

학술논문 초록 작성에 대한 대형 언어모델 적용 가능성 탐색

일반대학원 문헌정보학과 문헌정보학전공
석사과정 1학기차 하인영

논문 개요

- 저자
 - 김유미: 제1저자, 경북대학교 박사과정생
 - 석사 논문 <독해 수준과 텍스트 난이도를 고려한 개인화 정보 검색 성능의 향상>
 - 양승원: 공동저자, 루이지애나 주립 대학 CCT 연구소 부교수
 - 학술논문 <지도 기반 데이터 시각화 플랫폼 사용성 평가 기준 개발 및 적용 연구> 공동저자
 - 이종욱: 교신저자, 경북대학교 문헌정보학과 부교수
 - 연구 동향, 계량서지학
- 2025-10-20 접수→2025-10-31 1심→2025-11-10 게재 확정
 - 한국문헌정보학회지 제59권 4호 발행

본문 목차

1. 서론

2. 개념적 배경 및 관련 연구

2.1 학술논문 초록의 개념과 유형

2.2 LLM을 활용한 학술 텍스트 요약

2.3 한국어 문서 요약과 품질 평가 연구

3. 연구방법

3.1 데이터셋 개요

3.2 프롬프트 설계 및 유형 구분

3.3 유사도 측정

3.4 전문가 인식 조사

4. 연구 결과

4.1 초록의 글자 수 비교

4.2 유사도 분석

4.3 전문가 인식 조사

5. 논의 및 결론

본문 읽기: 서론

- 전개



초록의 중요성	연구의 핵심 내용을 담고 있어 다양한 분야에서 유용하게 쓰임
초록 품질 향상을 위한 연구	초록 구조와 작성 원칙 제시, 핵심 요소 정의 등을 시도해왔음
초록 자동 생성 연구	TextRank, LSA, Pointer-Generator Network와 같은 기법이 초기에 제시됨
연구의 필요성	LLM 적용 연구의 미비함
기대 효과	초록 품질 개선 방법 제시
분석에 사용한 데이터 및 연구 방법 개략	

본문 읽기: 개념적 배경 및 관련 연구

2.1 학술논문 초록의 개념과 유형

본문 발췌: 학술논문의 초록은 논문의 핵심 내용을 간결하면서도 포괄적으로 제시하는 요약으로, 독자가 논문의 내용을 빠르게 파악하고 자신의 관심사와의 관련성을 판단할 수 있도록 돕는다(Klimova, 2015).

지시적 초록: 문헌이 무엇에 관한 것인지를 기술

정보적 초록: 연구의 목적, 방법론, 결과, 결론, 권고사항까지 구체적으로 기술

→혼재되어 나타남

본문 읽기: 개념적 배경 및 관련 연구

2.2 LLM을 활용한 학술 텍스트 요약

LLM의 요약 성능이 향상되고 있으나, 환각 문제가 발생함

→*프롬프트 엔지니어링 / 생성 결과의 신뢰성을 평가할 방안을 연구

*본문 발췌: LLM으로부터 원하는 결과를 얻기 위해서는 프롬프트가 구체적인 맥락 정보를 포함해야 하며, 이러한 정교한 프롬프트를 구성하는 과정과 기술을 ‘프롬프트 엔지니어링’이라 한다(Marvin et al., 2024).

본문 읽기: 개념적 배경 및 관련 연구

2.3 한국어 문서 요약과 품질 평가 연구

- 텍스트 요약 연구가 선행되어 왔으나, 특정 주제 분야에 한정됨
- 요약문 대상 정량적 품질 평가 방안에 관한 연구: 학술 초록을 대상으로 한 실증적 연구는 미비한 편임

본문 읽기: 연구방법

❖ 인간초록(원초록) vs AI초록(생성초록)

