

노동정책연구  
2025. 제25권 제3호 pp.139~164  
한국노동연구원  
<http://doi.org/10.22914/jlp.2025.25.3.006>

연구논문

## RAG-LLM 기반 KLIPS 지식 챗봇 개발

민인식\*

본 연구는 한국어 특화 AI 모델인 EXAONE 4.0 모델을 활용하여 한국노동패널(KLIPS) 전문 지식탐색을 위한 검색증강생성(RAG) 챗봇을 개발하였다. 해당 모델을 Google Gemini와 Chat GPT 등 최신 글로벌 LLM과 비교 분석하는 것을 목적으로 한다. KLIPS 공식 문헌으로 구축된 지식베이스와 EXAONE 4.0(1.2B) 모델을 결합한 RAG-LLM 시스템을 설계한 후, KLIPS 사용자의 실제 분석 환경인 Stata에서 사용가능하도록 smart\_klips\_exaone 모듈을 개발하였다. 벤치마크 질의응답 비교 결과, 답변의 범용성·유창성 측면에서 글로벌 모델이 일부 우세를 보였으나, KLIPS 고유 용어의 재현 정확성과 환각 억제 능력에서는 RAG-EXAONE 모델이 우위를 나타냈다. 이는 전문 연구지원 도구의 핵심 효용성이 사실적 정확성과 안전성에 있음을 고려할 때 본 연구에서 제시한 모델이 실제 연구환경에서 실용적 가치를 지닐 수 있음을 시사한다. 또한 본 연구는 지식베이스를 점진적으로 고도화하는 '인간-AI 협업 기반 선순환 구조'를 향후 발전 방향으로 제시하였다.

핵심용어 : 한국노동패널(KLIPS), 검색증강생성(RAG), 대규모 언어모델(LLM), 엑사원(EXAONE), Stata

논문접수일 : 2025년 7월 25일, 심사의뢰일 : 2025년 8월 5일, 심사완료일 : 2025년 9월 1일

\* 경희대학교 경제학과 교수(imin@khu.ac.kr)

## I. 서론

### 1. 연구의 배경과 필요성

한국노동패널(Korean Labor and Income Panel Survey : KLIPS)은 1998년부터 한국의 대표 가구와 그 구성원인 가구원을 매년 추적·조사해 온 대표적인 패널데이터이다. 이를 활용한 자료는 고용, 소득, 교육, 가계 자산 등 노동경제 관련 연구의 기초 자료이자, 정부의 노동·복지 정책 수립의 근거 자료로 사용되고 있다(김용원·김정덕, 2025). KLIPS는 제27차(2024년 조사) wave까지 진행된 장기패널 구조 덕분에 개인·가구 수준의 동태적 변화를 식별할 수 있고, 더 나아가 세대 간 이동성 및 자산 이전을 연구할 수 있는 거의 유일한 패널데이터이다(최지은·홍기석, 2011; 홍성은·이슬기, 2025; 권다운, 2024).

높은 활용과 확장 가능성에도 불구하고 27년 동안 축적된 데이터의 방대한 분량(1~26차 wave만 고려하였을 때, 80개 파일)과 매년 사소하게 변경되는 설문, 코딩 체계로 인해 신규 연구자뿐 아니라 숙련된 연구자도 KLIPS 체계에 대한 이해가 쉽지 않다. KLIPS를 조사·발표하는 한국노동연구원(KLI)에서는 이러한 점을 감안하여 노동패널 자료추출 시스템인 smart\_klips를 웹상에서 제공하고 있다.<sup>1)</sup> 자료추출 시스템은 연구자가 원하는 변수와 차수를 선택하여 하나의 패널자료로 가공하는 시스템이다. 자료추출 시스템은 ‘패널데이터 구축’이라는 목적만 달성할 수 있기 때문에 변수 탐색과 정확한 정의 확인, 조사설계(가중치, 추가 표본) 등을 확인하기 어렵다. 물론 기존 사용자 가이드(User’s Guide)와 FAQ를 통해 정보를 제공하고 있지만 ‘정적 문서’ 형태로만 제공하는 단점이 있다. 따라서 연구자가 원하는 맥락과 수준에 맞추어 즉시 응답하지 못하는 한계가 여전히 존재한다.

연구자들이 일일이 방대한 코드북과 매뉴얼을 확인하거나 정형화된 질의응답에

1) <https://smartklips.kli.re.kr/klips/smartklips>(검색일 : 2025. 7. 10).

의존해야 하는 과정은 분석 초기 단계에서 상당한 시간과 노력을 소모하게 만든다. 2022년 Open AI의 Chat-GPT 출현 이후 대규모 언어모델(Large Language Model : LLM)은 자연어 질의에 대한 동적 응답을 제공함으로써 학술연구 분야에서도 획기적인 변화를 가져오고 있다. 지난 27년 동안 KLIPS의 방대한 지식체계를 이러한 LLM과 통합하여 KLIPS의 다양한 정보(변수 정의, 조사설계, 선행연구 정보 등)를 실시간으로 제공할 수 있는 지능형 연구 보조시스템을 구축·검증할 필요가 있다.

## 2. 연구의 목적 및 차별성

최근 Chat GPT-4o, Gemini 2.5 Pro와 같은 대규모 언어모델(LLM)은 연구를 보조하는 시스템으로서 불가분 관계로 활용되고 있다(Boiko et al., 2023; Luo et al., 2024). 따라서 KLIPS 분석에서도 LLM을 활용하는 것은 자연스러운 흐름이다. 그러나 일반 LLM을 KLIPS 분석에 즉시 투입하기에는 다음과 같은 구조적 제약이 존재한다. 첫째, 환각(hallucination) 현상이다. 가령 존재하지도 않은 변수명, 조사연도, 분석 절차를 마치 사실처럼 제시할 가능성이 있다. 둘째, 지식 시차가 존재한다. 미리 사전학습된 이후 추가된 최신 조사 내용이나 가중치 개편을 반영하지 못한다. 셋째, 근거 결여 가능성이다. 답변의 출처를 명시적으로 인용하지 않아 재현 가능성을 저해한다. 특화된 패널데이터를 다룰 때, 환각·지식 단절은 분석 오류로 직결될 수 있으므로 이를 보완할 수 있는 체계적 접근이 필요하다.

최근 정보검색(retrieval) 모듈을 LLM 앞단에 결합하여 ‘질의 ⇒ 관련 문헌 자동 소환 ⇒ 근거와 함께 답변 생성’ 과정을 수행하는 RAG(Retrieval-Augmented Generation)-LLM 아키텍처가 주목받고 있다(Lewis et al., 2020). RAG 시스템을 통해 외부 지식베이스에 기반해 최신 업데이트된 정보와 도메인 특화 정보를 실시간 주입할 수 있다. 모델이 사용한 근거 문헌을 함께 제공함으로써 출처 투명성을 확보할 수 있다. 국내 데이터 맥락에 맞는 고성능 LLM과 RAG를 결합하여, KLIPS처럼 한국 고유의 노동 관련 용어와 제도적 맥락을 이해하는 AI 연구보조 시스템을 구축하고자 한다.

본 연구에서는 노동경제·사회과학 연구자가 주로 사용하는 Stata 통계패키지에서 RAG-LLM을 활용하는 시스템을 제시하고자 한다. Stata 내부에서 자연어로

질문하고 즉시 코드·해설·출처를 반환하는 명령어를 제공할 수 있다면 연구 생산성을 실질적으로 향상시킬 수 있을 것으로 기대한다.

