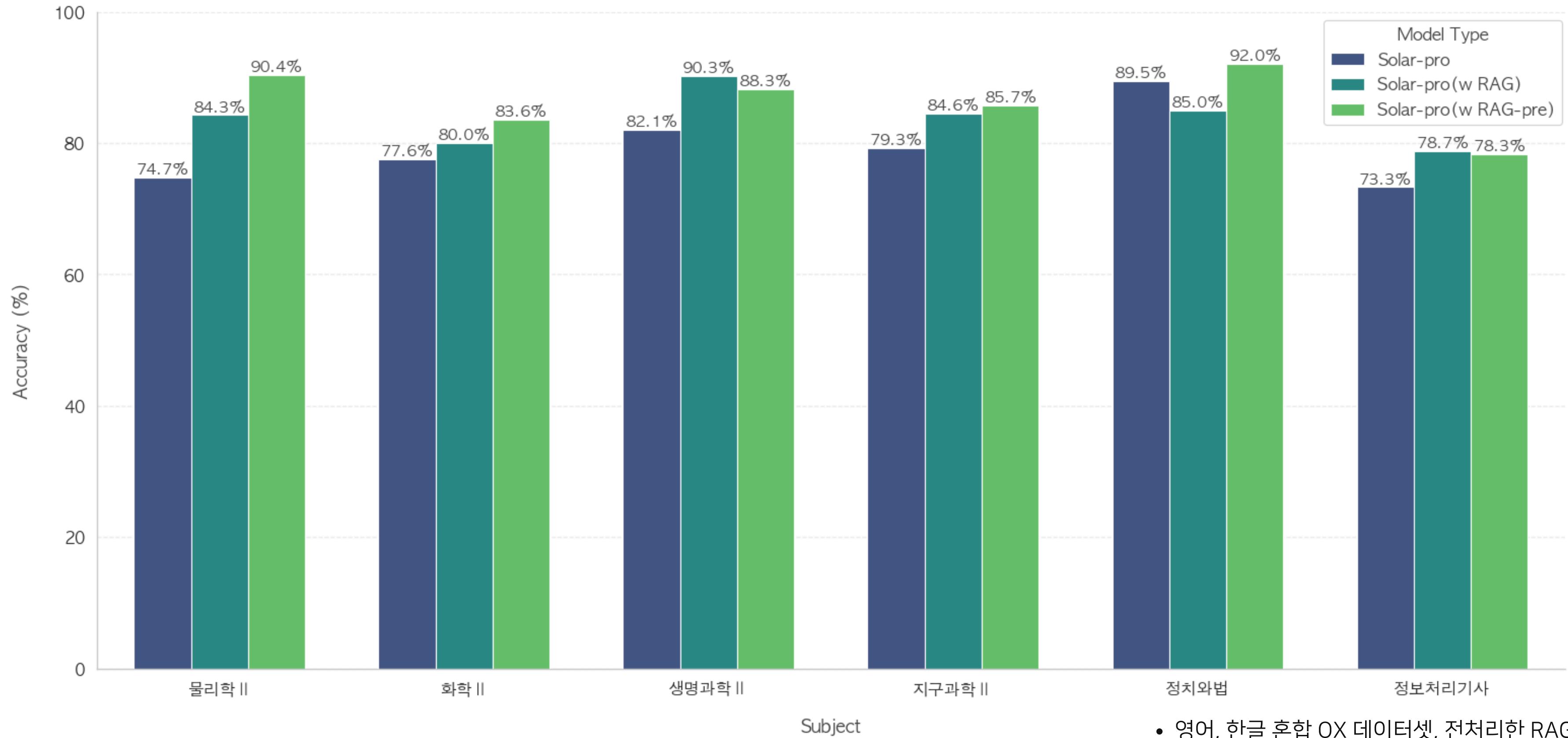
Chunk Size

- 250
- 500
- 1000



- 영어, 한글 혼합 OX 데이터셋, 전처리한 RAG

## Model Performance Comparison by Subject (Chunk Size=250, Chunks=1)



# 실험 과정 - RAG 유무에 따른 피드백/해설 비교

	LLM	LLM+RAG
피드백 설명	피드백이 상대적으로 일반적인 설명에 그침.	피드백이 보다 정교하고 정확함
피드백 내용	<p>피드백: '거성 또는 초거성 단계'에서 '질량이 클수록 수명이 길어지고 더 밝으며 온도는 낮아짐'이라고 설명한 부분은 잘못된 내용입니다. 실제로는 질량이 클수록 수명이 짧아지고 더 밝으며 온도는 높아집니다.</p> <p>해설: '거성 또는 초거성 단계'에서 질량이 클수록 수명이 짧아지고 더 밝으며 온도는 높아집니다.</p>	<p>잘못된 내용: "질량이 클수록 수명이 길어지고 더 밝으며 온도는 낮아짐."</p> <p>고칠 내용: "질량이 클수록 수명이 짧아지고 더 밝으며 표면 온도는 낮아짐."</p> <p>해설: 별의 질량이 클수록 핵융합 반응이 더 빠르게 일어나며, 이는 별의 수명을 단축시킵니다. 그러나 질량이 큰 별은 더 밝게 빛나며, 표면 온도는 낮습니다. 따라서 "질량이 클수록 수명이 길어져" 대신 "질량이 클수록 수명이 짧아져"라고 표현해야 하며, "온도는 낮아짐" 대신 "표면 온도는 낮아짐"이라고 표현해야 합니다.</p>
퀴즈 설명	문제의 난이도가 균일하지 않고 단순화되는 경향	정답에 대한 해설이 구체적이고 교과서적 표현을 활용
퀴즈 내용	<p>주계열성 단계에서 별은 내부의 중력과 외부로 나가는 방사압이 균형을 이루지 못하며 진동을 반복하나요? (O)</p> <p>설명: 주계열성 단계에서 별은 내부의 중력과 외부로 나가는 방사압이 균형을 이루지 못하며 진동을 반복한다.</p>	<p>주계열성 단계에서 별은 내부의 중력과 외부로 나가는 방사압이 균형을 이루며 진동을 반복한다. (O/X)</p> <p>정답: X</p> <p>설명: 주계열성 단계에서 별은 내부의 중력과 외부로 나가는 방사압이 균형을 이룬다. 진동은 거성 또는 초거성 단계에서 발생한다.</p>