❖ 인간초록(원초록) vs 논문의 본문

❖ AI초록(생성초록) vs 논문의 본문

✓ 표면적 어휘기반 유사도

✓ 의미 기반 유사도

본문 읽기: 연구방법

분석 데이터: 2022년-2024년 한국문헌정보학회지 수록 한글 논문 204편

TF*IDF 표면적 단어기반 유사도→형태소 분석기를 이용하여 데이터 전처리 과정을 거침

BERTScore: 의미 기반 유사도→비정형적 텍스트를 분석함

AI초록: Python PyMuPDF 라이브러리, GPT-4o와 Open API 키를 통해 Python 환경에서 생성함

본문 읽기: 연구방법

★
다국어
프롬프트에
관한
선행연구

〈표 1〉 ChatGPT 기반 초록 생성을 위한 프롬프트 유형

프롬프트 유형	내용
Case 1 (Base)	Please write an abstract for the following academic paper. The abstract must be written in Korean as a single paragraph without any breaks and must be between 400 and 600 characters in length.
Case 2 (Base-Korean)	다음은 한 편의 학술논문 본문입니다. 이 내용을 바탕으로 초록을 한국어로 작성해 주세요. 초록은 400자 이상 600자 이내로, 단락 구분 없이 연속적으로 작성해 주세요.
Case 3 (Structure)	Base Promprt + The abstract must include the research purpose, methodology, and main findings.
Case 4 (Few-shot)	Base Prompt + Here are two examples of well-written academic abstracts in Korean: Example 1, Example 2 ¹⁾ ¹⁾ 2021년 발표된 한국문헌정보학회지 논문 중 피인용 횟수 상위 5편을 대상으로, 전문가 검토를 통해 두 편을 선정함
Case 5 (Role)	Base Prompt + You are a senior researcher with over 20 years of experience in the field of Library and Information Science.
Case 6 (All combined)	Base Prompt + Structure + Few Shot + Role

본문 읽기: 연구방법

전문가 서면 조사

전문가 구성: 현직 교수 5인, 학술지 게재 경험이 있는 박사과정생 5인

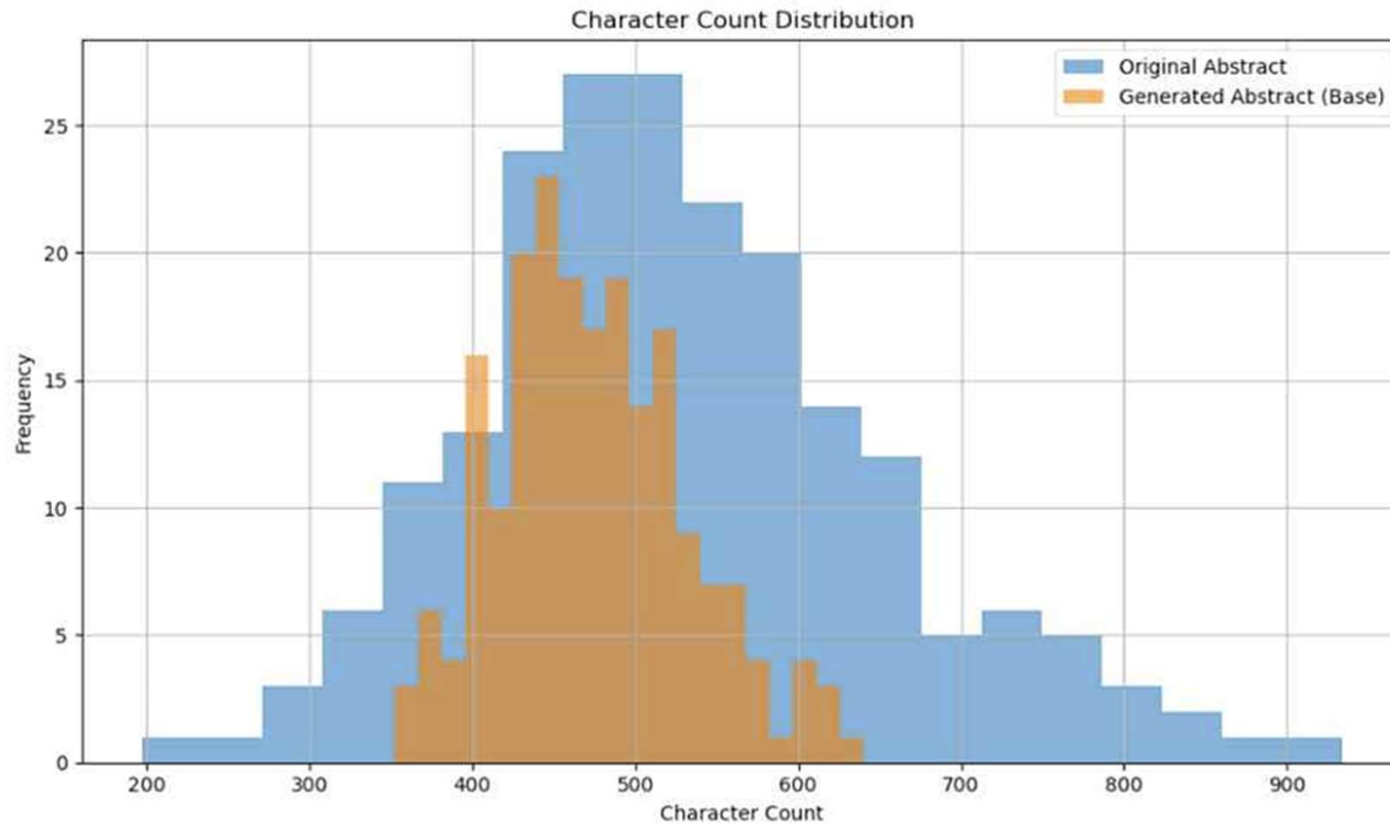
인간초록 vs AI초록의 TF-IDF 코사인 유사도 기준 상위 5편·하위 5편, 총 10편의 논문에 해당하는 인간초록과 AI초록 한 쌍씩을 배정함