### 3. 연구방법 및 구성

본 연구에서는 KLIPS에 특화된 RAG-LLM 시스템을 구축하기 위해 사용자 가이드(User's Guide), 코드북, 관련 논문 등을 도메인 지식베이스로 구축한다. 이러한 지식베이스에 기초하여 답변을 생성하는 LLM 모델은 LG EXAONE 4.0을 선택하였다. 마지막으로 RAG-LLM 시스템이 Stata에서 구현될 수 있도록 smart\_klips\_exaone 명령어를 소개한다. 구축된 시스템의 평가를 위해 표준적인 질문셋(set)을 통해 정확성, 환각 가능성, 근거충실도 지표를 제시한다.

논문의 구성은 다음과 같다. 제II장에서는 LLM과 RAG 기술적 원리를 개관하고 전문 도메인 특화 RAG 연구의 국내외 사례를 검토한다. 제III장에서는 EXAONE 기반 RAG 시스템 구축 과정을 구체적으로 설명하고 작동 흐름을 기술한다. 제IV장에서는 모델성능 평가 및 활용 사례를 구체적으로 제시한다. 제V장에서는 KLIPS 데이터 접근성 제고 및 소버린(sovereign) AI 기반 연구로서 정책적 의미를 논의하고 후속 연구과제를 제안한다.

## II. 이론적 배경 및 선행연구 검토

### 1. KLIPS 데이터 활용 연구 동향

1998년 1차 wave로 조사가 시작된 이후 2025년 7월 현재 26차 wave(2023년 조사)까지 공식적으로 공개되었다. 노동패널 홈페이지의 '자료 활용 리스트'에 따르면 1998년 6편에서 시작된 관련 연구는 2023년 222편, 누적 3,592편에 이르고 양적으로 크게 성장하고 있다<sup>2)</sup>. 출판 형태로는 학술지 논문이 1,833편(51%)

2) [https://www.kli.re.kr/board.es?mid=a40606000000&bid=0029&act=view&list\\_no=134237&nPage=1](https://www.kli.re.kr/board.es?mid=a40606000000&bid=0029&act=view&list_no=134237&nPage=1)(검색일 : 2025. 7. 10).

으로 가장 많고, 학회, 토론회 발표문(21%)과 학위논문(14%)의 다양한 방식으로 공개되고 있다. KLIPS 데이터가 학술·교육 양쪽 분야에서 모두 다루어지고 있음을 예상할 수 있다.

주제 측면에서는 가계경제·세대 연구(16.1%), 노동공급·이동(14.6%), 불평등·빈곤(10%) 그리고 임금·소득(9%)이 주를 이루고 있다. KLIPS는 가구 경제, 노동시장 변화, 분배구조 분석에 특화된 패널데이터임을 확인시켜 준다. 최근에는 삶의 질, 주관적 만족도, 정책평가 연구가 빠르게 증가하여 연구 스펙트럼이 사회 정책과 보건 영역까지 확장되고 있다. 방법론 역시 기술통계분석에서 시작하여 패널 고정효과, 이중차분(DID), 이산선택모형, 생존분석이 주로 사용되었고 최근 연구에서는 머신러닝 기반 인과추론, 텍스트 마이닝, 네트워크 분석 등 AI 시대에 걸맞은 접근법이 활용되며 분석방법 다변화가 가속화되고 있다.

KLIPS를 활용한 연구는 양적·질적으로 비약적으로 성장하고 있지만 wave가 추가될수록 변수 변경, 가중치 적용, 데이터 구조 파악과 코드북 검색의 높은 학습 비용이 증가하고 있다. 이와 같이 연구분석 과정에서 숨겨진 장벽을 해소하기 위해 도메인 특화된 RAG-LLM 시스템에 기초한 AI 연구보조 시스템이 필수적이라는 문제의식이 자연스럽게 도출된다.

## 2. 대규모 언어모델(LLM)과 검색증강생성(RAG)

LLM은 방대한 텍스트 데이터로부터 통계적 패턴을 학습하여 인간의 언어와 유사한 텍스트를 생성하고 이해하는 인공지능 모델이다(Bommasani et al., 2021). 트랜스포머(Transformer)<sup>3)</sup> 아키텍처를 기반으로 하는 LLM은 주어진 프롬프트의 다음에 이어질 단어를 확률적으로 예측하는 방식으로 작동하며 이러한 능력은 질의응답, 번역, 요약, 창작 등 다양한 작업에서 뛰어난 성능을 보여주고 있다(Brown et al., 2020). 그러나 서론에서 언급하였듯이 LLM은 사전 학습된 데이터에만 의존하기 때문에 최신 정보나 특정 전문 분야의 지식에 취약하며 사실과 다른 내용을 그럴듯하게 생성하는 환각 현상이라는 약점을 내포하고 있다. 특히 사실 기반의 정확성과 재현성이 중요한 학술연구에서 이러한 단점은 LLM의 직접

3) 문장을 토큰(token)이라는 작은 단위로 나눈 후 모든 토큰 쌍의 관계를 한꺼번에 비교하여 '지금 이 단어가 다른 단어들과 얼마나 관련 있는지' 점수로 계산한다.

적 활용을 제약하는 주요 요인이다.

이러한 한계를 극복하기 위해 검색증강생성(RAG)이 제안되었다. RAG는 답변을 생성하기 전에 신뢰할 수 있는 지식 데이터베이스에서 사용자의 질문과 관련된 최신 정보를 검색(Retrieval)하고 이를 근거 자료로 활용하여 답변 생성을 증가하는 방식이다(Gao et al., 2023). RAG-LLM 접근법은 정확성과 최신성이 요구되는 여러 전문 분야에서 빠르게 도입되고 있다. 금융 분야에서는 뉴스, 공시, 소셜 미디어 등을 RAG의 지식 소스로 활용하여 금융시장의 정성적 정보를 분석하고 투자 심사를 예측하는 연구가 활발하다(Zhang et al., 2023). 법률 및 의료 분야에서 방대한 판례나 최신 의학 논문을 기반으로 사용자의 질의에 정확한 답변을 제공하는 시스템이 개발되고 있다(Wiratunga et al., 2024).

사회과학 분야에서도 RAG-LLM의 활용 가능성은 점차 주목받고 있다. Schilde and Heppner(2024)는 독일의 감정평가 데이터베이스를 벡터 DB로 인덱싱한 후 GPT-4o+RAG를 구축하여 전문가 질의에 근거·출처를 포함한 답변을 제공한다. Sun et al.(2024)은 사회적 논쟁 트윗+뉴스 콘텐츠를 복합 검색한 후 LLM에 주입하여 정책 메시지에 대한 미래 댓글 분포 정확도를 10% 이상 향상시킬 수 있다고 분석하였다. Yu and Chen(2024)은 홍콩의 법령, 행정지침 80만 건을 다중 인덱스로 구축하여 민원 Q&A에서 기존 LLM 대비 사실 일치율과 설명 가능성을 동시에 개선할 수 있음을 보였다. Zhu et al.(2024)은 미국의 중환자 데이터 셋(약 4만 건 입원사례)에서 사망위험 및 재입원 예측 실험을 하였다. 환자 행태 연구에서 흔히 직면하는 텍스트, 수치 혼합 데이터 해석 문제를 해결하기 위해 RAG 전략을 제시하였고 외부 지식을 통한 환각 억제 설계를 통해 정책·임상 의사결정의 예측 정확도를 높여 사회적 비용 절감에 기여할 수 있다고 논의하였다.

최근 국내에서도 RAG-LLM 시스템 활용 가능성에 주목하고 있다. 홍지윤 외(2025)는 규칙 기반 금융대출 심사의 한계를 극복하기 위해 외부 금융문서를 검색 증강으로 주입한 LLM 기반 심사 시스템을 설계 및 구현하고 고객 맞춤형 설명과 행동 피드백을 자동 생성하도록 하였다. 실제 고객 데이터를 활용한 실험에서 기존 방식 대비 정보검색, 텍스트 생성 정확도를 유의하게 향상시킬 수 있음을 보였다.

