

Design and Implementation of a Generative AI Ethics Q&A Chatbot Using RAG

Wan-Je Gil*, In-Soo Shin*

*Professor, Dept. of AI Convergence Education, Graduate School of Education, Dongguk University, Seoul, Korea

[Abstract]

This study designed and implemented an ethical Q&A chatbot using the Retrieval-Augmented Generation (RAG) technique to support the ethical use of generative AI. Based on the evaluation using the PDF document "Generative AI Ethics Guidebook (2023)", the RAG-based chatbot demonstrated superior response performance in handling unstructured data (such as tables and diagrams) within the document compared to existing generative AI models (ChatGPT, Gemini, DeepSeek, etc.). Using the RAGAS(Retrieval-Augmented Generation Assessment Score) evaluation framework, key metrics—including Context Recall and Factual Correctness—were analyzed. The results confirmed that the RAG chatbot enhances the reliability of document-based information retrieval and can be effectively used in AI ethics education. This research provides a foundational methodology for developing AI ethics guidelines and applying them in educational settings.

▶ **Key words:** Generative AI, Retrieval-Augmented Generation (RAG), Ethical AI, Q&A Chatbot, AI Ethics Guidelines

[요약]

본 연구는 생성형 AI의 윤리적 활용을 지원하기 위해 검색 증강 생성(RAG) 기법을 적용한 윤리 Q&A 챗봇을 설계·구현하였다. ‘생성형 AI 윤리가이드북(2023년)’ PDF 문서를 근거로 한 답변 성능을 평가한 결과 기존 생성형 AI 모델(ChatGPT, Gemini, DeepSeek)과 비교하여, RAG 기반 챗봇은 문서 내 비정형 데이터(표, 다이어그램 등)에서 우수한 응답 성능을 보였다. RAGAS(Retrieval-Augmented Generation Assessment Score) 평가 프레임워크를 활용해 주요 지표(Context Recall, Factual Correctness 등)를 분석한 결과, RAG 챗봇은 문서 기반 정보 검색의 신뢰성을 높이며 윤리 교육에 효과적으로 활용될 수 있음을 확인하였다. 본 연구는 AI 윤리 교육과 가이드라인 개발에 기초 자료를 제공한다.

▶ **주제어:** 생성형 AI, 검색 증강 생성(RAG), 윤리적 AI, Q&A 챗봇, AI 윤리 가이드라인

- First Author: Wan-Je Gil, Corresponding Author: In-Soo Shin
*Wan-Je Gil (jaygil8755@gmail.com), Dept. of AI Convergence Education, Graduate School of Education, Dongguk University
- *In-Soo Shin (9065031@hanmail.net), Dept. of AI Convergence Education, Graduate School of Education, Dongguk University
- Received: 2025. 08. 04, Revised: 2025. 08. 28, Accepted: 2025. 09. 02.

I. Introduction

최근 대규모 언어 모델(Large Language Model, LLM)의 발전과 함께 생성형 AI(Generative AI)가 다양한 분야에서 활용되고 있다. 특히 교육 분야에서는 AI를 활용한 맞춤형 학습, 자동 피드백, Q&A 챗봇 등의 기술이 도입되면서 학생들의 학습 효율성을 높이고 교사의 업무 부담을 줄이는 데 기여하고 있다. 그러나 생성형 AI의 활용이 급격히 늘어나면서, 생성된 콘텐츠의 신뢰성과 윤리적 문제가 중요한 연구 과제로 떠오르고 있다. AI가 제공하는 정보의 신뢰성을 높이는 방안으로 검색 증강 생성(Retrieval-Augmented Generation, RAG) 기법이 주목받고 있다.

RAG는 LLM이 기존의 학습 데이터만을 활용하는 것이 아니라, 외부 데이터베이스에서 실시간으로 관련 정보를 검색하여 보다 정확하고 신뢰할 수 있는 답변을 생성하는 기술이다. 이 기법은 특히 AI가 최신 정보를 반영하거나, 특정 도메인의 정밀한 데이터를 기반으로 응답해야 하는 상황에서 중요한 역할을 한다. 하지만 교육 현장에서 RAG를 적용할 경우, AI가 제공하는 정보의 정확성을 평가하고, 윤리적 문제를 고려하는 체계적인 시스템이 필요하다.