- RAG를 활용하면 피드백의 정확성, 신뢰도, 교과서적 표현 반영이 향상됨.
- OX 퀴즈의 품질이 개선되며, 설명이 구체적이고 학습 효과가 높아짐.
- 범용적인 개념을 많이 다루는 수능 수준의 퀴즈라면 RAG 없이도 충분할 가능성이 높지만, 잘못된 개념을 교정하거나 보다 정교한 문제를 만들려면 RAG가 도움될 수 있음.

# 실험 과정 - RAG 문서 로더

Name	Result	Time
PyPDFLoader	학습 목표 01아리스토텔레스가 말하는 “인간은 정치적 동물(Homo Politicus)이다.”에서 정치란	1.5s
PyPDFLoader with OCR	학습 목표 01아리스토텔레스가 말하는 “인간은 정치적 동물(Homo Politicus)이다.”에서 정치란	36.7s
PyMuPDFLoader	학습 목표 01 아리스토텔레스가 말하는 “인간은 정치적 동물(Homo Politicus)이다.”에서 정치란	0.2s
PyPDFium2Loader	학습 목표 01 아리스토텔레스가 말하는 “인간은 정치적 동물(Homo Politicus)이다.”에서 정치란	0.2s
PDFMinerLoader	매우 성능이 낮음 (notebook 출력 참고)	2.0s
PDFPlumberLoader	01 정치의 의미와 기능 아리스토텔레스가 말하는 “인간은 정치적 동물(Homo Politicus)이다.”에서 정치란 어떤 의미를 담고 있을까? •정치의 의미를 다양한 관점에서 이해할 수 있다. 학습 목표 •구체적 사례를 통해 정치의 기능을 설명할 수 있다. 아리스토텔레스 ► (Aristoteles, 기원전 384~기원전 322)	2.8s

여러 PDF 문서 로더 비교를 통해 가장 적합한 PyMuPDFLoader를 선택함

## PyMuPDFLoader

- PDF 문서의 각 페이지를 개별 Document 객체로 변환
- 페이지의 텍스트 내용과 메타데이터가 포함
- 속도 최적화가 되어 있어 빠르게 PDF를 처리 가능

[GitHub issue #75](#)