이광우·김수균(2024)은 국내 기업의 기밀 보호와 사실성 확보 문제를 해결하기 위해 한국어 문장 임베딩 기반 지식 DB+최적화 검색+폐쇄망 운용으로 구성된 RAG 질의응답 시스템을 설계·구현하였다. 사용자가 지식 DB를 직접 관리해 최

신 업무 정보를 즉시 반영할 수 있고, 폐쇄망 환경에서 동작함으로써 정보 유출 위험을 최소화해 기업 현장에서 생성형 AI 도입을 실질적으로 지원할 수 있음을 보였다.

한선의 · 이균희(2025)는 신용카드사의 주요 수익원인 할부 이용 여부를 예측하기 위해 RAG-LLM을 제안하였다. 소규모 · 신규 고객 집합에서는 RAG-LLM이 기존 머신러닝 방법론에 비해 탁월한 예측력과 고객 특성 해석력을 동시에 제공함을 입증하였다.

### 3. 본 연구의 차별적 요소

국내외 다양한 전문 분야에서 RAG-LLM의 효용성을 입증하는 연구가 활발하게 진행되고 있다. 본 연구는 다음과 같은 세 가지 측면에서 기존 선행연구와 차별화 지점이 존재한다.

첫째, 도메인 지식의 특수성이다. 기존 사회과학 분야 관련 연구에서는 뉴스, 소셜미디어, 법령, 행정문서 등 사회 현상에 대한 내용적 정보를 담고 있는 비정형 데이터를 지식베이스로 활용하였다. 반면 본 연구에서는 KLIPS 사용자 가이드, 코드북, 방법론 보고서 등 절차적 · 기술적 정보를 지식베이스로 삼는다. 연구자들이 ‘KLIPS 데이터를 어떻게 정확하게 사용할 것인가’라는 근본적인 질문에 답하기 위함이다. 복잡한 구조를 가진 특정 패널데이터의 활용 방법론 자체를 지원하는 RAG-LLM 시스템을 구축하고 그 효용성을 검증하는 것은 ‘연구방법론’ 지원이라는 새로운 영역을 개척한다.

둘째, 한국어 특화 LLM의 전략적 선택이다. 본 연구에서는 LG AI 연구원이 개발한 한국형 프론티어 모델인 EXAONE 4.0을 핵심 엔진으로 채택한다.<sup>4)</sup> 이는 KLIPS에 포함된 ‘실업급여’, ‘육아휴직’, ‘국민기초생활보장법’ 등 한국의 노동 및 복지 제도와 관련된 전문 용어의 미묘한 뉘앙스를 보다 정확하게 포착할 수 있을 것으로 기대한다. 더 나아가, 이는 국내 기술로 개발된 ‘소버린 AI’ 모델이 전문화된 학술연구 분야에서 글로벌 모델과 대등한, 또는 그 이상의 성능을 보일 수 있는지 검증하고자 한다.

4) “AI에 진심인 LG, 오픈 AI보다 먼저 하이브리드 AI 모델 공개”, <https://v.daum.net/v/20250715100228796>(검색일 : 2025. 7. 15).

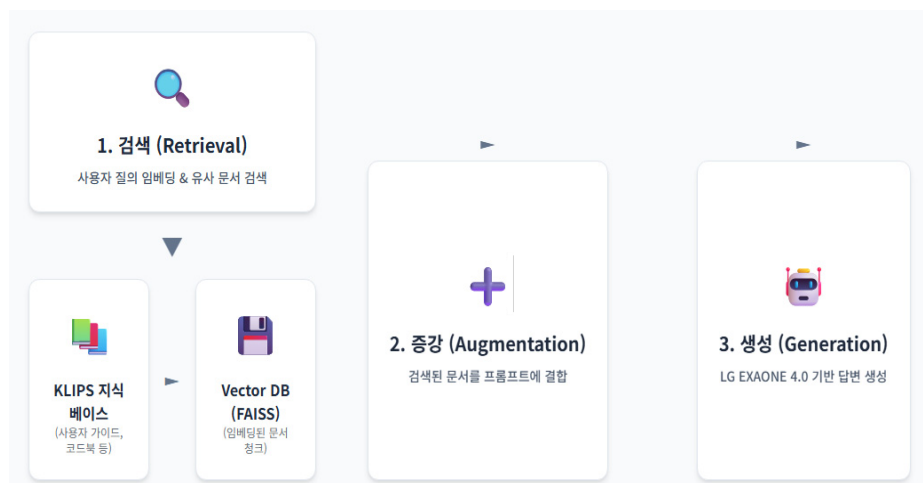
셋째, 연구 워크플로와 직접적인 통합이다. 기존 많은 RAG 시스템은 웹 기반 채팅 인터페이스나 별도의 API 형태로 제공된다. 따라서 연구자가 분석 작업 중 필요할 때마다 작업환경을 벗어나 정보를 찾아야 하는 불편함이 존재한다. 사회과학 연구자들이 가장 널리 사용하는 Stata 통계패키지 내에서 직접 RAG-LLM을 호출할 수 있는 모듈을 구현함으로써 이러한 단점을 극복하고자 한다. 이는 AI 연구 보조시스템을 단순히 참고 도구가 아니라 실제 데이터 분석 과정과 유기적으로 결합한 내장형(built-in) 지원 도구로 전환시키는 실용적 혁신이다.

### Ⅲ. 연구의 설계 및 시스템 구축

#### 1. RAG 지식베이스 구축

앞서 논의했듯이 본 연구에서 제안하는 KLIPS 지식 챗봇은 RAG 아키텍처를 기반으로 한다. RAG와 LLM이 결합되는 과정은 [그림 1]에서 도식화하였다. 구축된 지식베이스에서 검색하여 근거 자료를 LLM에 보내면 EXAONE 4.0에서 근거 자료와 LLM 생성 능력을 결합하여 답변을 완성하게 된다.

[그림 1] RAG-LLM 도식화



1단계에서 구축하는 지식베이스의 품질과 이를 효율적으로 검색하는 능력에 따라 RAG 시스템의 성능이 결정된다. 따라서 KLIPS와 관련된 가장 신뢰도 높은 공식 문헌과 핵심 연구자료를 선별하여 지식베이스를 구축하고자 한다. 본 연구에서 수집된 KLIPS 관련 자료와 문서는 <표 1>에서 정리한다.

수집된 데이터는 거의 대부분 KLIPS 홈페이지에 게재된 공식 자료를 사용하였다. 해당 문헌들은 KLIPS 데이터의 거의 모든 측면을 포괄하고 수천 페이지 분량으로 시스템이 광범위한 질의에 대응할 수 있는 기초를 형성한다. 노동패널 홈페이지에서 찾을 수 있는 공식 문서이고 pdf와 json 문서 형태로 변환하여 사용하였다.<sup>5)</sup>

수집된 원본 문서는 그 상태 그대로 사용하는 것은 불가능하고 LLM의 제한된 컨텍스트 창에 맞춰 효율적으로 검색이 가능하도록 작은 텍스트 조건으로 분할하는 청킹(chunking) 과정을 거쳐야 한다. KLIPS 관련 문서는 일반적인 서술 외에도 변수 목록, 코드 값 등 구조적인 정보가 혼재되어 있으므로 그러한 구조를 고려한 청킹 과정이 필요하다. 청킹 과정의 유효성(validity)을 확보하기 위해 학습자료를 pdf와 json 파일로만 저장하였다. 특히 excel 파일은 하나의 row가 하나의 정보를 가질 수 있도록 수정한 후 json 파일로 변환하는 과정을 거쳤다.<sup>6)</sup>

충분한 문맥을 확보하면서 검색 정확도를 높이기 위해 청크 크기는 1,000자(코드북, 변수 설명 파일에서는 300자), 청크 간 중첩은 200자(코드북, 변수설명 파일에서는 50자)로 설정하였다.<sup>7)</sup> 이는 하나의 청크가 지나치게 많은 정보를 담아 핵심 의미가 희석되는 것을 방지하고 청크 경계에서 문맥이 끊기는 현상을 최소화하여 의미적 연속성을 확보하는 데 필요하다.