평가 항목 이해가능성, 구조, 정보의 선택, 문법·맞춤법, 간결성에 관한 리커트 5점 척도 부여

하나의 논문에 대하여 생성된 인간초록과 AI초록을 구분하고, 근거를 서술하도록 함

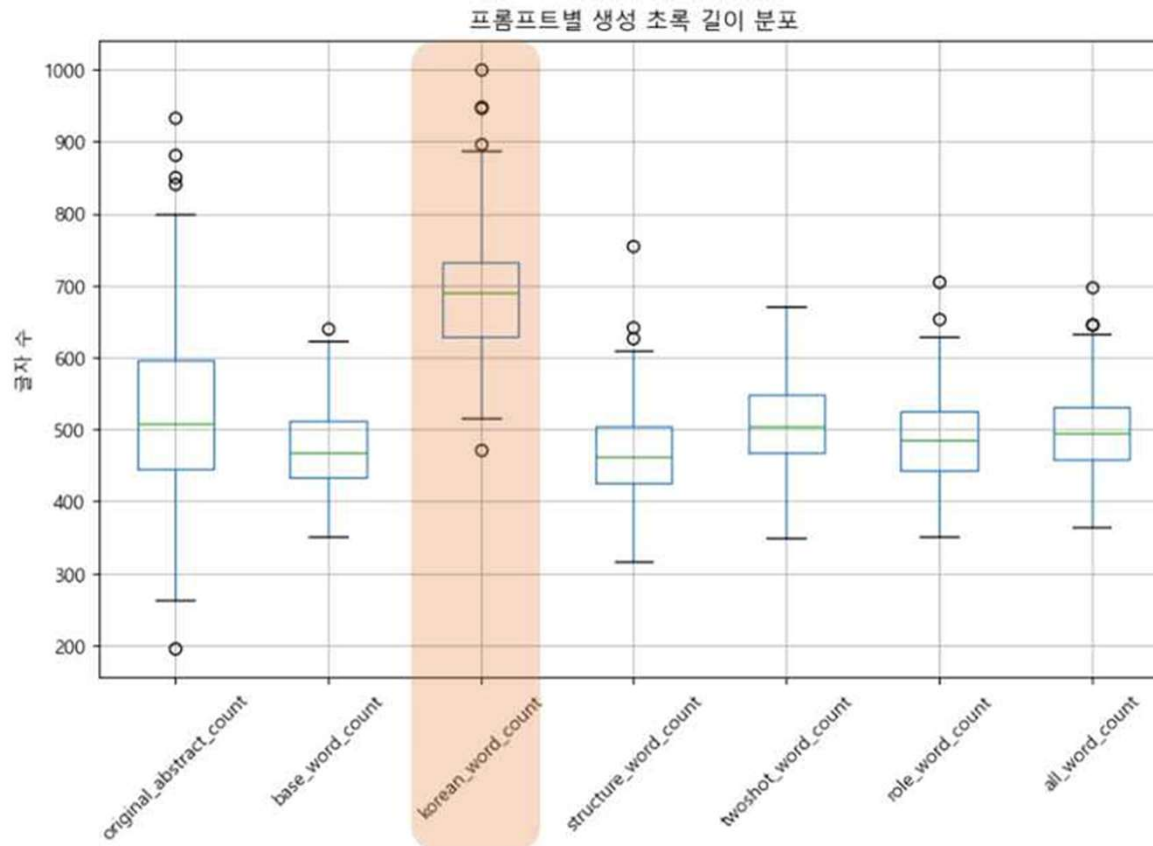
교수 5인을 대상으로 한 서면 면담 → 문항: ChatGPT나 Gemini 등 생성형 AI 활용 경험과 방식, 초록 작성 시 AI 활용의 장점과 한계, 초록 및 원고 작성에서 AI 활용이 허용될 수 있는 범위, AI 활용 사실의 명시 여부 및 표기 방식에 대한 견해