# 실험 과정 - RAG 텍스트 분할

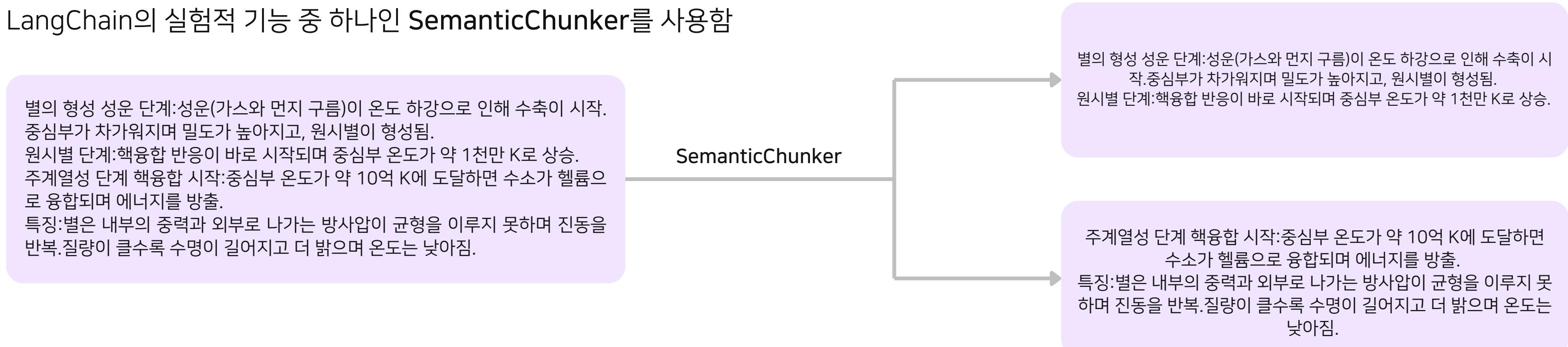
## 의미 기반 분할(SemanticChunker)

- 텍스트를 의미론적 유사성에 따라 분할하는 도구로, 문장 간의 유사성을 계산하여 문서를 의미 있는 청크로 나눔
- 업스테이지의 임베딩을 이용하여 텍스트의 유사성을 판단함

## 선택 이유

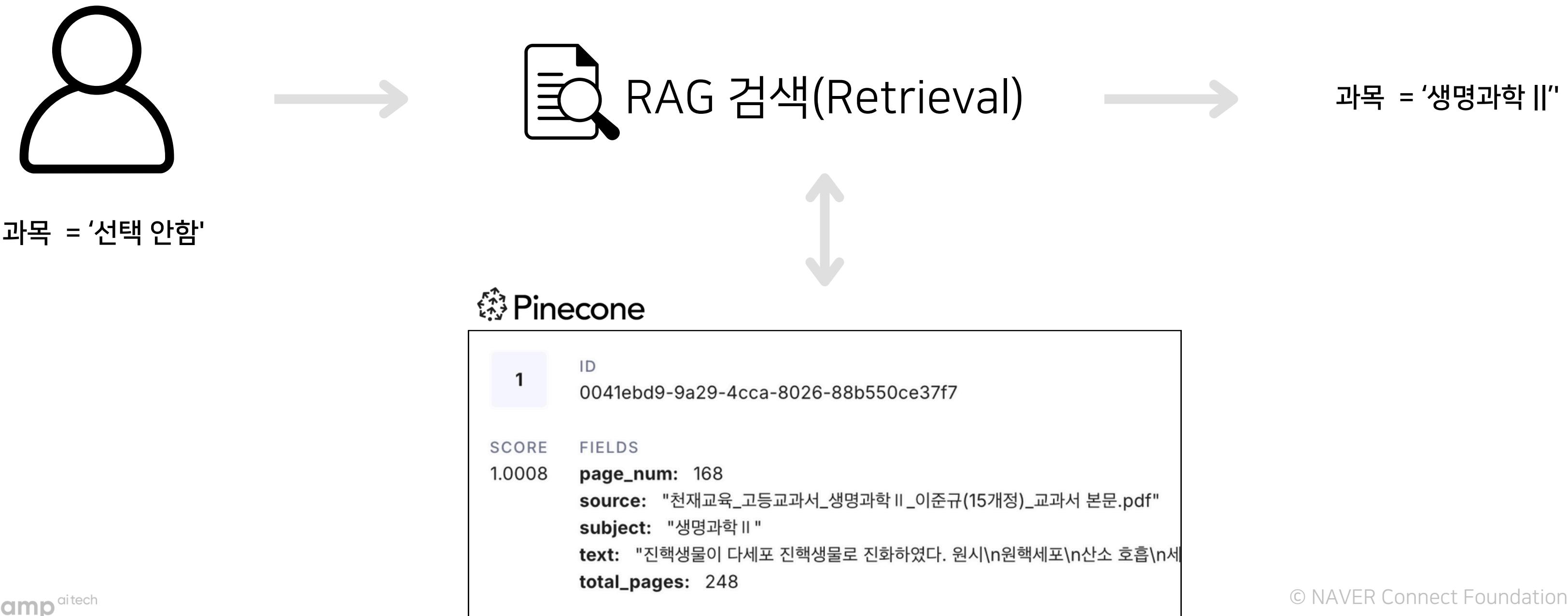
- 고정된 크기(문자 수)로 자르는 경우, 문장의 의미가 고려되지 않음
- 유저가 작성하는 백지의 경우, 문장기호가 제대로 되어 있지 않을 가능성이 높음
- 의미 기반을 사용할 경우,
  - 임베딩 사용으로 인한 비용이 들 수 있지만 이건 소량이라고 판단함
  - 출력되는 문장의 길이(토큰 수)가 고정되어 있지 않는 것은 문제가 아니라고 판단함