지식베이스 구축의 2번째 단계에서는 분할된 청크를 컴퓨터가 이해할 수 있는 수치벡터로 변환하는 임베딩(embedding) 과정을 거치게 된다. 임베딩은 텍스트의 의미론적, 문법적 정보를 벡터 공간에 투영하는 과정으로 임베딩 함수는 각 청크를  $d$  차원의 벡터  $v_i$ 로 변환한다. 본 연구의 임베딩에서는  $d = 768$ 차원의 벡터

5) 17개 파일, 3,015페이지, 98만 단어, 390만 글자(공백 포함)의 분량을 담고 있다.

6) 원본이 pdf인 경우는 pdf 형식을 그대로 사용하였고, klips\_faq.json, klips\_crawled.json은 저자가 파이썬 크롤링 코드를 따로 작성하여 지식베이스 문서로 완성하였다. 또한 코드 등 엑셀파일 역시 json 파일로 변환하는 파이썬 코드를 활용하여 생성하였다.

7) Gao et al.(2023)과 Jain and Goyal(2024)에서 청킹의 중요성과 상충관계 그리고 중첩 비율을 참고하였다.

가 생성된다. 한국어로 작성된 KLIPS 문헌의 복잡하고 전문적인 의미를 정확하게 벡터화하기 위해서 다양한 한국어 자연어 처리에서 높은 성능이 입증된 “jhgan/ko-sbert-nli”을 선택하였다.<sup>8)</sup>

prompt에 속하는 사용자의 질문 벡터  $v_q$ 와 가장 유사한 문서 벡터  $v_i$ 를 찾기 위해서는 코사인 유사도(cosine similarity)를 계산한다. 코사인 유사도는 두 벡터가 이루는 각도의 코사인 값으로 1에 가까울수록 의미적으로 유사함을 의미한다. 질문 벡터  $v_q$ 와 코사인 유사도 점수가 가장 높은 문장 벡터  $v_i$ 를 찾아서 사용자 질문에 가장 관련성 높은 정보라고 판단한다. 따라서 RAG 시스템은 해당 문장을 LLM에게 전달하게 된다. 지식베이스의 모든 문장은 임베딩된 벡터 데이터베이스로 저장된다. Facebook AI에서 개발한 FAISS(Facebook AI Similarity Search) 형식으로 저장하였다.

〈표 1〉 KLIPS-RAG 구축을 위한 지식베이스 문서

파일 형식	문서명	문서 내용(자료시점)
pdf	1~26차년도 유저가이드	KLIPS 1~26차 유저가이드(2024년)
	1~26차년도 통합설문지 (가구용, 개인용)	1~26차 설문지 : 가구용과 개인용 (2024년)
	1~26차년도 통합설문지 (부가조사용)	1~26차 설문지 : 부가조사용 (2024년)
	2017-무응답처리_처리_연구_최종보고서	무응답처리 연구보고서(2017년)
	패널자료품질개선	패널자료 품질개선 연구보고서 VII (2017년)
	표준 추가 연구용역	표본 추가 연구용역 최종보고서 (2009년)
	패널자료품질개선_X	패널자료 품질개선 연구보고서 X (2020년)
	smart_klips_v4_help_file	smart_klips_v4 모듈 도움말 파일
	2023년 KLIPS 기초분석 보고서	26차 KLIPS 기초분석 보고서 (2024년)
	KLIPS 가중치연구 2013	한국노동패널 가중치 연구(2013년)

8) <https://huggingface.co/jhgan/ko-sbert-nli>(검색일 : 2025. 7. 10)에서 관련된 기술문서를 확인할 수 있다.

〈표 1〉의 계속

파일 형식	문서명	문서 내용(자료시점)
json	crawled_data	KLIPS 홈페이지 전체 목차와 내용을 크롤링한 파일(2025년)
	klips_faq	KLIPS 홈페이지에서 제공하는 32개 FAQ 질문과 답변(2025년)
	klips_codebook_revised_직업력	1~26차 코드북 중 직업력 변수와 변수 설명(2024년)
	klips_codebook_revised_개인용	1~26차 코드북 중 개인용 변수와 변수 설명(2024년)
	klips_codebook_revised_가구용	1~26차 코드북 중 가구용 변수와 변수 설명(2024년)
	KLIPS_자료활용리스트	KLIPS 홈페이지에서 제공하는 KLIPS 데이터 활용 학술문헌(2025년)
	smart_klips_v4_가공변수_코드북	smart_klips_v4 모듈에서 사용하는 가공변수 코드북

## 2. RAG-LLM 시스템 설계

앞서 구축한 KLIPS 데이터베이스는 [그림 1]에서 ‘검색’ 단계에 해당하는 외부 기억장치 역할을 한다. 본 절에서는 이러한 검색 모듈과 LLM을 결합해 최종적으로 사용자의 질문에 답변하는 ‘생성(generation)’ 단계를 설명한다. 이 과정에서 어떤 LLM을 사용할 것인지는 시스템 전체의 성능과 정체성을 결정하는 핵심적인 설계 변수이다.

본 연구에서는 OpenAI의 Chat-GPT 4o 또는 Google Gemini와 같은 글로벌 모델 대신 LG AI 연구원이 2025년 7월 공개한 한국어 특화모델인 EXAONE 4.0을 핵심 LLM으로 채택하였다.<sup>9)</sup> 현재 EXAONE 4.0은 32B(320억 개 매개변수) 모형과 1.2B(12억 개 매개변수) 모형이 공개되어 있다. 본 연구에서는 개인용 CPU를 사용하는 점을 감안하여 소규모(mini) 모형인 1.2B를 선택하였다.

9) <https://huggingface.co/LGAI-EXAONE/EXAONE-4.0-1.2B>(검색일 : 2025. 7. 10)에서 관련된 기술문서를 확인할 수 있다.

〈표 2〉에서는 LG EXAONE과 유사한 한국형 AI 모델을 비교하고 있다. 특히 LG AI와 SKT AI 모델은 2025년 8월 초 국가대표 AI에 선발되어 그 우수성을 인정받았다.<sup>10)</sup>

EXAONE 4.0을 선택한 전략적 근거는 다음과 같다. 첫째, 한국어 및 도메인 특화 성능의 우수성이다. 글로벌 LLM은 방대한 영어 데이터를 중심으로 학습되었지만 EXAONE은 대규모의 고품질 한국어 데이터를 기반으로 학습되어 있다. 따라서 KLIPS에 등장하는 한국 노동시장의 고유한 용어와 제도적 차이를 탁월하게 이해할 수 있다. 따라서 고도로 전문화된 KLIPS 지식도메인을 다루는 본 연구의 목적에 부합한다.

둘째, EXAONE은 접근성에서 장점이 있다. 글로벌 LLM은 고가의 상용 API 사용료를 지급해야만 작업이 가능하다. 그러나, EXAONE의 경우 2025년 7월 현재 허깅페이스(huggingface)에 오픈 모델로 공개하였다. 따라서 KLIPS를 관장하는 노동연구원 입장에서도 비용 부담 없이 KLIPS RAG-LLM 시스템 개발에 집중할 수 있다.