본문 읽기: 연구 결과



〈그림 1〉 원초록과 생성초록(Base 프롬프트) 글자 수 분포

본문 읽기: 연구 결과



<그림 2> 프롬프트 유형별 생성초록 글자 수 분포

본문 읽기: 연구 결과

〈표 2〉 프롬프트 유형별 원초록과 생성초록 간의 유사도 평균 비교

No	프롬프트 유형	원초록 vs. 생성초록		본문과의 비교(TF-IDF 유사도)	
		TF-IDF 유사도	BERTScore	본문 vs. 생성초록	본문 vs. 원초록
1	Base	0.6263	0.8962	0.7087	0.7146
2	Base-Korean	0.6436	0.8922	0.7412	
3	Structure	0.6377	0.8981	0.6982	
4	Few-Shot	0.6329	0.8952	0.7131	
5	Role	0.6331	0.8957	0.7158	
6	All combined	0.6371	0.8973	0.7058	

선행연구와 다른 결과

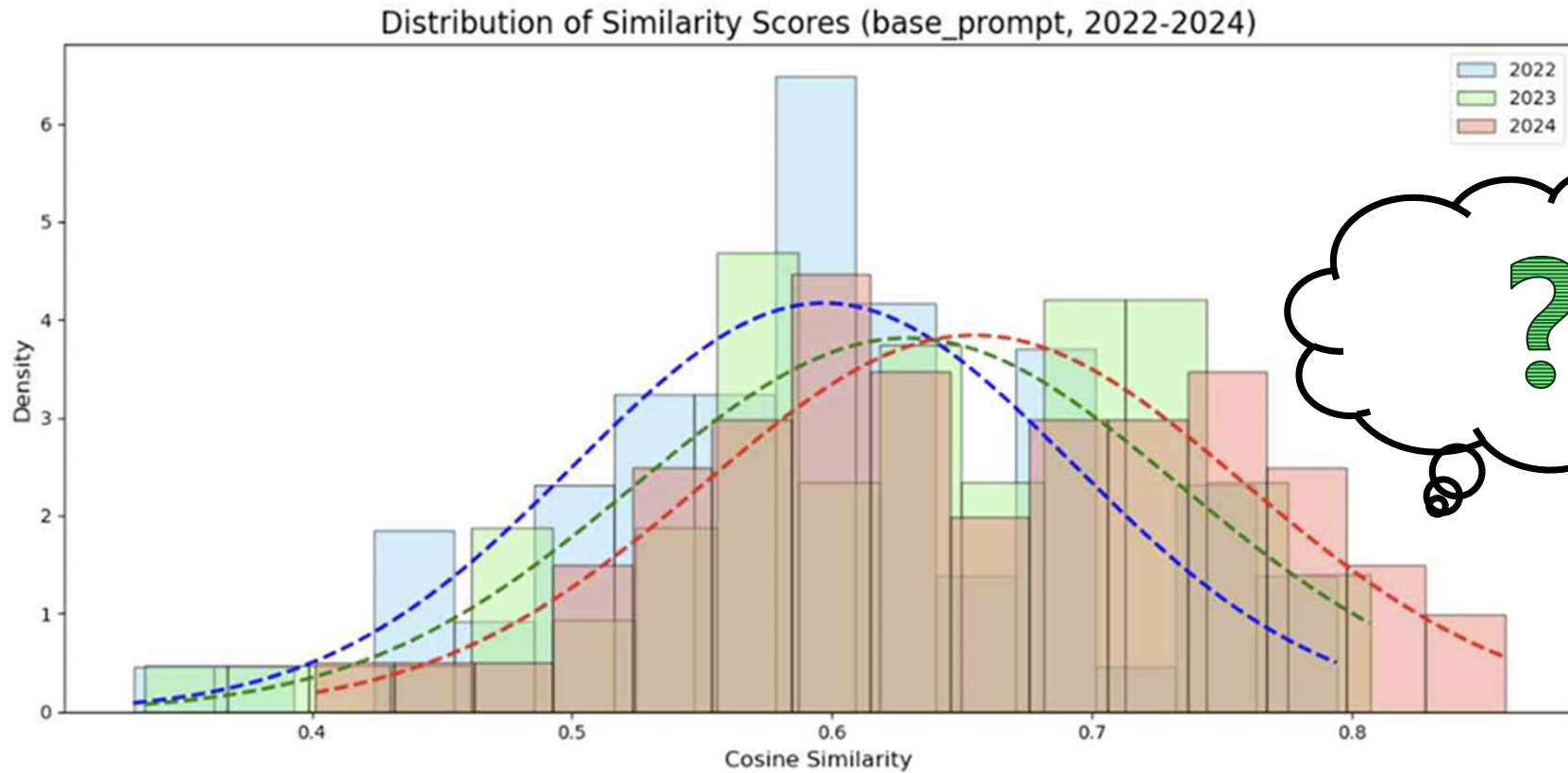
〈표 3〉 연도별 TF-IDF 및 BERTScore 기반 유사도 평균(Base Prompt 기준)