LangChain의 실험적 기능 중 하나인 SemanticChunker를 사용함



# 실험 과정 - 과목 자동 분류

과목을 '선택 안함'으로 할 경우,  
RAG의 검색(Retrieval) 과정에서 벡터 DB(Pinecone)의 subject를 가져와서 자동으로 과목을 지정해줌



# 실험 과정 - OX퀴즈, 사지선다

## 초기 프롬프트 (Baseline)

```
content = f"""
아래의 텍스트를 기반으로 O/X 퀴즈 5개를 만들어주세요.
응답은 JSON 형식으로 반환하세요:
[
    {"question": "질문 내용", "answer": "O 또는 X", "explanation": "정답
에 대한 설명"}, ...
] """
```



수준이 일정 이하인  
너무 간단한 문제들

## 수능형 프롬프트 (First Iteration)

```
content = f"""
아래의 텍스트를 기반으로 수능에 나올만한 문제의 수준으로
OX 퀴즈 5개를 만들어주세요. 응답은 JSON 형식으로만 반환해야 하세요
[
    {"question": "질문 내용", "answer": "O 또는 X", "explanation": "정답
에 대한 설명"}, ...
] """
```



난이도는 수능 문제에 가깝지만  
약간 간단한 단어맞추기 문제들

## 최종 개선된 프롬프트 (Final Version)

```
content = f"""
아래 텍스트를 기반으로 수능 수준의 O/X 퀴즈 5개를 만들어주세요.
다음 조건을 충족해야 합니다:
1. 질문은 개념 이해와 문제 해결 능력을 평가할 수 있어야 합니다.
2. 정답이 "O"인 문제와 "X"인 문제의 비율은 균형 있게 구성해주세요.
3. 난이도는 수능 수준에 맞춰 구체적이고 사고를 요하는 내용을 포함해야 합니다.
4. 각 질문의 정답에 대한 설명은 간결하지만 충분히 납득 가능하게 작성해주세요.
"""
```



개념 이해와 사고를 요하는  
내용을 포함한 문제들

# 실험 과정 - OX 퀴즈, 사지선다

생성된 퀴즈 예시: {

    "question": "조석 현상은 지구의 물질에 작용하는 달과 지구의 인력에 의해 일어나며, 이로 인해 바닷가의 해수면은 주기적으로 상승과 하강을 반복한다. (O/X)",

    "answer": "O",

    "explanation": "정답. 조석 현상은 달과 지구의 인력에 의해 발생하며, 바닷가의 해수면이 주기적으로 상승과 하강을 반복한다."

}

...

==== 테스트 결과 요약 ===

Configuration: temp=0.3, top\_p=0.9

Token usage: 1012

Configuration: temp=0.7, top\_p=0.9

Token usage: 976

Configuration: temp=1.0, top\_p=0.9

Token usage: 931

Configuration: temp=0.7, top\_p=0.5

Token usage: 916

...