〈표 2〉 한국형 AI 모델 비교

AI 모델	특징	국가대표 AI 선발 여부
LG EXAONE	한국어 특화 초거대 언어모델, 전문가용 모델을 지향하며, 대규모 한국어 데이터 학습을 통해 높은 수준 언어 이해 및 생성 능력	선발
KT 믿음	KT가 개발한 초거대 언어모델, 환각 현상 최소화에 중점, B2B 및 공공분야에 특화	탈락
카카오 KANANA	GPT-3 모델에 기반하여 문맥 기반과 자연스러운 문장 생성능력 우수	탈락
SKT 에이닷	다양한 기업 및 학계와 협력하여 한국어와 문화에 특화된 AI 기술	선발

다만 글로벌 생태계를 가진 LLM에 비해 EXAONE은 상대적으로 사용자 커뮤니티

10) [https://www.chosun.com/economy/tech\\_it/2025/08/05/QNNBN6NYYRG27IULDUFVGA6GX1/](https://www.chosun.com/economy/tech_it/2025/08/05/QNNBN6NYYRG27IULDUFVGA6GX1/)  
(검색일 : 2025. 8. 5).

니티나 서드파티(third-party) 도구가 많지 않으며, 본 연구에서 선택한 EXAONE 4.0-1.2B는 전문가 모델(32B)보다 장문 추론 및 복합 연산에서 답변 깊이가 떨어질 가능성이 있다.

### 3. Stata 연동 명령어 개발

제II장 제3절에서 논의하였듯이 본 연구의 독창적 기여 중 하나는 RAG-LLM 시스템이 Stata 내에서 구현되도록 Stata 명령어를 같이 개발하여 제공하고 있다는 것이다. 대부분 사회과학 연구자는 Stata, R, SPSS 등 특정 통계패키지를 주된 연구분석 환경으로 사용한다. 연구 과정에서 도움을 받기 위해 분석 프로그램, 웹 브라우저, 문서 뷰어를 반복적으로 오가는 것은 연구 흐름을 방해하고 비효율성을 유발할 수 있다. 본 연구에서 제안한 AI 연구보조 시스템의 효용성을 극대화하기 위해 연구자의 기존 워크플로에서 벗어나지 않고 분석 과정 내에서 직접 상호작용할 필요가 있다.

이러한 문제의식을 바탕으로 Stata 내에서 직접 KLIPS 지식 챗봇을 호출할 수 있는 명령어 `smart_klips_exaone`을 개발하였다. 이 명령어의 prompt 옵션에서 '자연어 질문' 형식으로 입력하면 LLM 시스템과 통신하여 얻은 답변 결과를 Stata 결과 창에서 바로 출력해 주는 인터페이스 역할을 수행한다. 이를 통해 연구자는 데이터 처리나 통계분석 중 궁금한 점이 생겼을 때 작업 맥락을 유지한 채 즉시 AI의 도움을 받을 수 있다.

기술적으로 이러한 통합은 Stata 17버전 이후 Python과의 연동 기능 덕분에 가능하다. 본 시스템의 핵심인 RAG 로직인 FAISS 벡터 데이터베이스 검색, EXAONE 4.0 모델과의 통신 등은 파이썬 AI 관련 라이브러리를 활용해야 한다. ado 파일 내에서 파이썬 블록(block)을 실행하고 답변 결과를 다시 Stata로 내보내서 결과 창에 표시하게 된다(〈표 3〉 참고). 이러한 하이브리드 접근 방식은 Stata의 강력한 데이터 관리 및 계량분석 기능과 AI 개발에 최적화된 파이썬 생태계를 동시에 활용하여 각 언어의 장점을 극대화할 수 있다. 특히 파이썬 코딩에 익숙하지 않은 Stata 사용자들이 별도의 학습 없이 최신 AI 기술의 혜택을 온전히 누릴 수 있어 연구의 대중화와 생산성 향상에 기여할 수 있다.

〈표 3〉 smart\_klips\_exaone 프로세스

단계	주체	역할
질의 입력	Stata 사용자	smart_klips_exaone 명령어 실행
검색	Python RAG 시스템	질문을 받아 KLIPS 벡터 DB에서 관련 문서 검색
생성	Exaone 4.0	검색된 문서와 질문을 바탕으로 최종 답변 생성
결과 반환	Stata 사용자	Stata 결과 창에 답변 및 출처 확인

## IV. 모델 성능 및 활용

### 1. 평가방법 설계 및 결과

본 절에서는 제Ⅲ장에서 설계 및 구축한 smart\_klips\_exaone 시스템의 성능을 객관적으로 평가하고자 한다. 평가 대상은 EXAONE 4.0 기반 RAG 시스템과 검색증강 기능이 없는 글로벌 최상위 LLM 기반 모델과 직접 비교하는 방식으로 진행된다. 이를 통해 검색증강 & 한국어 특화 LLM이 KLIPS와 같은 전문 도메인에서 가지는 상대적 우위와 실효성을 검증하고자 한다.

평가를 위한 질의 프롬프트는 〈표 4〉에서 6가지 유형을 예시적으로 설정하였다.

〈표 4〉에 제시된 prompt를 그대로 3개 모델에 동일하게 사용한 결과는 〈표 5〉에서 정리한다. 분석결과, RAG-EXAONE 모델은 전반적으로 안정적인 성능을 보였다.<sup>11)</sup> 도메인 특화 용어의 정확성과 환각 억제 측면에서 RAG-EXAONE 모델의 비교우위가 관찰된다. 특히, FAQ, 변수 정의 그리고 한국 제도 맥락에서 안정적이다. 반면 GPT-4o는 대부분의 항목에서 부족한 답변이 관찰되고 특히 KLIPS 코드북에 없는 변수 이름을 임의로 사용하는 현상이 대부분 답변에서 나타난다. 그러나, 이러한 환각 현상은 Google Gemini에서는 상대적으로 나타나지 않는다는 것을 확인하였다. 그러나 절차·방법론(가중치) 관련 질문은 세 모델 모두 취약하다는 것을 보여준다. 문서 태그 기반이라 수식에 대한 검색이 미흡하다.

11) smart\_klips\_exaone 답변 결과는 부록에서 구체적으로 제시한다.

추후 CoT(Chain of Thoughts) 프롬프트 보강이 필요한 것으로 예상된다.

〈표 4〉 벤치마크 질문 유형 구성

질문 번호	유형	예시 질문
Q1	FAQ 확인형	KLIPS 홈페이지에 제시된 FAQ에 대한 답변 확인: 노동패널자료의 정규/비정규직 비중은 왜 통계청에서 발표하는 수치와 다를까?
Q2	사실 확인형	특정 변수명, 변수 정의를 묻는 질문: 가구소득을 구성하는 세부 항목과 변수명을 알려주세요.
Q3	절차 안내형	데이터 처리 방법이나 가중치 적용 절차: 추가 표본의 가중치 결정에 대해 설명해 주세요.
Q4	개념 비교형	한국적 맥락 개념: '국민기초생활보호대상자'에 대한 지원금은 어떤 소득 항목에 포함되는가?
Q5	정보 종합형	여러 문서에 흩어진 정보를 종합적으로 판단: 장애인 가구 식별을 위해 적용되는 변수(개인 기준)가 언제부터 도입되었나요? 그리고 횡단 가중치 설정 방법은?
Q6	환각 유도형	존재하지 않는 변수나 잘못된 정보를 포함한 질문: 27차 데이터에 있는 h_sat 변수에 대한 정의는?

종합하면, RAG 아키텍처를 통해 전반적인 환각 현상은 크게 감소하였다. 또한, KLIPS라는 특정 도메인의 전문 용어와 한국적 맥락을 정확히 이해하고 재현하는 능력은 EXAONE 기반 모델이 글로벌 모델 대비 충분히 견줄 만하다. 연구 목적에 따라 RAG-LLM의 강점을 전략적으로 선택해야 할 필요성을 보여준다.