연도	원초록 vs. 생성초록		본문과의 비교(TF-IDF 유사도)	
	TF-IDF 유사도	BERTScore	본문 vs. 생성초록	본문 vs. 원초록
2022	0.597	0.893	0.690	0.697
2023	0.629	0.896	0.715	0.717
2024	0.655	0.900	0.722	0.731

시간이
지날수록
의미 기반
유사도가
상승함

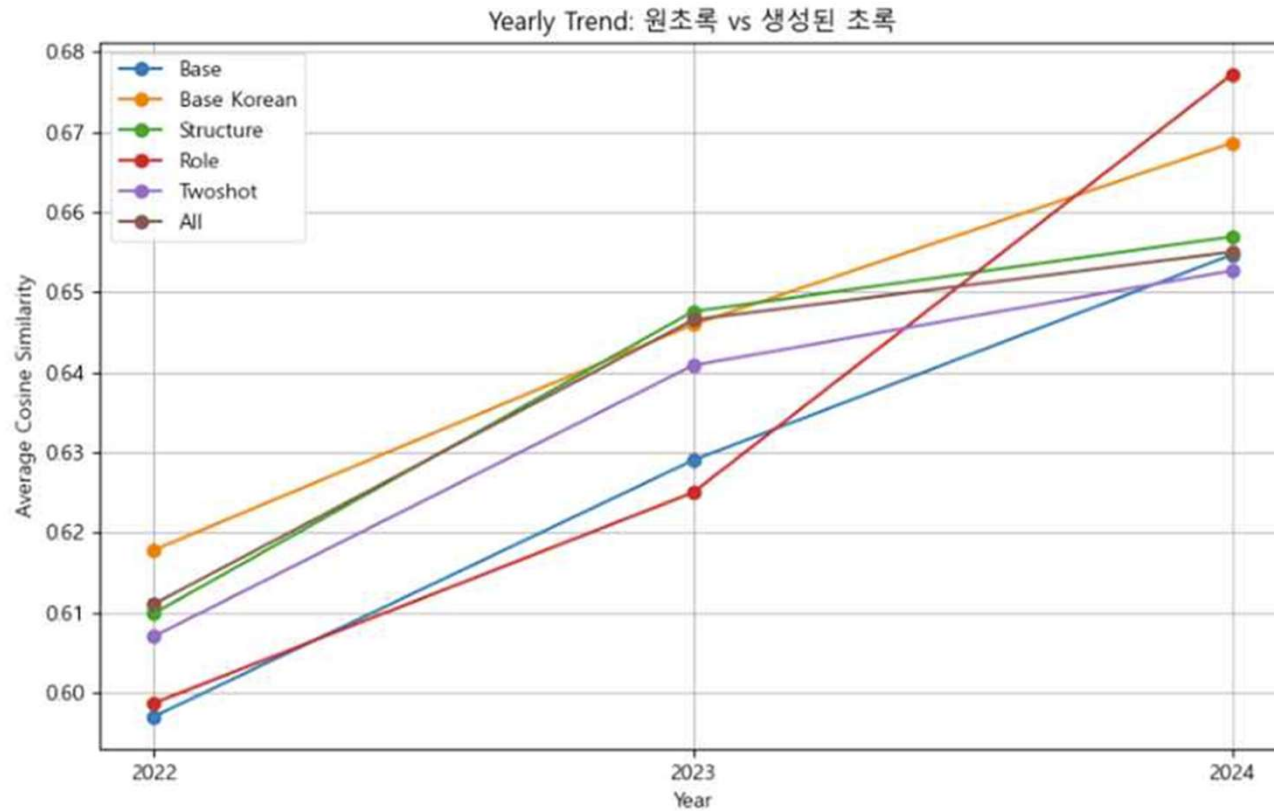


본문 읽기: 연구 결과



〈그림 3〉 연도별 원초록과 생성초록 간 유사도 분포

본문 읽기: 연구 결과



<그림 4> 프롬프트 유형에 따른 원초록과 생성초록 간 유사도 연도별 추이

본문 읽기: 연구 결과_전문가 인식 조사

인간초록 vs AI초록 구분 정확도 60%

-박사과정생 정확도 50%, 교수진 정확도 70%

→연구 경험이 풍부한 교수진이 더 정확히 구분했으나, 전반적으로 인간초록과 AI초록을 식별하기가 쉽지 않음을 나타냄

본문 읽기: 연구 결과_전문가 인식 조사

인간초록과 AI초록을 올바르게 구분한 참가자

-이해가능성, 문법, 간결성 항목에서 AI초록에 인간초록보다 높은 점수를 부여함

→AI초록이 문체적으로 명료하고, 문법적으로 정제돼 있다는 인상을 응답자에게 주었음을 알 수 있음

인간초록과 AI초록을 올바르게 구분하지 못한 참가자

-이해가능성 항목에서 인간초록에 AI초록보다 높은 점수를 부여함

(=인간초록을 AI초록으로 오인하고 AI초록이 이해하기 쉬웠다고 평가한 것)

-문법, 구조, 간결성 항목에서 AI초록에 인간초록보다 높은 점수를 부여함

(=AI초록을 인간초록으로 오인하고 인간초록이 문법, 구조, 간결성 항목에서 낮다고 평가한 것)