공식 웹사이트에서 확인된 파라미터들로 튜닝

초기 설정(Configuration: temp=0.3, top\_p=0.9)

중간 조정 (Configuration: temp=0.7, top\_p=0.9)

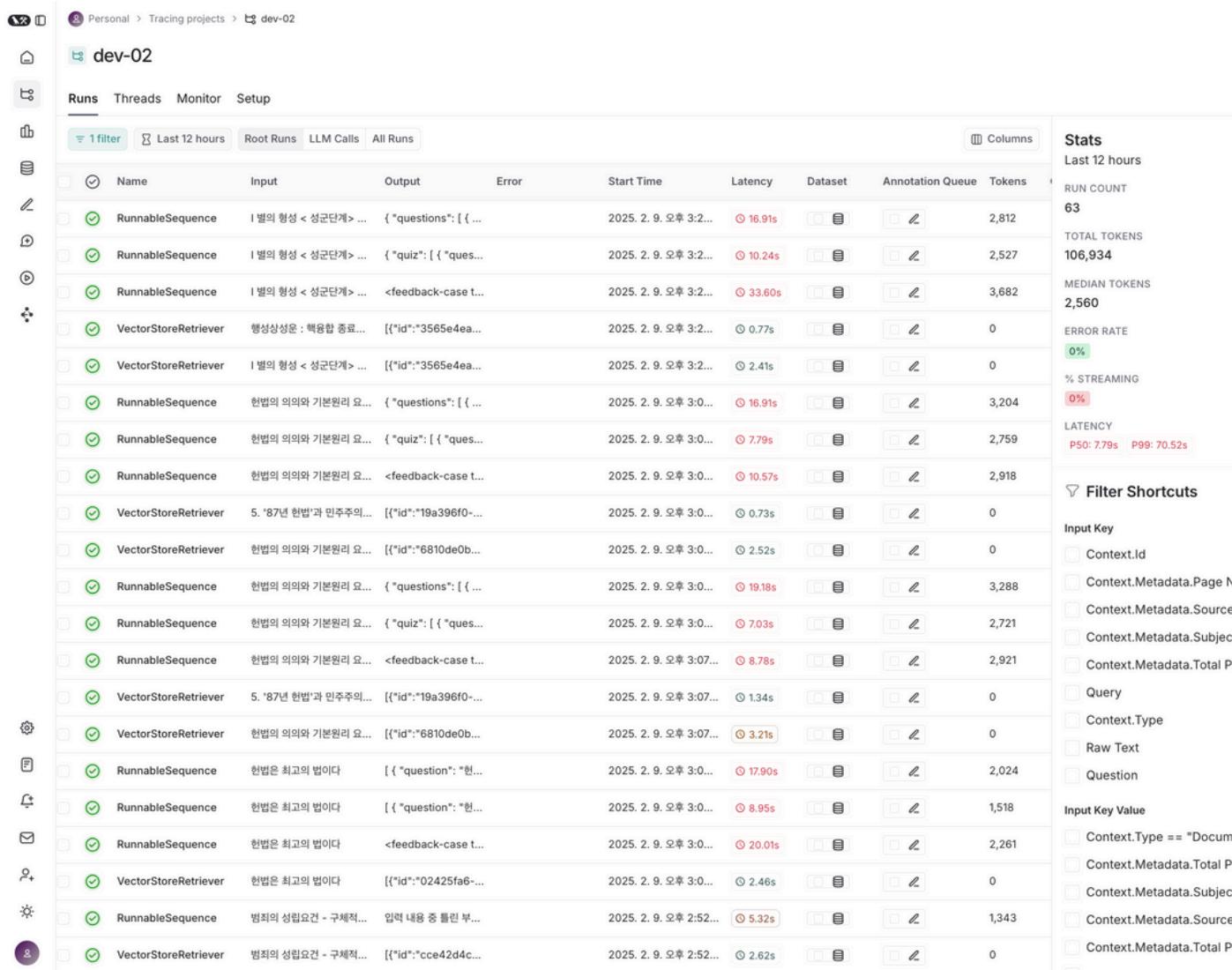
높은 설정 (Configuration: temp=1.0, top\_p=0.9)

각 설정에 따른 Token usage 분석 결과,  
temperature=0.7, top\_p=0.9 설정이 토큰 사용량과 퀴즈  
품질 면에서 가장 효율적인 것으로 확인됨.