〈표 5〉 벤치마크 질문에 대한 LLM 답변 내용

번호	LLM	주요 평가
Q1	RAG-EXA	해당 질문에 정답은 명확하게 KLIPS FAQ에 제시되어 있음. <sup>12)</sup> 3개 모델의 답변과 FAQ 정답을 AI가 비교한 결과 RAG-EXA가 정답과 가장 유사하고 Chat-GPT가 정답과 거리가 멀다고 판단함.
	Chat-GPT	
	Gemini	
Q2	RAG-EXA	가구소득 세부 항목을 정확히 구분해 줌. 다만 각 변수에 대한 영문변수명을 제시하지 않음. prompt의 '변수명'을 한글 변수명으로 이해한 것으로 예상됨.
	Chat-GPT	KLIPS 코드북에 제시된 구분을 사용하지 않음. 코드북에는 없는 "사업소득"이라는 표현을 사용함.
	Gemini	가구소득 세부항목을 정확히 구분해 줌. 다만 영문변수명은 임의로 지정함.

〈표 5〉의 계속

번호	LLM	주요 평가
Q3	RAG-EXA	가중치에 대한 표면적인 답변만 제시함.
	Chat-GPT	가중치 설정에 대한 구체적인 답변 없음. 가중치 변수 이름 잘못됨.
	Gemini	가중치에 대한 표면적인 답변만 제시함.
Q4	RAG-EXA	'정부보조금 이외로 별도로 표시'된다고 답변함.
	Chat-GPT	KLIPS 코드북에서는 사용하지 않은 "공적이전소득" 표현 사용함. 코드북에 없는 변수 이름을 제시함.
	Gemini	"공적이전소득" 표현 사용함. 코드북에 없는 변수 이름을 제시함.
Q5	RAG-EXA	'해당 내용 없음'이라고 답변함.
	Chat-GPT	11차 조사라고 잘못 답변함. 횡단 가중치에 대한 표면적인 답변만 제시함.
	Gemini	장애여부 조사 시점을 알 수 없음. 횡단 가중치에 대한 객관적 설명 제시함.
Q6	RAG-EXA	환각현상 전혀 없음.
	Chat-GPT	27차 조사에 h_sat 변수가 있는 것처럼 답변함.
	Gemini	27차 조사는 공개 예정이고 h_sat 변수는 "존재 가능성"을 언급함.

주 : 1) smart\_klips\_exaone의 답변을 얻는 시간은 개인 노트북의 스펙에 따라 달라질 수 있으며, 저자의 노트북에 따르면 평균 1개 질문당 104초 정도 소요.

2) Chat-GPT-4o 모형, Gemini는 2.5 Pro 모형을 선택.

## 2. smart\_klips\_exaone 활용

본 연구에서 제시하는 RAG-LLM 성능이 완벽한 것은 아니지만 '정확성'과 '안정성'이라는 강점이 있다. 그러나 그 한계를 인정하고 이를 smart\_klips\_exaone을 통해 지속적으로 발전시켜 나가는 선순환 구조를 제안하고자 한다.

먼저 Stata에서 RAG-LLM을 실행하기 위해서는 다음과 같은 3개 파일을 다운

12) [https://www.kli.re.kr/board.es?mid=a40702000000&bid=0015&act=view&list\\_no=127925&tag=&nPage=1](https://www.kli.re.kr/board.es?mid=a40702000000&bid=0015&act=view&list_no=127925&tag=&nPage=1)(검색일 : 2025. 7. 10).

로드 받아서 설치해야 한다.<sup>13)</sup> 1, 2번 파일은 <표 1>에서 제시한 KLIPS 문서 자료를 청킹·임베딩한 내용을 저장하고 있다. 3번 파일은 RAG-LLM을 Stata 환경에서 실행하도록 만든 ado 파일이다. <표 6>의 3개 파일을 특정 폴더(윈도우즈 OS 기준, c:\ado\s 폴더)에 복사하여 저장한다.

<표 6> 실행을 위한 설치 파일 리스트

파일번호	파일명	파일 내용
1	index.pkl	피클(pickle) 파일로 index.faiss와 원본 텍스트 청크를 연결해 주는 mapping 역할
2	index.faiss	임베딩 결과가 저장된 인덱스 파일
3	smart_klips_exaone.ado	Stata 명령어 실행파일

본 시스템을 활용하여 정확한 변수 탐색, 신뢰도 높은 코드값 확인 그리고 연구 오류 방지를 위한 사실관계 교차 검증 등 연구의 전 단계에 걸쳐 실질적인 도움을 어느 정도 받을 수 있다. 그러나 <표 5>에서 확인할 수 있듯이 현재 시스템은 여러 문서에 흩어진 정보를 종합하여 추론하는 데에는 한계가 보인다. 이는 초기 지식 베이스가 원본 문서의 명시적 정보에 의존하기 때문이다. 이러한 한계를 극복하기 위해 smart\_klips\_exaone 자체를 데이터 수집 및 피드백 도구로 활용하여 지식 베이스를 점진적으로 고도화하는 도구로 활용하고자 한다.

다음과 같은 가상의 연구주제를 설정하고 이를 위해 <표 7>과 같이 10개의 smart\_klips\_exaone 명령문을 설정한다. 그 질문과 답변을 수집하여 엑셀파일(rag\_update1.xlsx)로 저장한다.

연구주제 : 26차(2023년) 조사 데이터를 이용하여 개인의 삶의 만족도와 주당 근로 시간의 관계를 분석하고자 한다.

13) 3개 파일은 논문심사 완료 후 KLIPS 연구자 누구나 사용할 수 있도록 공개 예정이다.

〈표 7〉 10개 질문과 답변을 수집하는 Stata 코드

```

smart_klips_exaone, prompt("26차(2023) 개인용 데이터에서 주당 근로시간 변수명을 알려줘")
local answer1=r(Answer)
local prompt1=r(Prompt)
smart_klips_exaone, prompt("26차(2023) 개인용 데이터에서 '삶의 만족도' 변수명을 알려줘")
local answer2=r(Answer)
local prompt2=r(Prompt)
smart_klips_exaone, prompt("'삶의 만족도' 변수의 응답 범주는 어떻게 구성되어 있고, 각 코드값은 무엇을 의미하나요?")
local answer3=r(Answer)
local prompt3=r(Prompt)
smart_klips_exaone, prompt("26차 데이터에서 분석에 필요한 통제변수인 '성별', '가구주 연령', '교육수준'의 변수명을 찾아줘")
local answer4=r(Answer)
local prompt4=r(Prompt)
smart_klips_exaone, prompt("26차 데이터에서 '종사상지위'(정규직, 비정규직 등)와 '월평균 임금' 변수명은 무엇인가요?")
local answer5=r(Answer)
local prompt5=r(Prompt)
smart_klips_exaone, prompt("'주당 근로시간' 변수의 결측치나 무응답에 해당하는 코드는 무엇인가요?")
local answer6=r(Answer)
local prompt6=r(Prompt)
smart_klips_exaone, prompt("26차 개인용 데이터를 사용하여 횡단면 분석을 할 때 적용해야 할 가중치 변수명은 무엇인가요?")
local answer7=r(Answer)
local prompt7=r(Prompt)
smart_klips_exaone, prompt("'삶의 만족도'를 종속변수로, '주당 근로시간'을 독립변수로 하고 '성별', '가구주 연령'을 통제하는 stata 회귀분석 코드 예시를 보여줘")
local answer8=r(Answer)
local prompt8=r(Prompt)
smart_klips_exaone, prompt("KLIPS에서 '정규직 근로자'는 구체적으로 어떻게 정의되나요?")
local answer9=r(Answer)
local prompt9=r(Prompt)
smart_klips_exaone, prompt("KLIPS 데이터를 사용하여 '근로시간'과 '만족도'를 함께 다룬 기존 연구 논문이 있다면 제목과 저자를 알려줘")
local answer10=r(Answer)
local prompt10=r(Prompt)
clear
set obs 10
gen strL prompt=""
gen strL answer=""