본문 읽기: 연구 결과_전문가 인식 조사

AI초록 판단 이유: 올바르게 구분한 응답자

- 정형화된 구조와 포괄적 표현 등을 사용하는 것을 AI초록의 특징으로 판단함
- 구체적인 연구 범위나 방법, 복잡하거나 덜 정제된 문장 구조를 인간 초록의 특징으로 판단함

AI초록 판단 이유: 올바르지 못하게 구분한 응답자

- 각자 기준이 다르거나, 기준을 명확히 설정하지 않은 채로 판단함

본문 읽기: 논의 및 결론

- 후속 연구 제안
 - 한-영 혼합 프롬프트
 - 다양한 모델 간 비교 실험
- 시적 특징을 찾아내려는 인식적 편향이 전문가 인식 조사에 영향을 미쳤을 가능성이 있음

연구의 의의와 한계

의의

- LLM을 활용한 학술논문 초록 작성에 대한 적용 가능성을 학문적·실천적으로 탐색함
- 학문적 시사점: AI초록의 본문에 대한 충실성을 정량적·정성적으로 분석
- 실천적 시사점: 인간-LLM 협업 초록 작성 가능성 제시, 작성 가이드라인 필요성 제기

한계

- 문헌정보학에 한정된 논문을 분석 대상으로 함→타 분야 적용 가능성을 탐색하기에는 무리가 있음
- 전문가 인식 조사에 참여한 표본 집단이 소규모임

토의 질문

1. 선생님들께서는 학술 활동에서 LLM을 어떻게 활용하고 계신가요?

2. 선생님들께서는 읽을 논문을 고를 때, 초록을 적극적으로 활용하시는 편인가요?

→Yes: 그렇다면, 논문을 고를 때에 있어 초록은 어떤 장점을 가지나요?

→No: 그렇지 않으시다면, 어떤 정보를 근거로 논문을 고르시나요?

+논의가 성숙된다면 초록이 갖추어야 할 요건, 논문을 고르는 기준, 좋은 논문은 무엇인지 등을 논의해보아도 좋겠습니다.