# LLM inference time 분석

LangSmith를 이용하여 12개의 데이터에 대해서 LLM inference time을 분석함

- RAG를 이용한 검색, 피드백 생성, OX 퀴즈 생성, 사지선다 문제 생성에 대해서 분석
- 피드백을 생성하는 부분에서 가장 시간이 오래 걸림



LangSmith

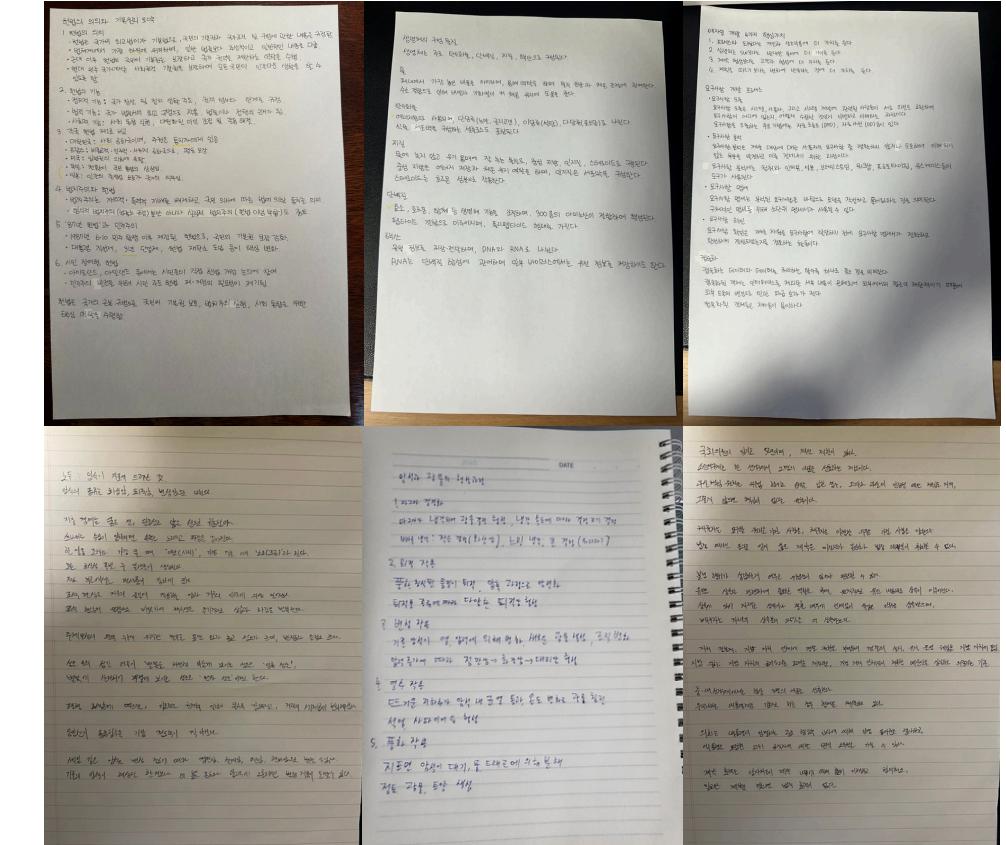


Table 1: LLM inference time 분석 (단위: s)

검색(Retrieval)	피드백	OX 퀴즈	사지선다
<b>합계</b>	<b>39.76</b>	<b>283.35</b>	<b>105.16</b>
<b>평균</b>	<b>3.31</b>	<b>23.61</b>	<b>8.76</b>

# References

- <https://wikidocs.net/251190> - <랭체인LangChain 노트> - LangChain 한국어 튜토리얼
- <https://if-blog.tistory.com/12821> - 교육부 블로그
- <https://www.etoday.co.kr/news/view/2363273> - 자기주도학습 관련 뉴스 기사
- <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci? sereArticleSearchBean.artild=ART002252978> - 백지공부 관련 논문
- <https://www.mk.co.kr/news/business/10942416> - 에듀테크 관련 뉴스 기사
- <https://www.etoday.co.kr/news/view/2285141> - 에듀테크 관련 뉴스 기사
- <https://if-blog.tistory.com/12060> - 교육부 블로그
- 미래N 사회문화 교과서
- <https://www.sinagong.co.kr/pds/001001001/past-exams> - 시나공 정치기
- <https://www.ebsi.co.kr/ebs/pot/poti/main.ebs> - 교과서 PDF 물화생지 정치와 법
- 나머지 사회과목들은 EBS 2022 수능완성
- 테스트 데이터 생성 시 사용
- EBS 2022 수능완성 + 시나공 정치기 다른 년도

---

# End of Document

## Thank You.

Thank you