```

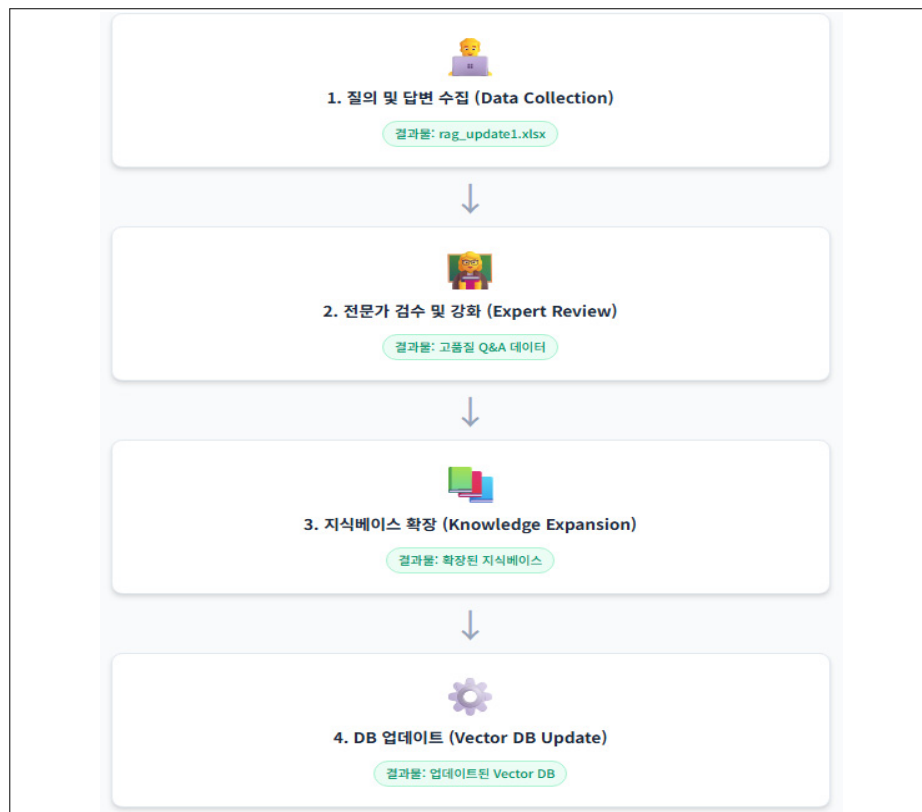
〈표 7〉의 계속

```

forvalues i=1/10 {
    replace prompt="`prompt`i'" in `i'
    replace answer="`answer`i'" in `i'
}
compress
save rag_update1, replace
export excel using rag_update1.xlsx, replace first(variable)
  
```

〈표 7〉의 마지막 Stata 코드에서 저장된 rag\_update1.xlsx 파일을 검토, 수정하여 index.faiss (임베딩 파일)을 업데이트하는 데 사용한다. 이러한 과정을 통해 더 특화된 KLIPS에 특화된 LLM을 구축할 수 있을 것으로 기대한다(그림 2) 참고).

[그림 2] RAG-LLM 개선 과정



## V. 결론 및 제언

### 1. 요약 및 시사점

본 연구는 장기간(26년) 조사가 진행된 방대하고 복잡한 KLIPS 데이터를 다루는 연구자들이 겪는 정보 접근의 어려움을 해소하고자, RAG 아키텍처와 한국어 특화 LLM(EXAONE 4.0)을 결합한 지식탐색 시스템 smart\_klips\_exaone 모듈을 개발하고 그 실효성을 검증하였다. Stata 환경에 직접 통합된 이 시스템은 연구자들이 기존의 작업 흐름을 벗어나지 않고 필요한 정보를 즉시 탐색할 수 있도록 설계되었다.

글로벌 최상위 LLM(GPT-4o, Gemini 2.5 Pro)과의 비교 평가를 통해 도출된 핵심적인 발견은, 전문 연구지원 도구의 효용성이 다차원적으로 평가되어야 한다는 점이다. 답변의 유창함이나 포괄성 측면에서는 글로벌 모델이 일부 강점을 보였으나, 연구의 엄밀성을 좌우하는 KLIPS 고유 용어의 정확한 재현과 연구의 신뢰도를 담보하는 환각 억제 능력에서는 RAG-EXAONE 모델이 비교우위를 나타냈다. 이는 연구지원 도구의 가치가 단순히 '그럴듯한 답변'이 아닌 '안전하고 정확한 답변'을 제공하는 데 있음을 시사한다. 또한, 본 연구는 국내 기술로 개발된 소버린 AI가 특정 전문 도메인에서 어떤 강점과 약점을 보이는지 실증적으로 진단함으로써, 향후 국내 AI 기술의 발전 방향에 대한 구체적인 시사점을 제공한다.

### 2. 한계 및 향후 연구방향

본 연구에서 개발한 시스템은 명백한 한계를 가지고 있다. 성능평가(표 5)에서 확인되었듯이, 여러 문서에 흩어진 정보를 종합하거나 복잡한 방법론적 절차를 깊이 있게 추론하는 데에는 아직 미흡한 성능을 보인다. 또한 환각 현상은 줄어들었지만 지식베이스에 존재하지 않으면 추론적 답변 능력이 부족하다.

이러한 한계를 극복하고 시스템을 지속적으로 발전시키는 방안으로 본 연구는 '인간-AI 협업 기반 선순환 구조(Flywheel Model)'를 제안할 수 있다. 이는 개발된 챗봇(smart\_klips\_exaone)을 연구자들이 질의응답 데이터를 수집하는 도구로 활용하고, 수집된 데이터를 전문가가 검수·강화하여 다시 지식베이스에 반영하는 모델이다. 이 순환 구조가 반복될수록, 시스템은 단순히 원본 문서를 검색하는 수준을 넘어 커뮤니티의 집단 지성이 축적된 진정한 도메인 전문가로 성장할 수 있다.

향후 연구로는 CoT(Chain-of-Thought)와 같은 고도화된 프롬프트 기법을 적용하여 추론 능력을 강화하고, 지식베이스를 KLIPS 활용 학술논문으로 확장하며, Stata 외에 R, Python 등 다른 통계 프로그램으로의 지원을 확대하는 과제가 남아있다.

## 참고문헌

- 권다운(2024). 「Is Having Well-off Parents a Part of the Resume?: Influence of Parent-Adolescent Conversations on Long-Term Educational Outcomes」. 『조사연구』 25 (1) : 23~50.
- 김용원·김정덕(2025). “노동패널 예산 삭감한 윤 정부의 난감함”. 더스쿠프 심층 취재. [https://v.daum.net/v/20250625092724201?s=print\\_news](https://v.daum.net/v/20250625092724201?s=print_news)(검색일 : 2025. 6. 26).
- 이광우·김수균(2024). 「국내 기업을 위한 RAG 구조 기반 질의응답시스템 설계」. 『한국컴퓨터정보학회논문지』 29 (7) : 81~88.
- 최지은·홍기석(2011). 「우리나라의 세대 간 소득 이동성 분석-아버지와 아들을 중심으로」. 『사회보장연구』 27 (3) : 143~164.
- 한선의·이군희(2025). 「RAG를 이용한 마케팅 활용에 대한 연구: 신용카드 할부 예측을 중심으로」. 『서비스경영학회지』 26 (1) : 95~123.
- 홍성은·이슬기(2025). 「세대 간 직업지위 차이와 청년 이직의도」. 『노동정책연구』 25 (2) : 57~86.

- 홍지윤 · 주아림 · 전수영(2025). 「RAG와 LLM을 활용한 금융 심사 설명력 향상 방안」. 『Journal of the Korean Data Analysis Society』 27 (2) : 465~476.
- Boiko, D. A., R. MacKnight, G. Kline, and G. Gomes(2023). “Autonomous Chemical Research with Large Language Models”. *Nature* 624 (7992) : 570~578.
- Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, and P. Liang(2021). “On the Opportunities and Risks of Foundation Models”. arXiv preprint arXiv : 2108.07258.
- Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, ... and D. Amodei(2020). “Language Models are Few-Shot Learners”. *Advances in Neural Information Processing Systems* 33 : 1877~1901.
- Gao, Y., Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, ... and H. Wang(2023). “Retrieval-Augmented Generation for Large Language Models : A Survey”. arXiv preprint arXiv : 2312.10997.
- Jain, P. and S. B. Goyal(2024). *Chunking Strategies for Rag Tutorial Using Granite*. IBM Developer. <https://www.ibm.com/think/tutorials/chunking-strategies-for-rag-with-langchain-watsonx-ai>(검색일 : 2025. 7. 15).
- Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, ... and D. Kiela(2020). “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. *Advances in Neural Information Processing Systems* 33 : 9459~9474.
- Luo, Z., M. Ghassemi, and F. Doshi-Velez(2024). “Large Language Models Surpass Human Experts in Predicting Neuroscience Results”. *Nature Human Behaviour* 8 (5) : 887~901.
- Schilde, L. and H. Heppner(2024). “Praktische Einblicke in Retrieval Augmented Generation(RAG)”. INNOQ Blog. <https://www.innoq.com/de/blog/rag-einblicke/>(검색일 : 2025. 7. 15).
- Sun, T., L. Liu, M. Sclar, M. Sap, and Y. Yang(2024). “Forecasting the

- Future of Online Debates with Large Language Models”. arXiv preprint arXiv : 2405.15850.
- Wiratunga, N., A. Y. Zhang, I. O. Nwogu, D. Corsar, and C. Pan(2024). “REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models”. arXiv preprint arXiv : 2402.07016.
- Yu, H. and Y. Chen(2024). “LAW-LLM: A RAG-based LLM for Chinese Legal Question Answering”. arXiv preprint arXiv : 2405.08812.
- Zhang, B., H. Yang, T. Zhou, A. Babar, and X. Y. Liu(2023). “Enhancing Financial Sentiment Analysis via Retrieval Augmented Large Language Models”. Proceedings of the Fourth ACM International Conference on AI in Finance : 349~356.
- Zhu, Y., C. Ren, S. Xie, S. Liu, H. Ji, Z. Wang, and C. Pan(2024). “REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models”. arXiv preprint arXiv : 2402.07016.

[부록] <표 5>의 smart\_klips\_exaone 답변결과

질문번호	답변결과 요약
Q1	<p>--- [답변] ---</p> <p>노동패널자료와 통계청 발표 수치의 차이는 주로 조사 방법론 및 데이터 수집 과정에서 발생하는 다양한 변수 때문으로 보입니다. 예를 들어, 노동패널은 특정 기간 동안의 실제 고용 상태를 더 정밀하게 추적하는 반면, 통계청의 통계는 더 넓은 표본 기반의 집계 데이터를 활용할 수 있어 결과가 달라질 수 있습니다. 또한, 통계청은 공식적인 통계 체계에 따라 분류하는 반면, 노동패널은 조사 설계에 따라 비정규직 정의나 범주가 상이할 수 있습니다.</p>
Q2	<p>### 가구소득 구성 세부 항목 및 변수명</p> <ol style="list-style-type: none"> <li>1. **근로소득**             <ul style="list-style-type: none"> <li>- 변수명 : 'h192102'(예시)</li> <li>- 내용 : 모든 가구원의 근로소득 합계값</li> </ul> </li> <li>2. **금융소득**             <ul style="list-style-type: none"> <li>- 세부 항목 :                     <ul style="list-style-type: none"> <li>- 'h192112-h192116' : 은행 이자/투자소득</li> <li>- 'h192122-h192126' : 사채 이자수입</li> <li>- 'h192134-h192138' : 주식/채권 매매차익</li> <li>- 'h192140' : 배당금</li> <li>- 'h192152-h192160' : 기타 금융소득</li> </ul> </li> </ul> </li> <li>3. **부동산소득**             <ul style="list-style-type: none"> <li>- 변수명 : 'h192126'(예시)</li> <li>- 내용 : 부동산 관련 소득(매매차익 등) 합계</li> </ul> </li> <li>4. **사회보험 수혜금**             <ul style="list-style-type: none"> <li>- 변수명 : 'h192136'(예시)</li> </ul> </li> <li>5. **이전소득**             <ul style="list-style-type: none"> <li>- 변수명 : 'h192152-h192160'(예시)</li> </ul> </li> <li>6. **기타소득**             <ul style="list-style-type: none"> <li>- 변수명 : 'h192142'(예시)</li> </ul> </li> </ol>
Q3	<p>추가 표본의 가중치 결정 방법은 다음과 같습니다.</p> <ol style="list-style-type: none"> <li>1. **표본 규모 설정** : 기존 패널 규모(7,000가구)에 추가로 6,000가구를 더해 총 1만 3,000가구로 구성합니다.</li> <li>2. **층별 비례 배분** : 기존 패널과 동일한 설계 방식(광역시·도 × 동·읍면부 × 주택 유형별 층별 배분)을 적용하여 가구 수를 비례적으로 분배합니다.</li> <li>3. **과부족 계산 및 보정** : 현재 유지되는 층별 패널 규모와 전체 목표 규모(1만 3,000가구) 간 차이를 분석해 추가 표본 수를 조정합니다.</li> <li>4. **가중치 산출** : 최종적으로 모집단 특성(예 : 연령, 성별, 가구 유형 등)을 반영해 각 가구 또는 계층별 가중치를 부여합니다. 이는 조사 목적에 따라 과학적·통계적 방법으로 계산됩니다.</li> </ol>

질문번호	답변결과 요약
Q4	“국민기초생활보호대상자 지원금은 가구용 설문지의 항목 (1) (국민기초생활보호대상자 지원금)에 별도로 기록되는 금액으로, 다른 정부보조금과 중복되지 않고 독립적으로 산정됩니다.”
Q5	주어진 참고 문서에서는 장애인 가구 식별 변수의 도입 시점이나 횡단 가중치 설정 방법에 대한 직접적인 정보를 찾을 수 없습니다. 따라서 이에 대한 정확한 답변을 제공하기 어렵습니다.
Q6	27차 데이터에 있는 h_sat 변수에 대한 정의는 제공된 참고 문서에서 확인할 수 없습니다. 정확한 정보를 확인하기 위해 추가적인 출처나 데이터셋을 참조해야 합니다.

Abstract

---

Developing a Knowledge Chatbot for KLIPS : Based on  
RAG-LLM Model

Min, Insik

This study develops a Retrieval-Augmented Generation(RAG) chatbot for specialized knowledge retrieval on the Korean Labor & Income Panel Study(KLIPS), using the Korean-specific AI model, EXAONE 4.0. The main purpose is to compare its performance against leading global LLMs such as Google Gemini and ChatGPT. We designed a RAG-LLM system that combines a knowledge base built from official KLIPS documentation with the EXAONE 4.0(1.2B) model, and developed the smart\_klips\_exaone module to ensure its usability within Stata. Benchmark Q&A comparisons revealed that while global models exhibited some superiority in general fluency, the RAG-EXAONE model demonstrated a distinct advantage in the accurate reproduction of KLIPS-specific terminology and in hallucination suppression. Considering that the utility of a professional research support tool lies in factual accuracy and safety, this finding suggests that the model proposed in this study may hold higher practical value in a real research environment. Furthermore, this study proposes a 'human-AI collaborative virtuous cycle' as a future direction for progressively enhancing the knowledge base.

Keywords : KLIPS, Retrieval-Augmented Generation(RAG), Large Language Model(LLM), EXAONE 4.0, Stata