교육 현장에서 생성형 AI를 도입할 때 가장 큰 도전 과제 중 하나는 AI 활용 역량 부족과 윤리적 가이드라인 부재이다. 기존 연구에 따르면, 교사들은 AI를 효과적으로 활용하는 데 필요한 AI 리터러시(AI Literacy)가 부족하며, 학생들이 AI를 올바르게 사용할 수 있도록 지도하는 능력 또한 필요하다[1]. 또한, AI가 생성한 콘텐츠의 신뢰성을 평가하고, 저작권 문제나 학문적 부정행위 등을 방지하는 윤리적 가이드라인이 미흡하다는 점이 지적되고 있다. AI 기반 챗봇이 학생들에게 즉각적인 피드백을 제공할 수 있는 장점이 있지만, 데이터 프라이버시 문제, AI의 의사결정 투명성, 잘못된 정보 제공 가능성과 같은 윤리적 문제를 유발할 수 있다[5]. 이에 따라 생성형 AI의 도입을 위한 명확한 윤리적 기준과 가이드라인을 수립하는 것이 필수적이다.

본 연구에서는 사례별로 잘 정리된 생성형 AI 가이드라인을 교육현장에서 효과적으로 활용하기 위한 방법을 제시한다. 가이드라인의 내용만을 근거로 답변하는 Q&A 챗봇을 설계하고 구현하기 위해 최신 RAG 기법을 적용하였다. ChatGPT, Gemini, DeepSeek 등의 상용 AI 모델이 제공하는 문서 기반 Q&A 성능을 비교 평가하였고, 이를 보완할 수 있는 맞춤형 RAG 기반 챗봇을 설계하였다. AI 챗봇이 PDF 문서를 분석하고, 문서 기반 질의응답을 수행할 수 있

도록 시스템을 설계하였고 OpenAI GPT-4o 모델을 활용하여, 문서에서 검색된 정보를 기반으로 신뢰성 높은 답변을 생성하는 구조를 구현하였다. 챗봇 성능 평가를 위해 RAGAS(Retrieval-Augmented Generation Assessment Score) 평가 프레임워크를 활용하여 자동 평가를 수행하고, 연구자가 직접 평가하여 신뢰성을 검증하였다. 연구의 주요 기여는 다음과 같다.

첫째, 교육 환경에서 생성형 AI를 활용하면서 궁금한 윤리적 문제를 해결하기 위해, 문서 기반 Q&A 챗봇 시스템을 구축하고 RAG 기법을 적용하여 신뢰도 높은 정보를 제공할 수 있도록 설계하였다.

둘째, ChatGPT, Gemini, DeepSeek 등 최신 생성형 AI의 상용 RAG 서비스의 한계를 살펴보고 더 뛰어난 성능을 위해서 맞춤형 RAG Q&A 챗봇이 필요하다는 점을 비교 실험을 통해 증명하였다.

마지막으로, 연구 결과를 바탕으로 교육자와 학생들이 AI를 안전하고 윤리적으로 활용할 수 있도록 AI 윤리 가이드라인을 더 풍부하게 개발하고 이를 적용할 수 있는 방안에 대한 연구가 필요하다는 점을 제시하였다.

본 논문의 구성은 다음과 같다.

제2장에서는 생성형 AI 윤리 및 RAG 관련 선행 연구를 검토하고, Langchain과 RAG의 개념과 원리에 대해 설명한다. 제3장에서는 RAG 기반 Q&A 챗봇의 시스템 설계를 설명하며, AI가 정보를 검색하고 응답을 생성하는 과정에 대해 기술한다. 제4장에서는 챗봇의 성능을 평가하는 방법과 RAGAS 기반 성능 분석 결과를 제시하며, 평가 결과를 통해 RAG의 신뢰성을 검증한다. 마지막으로, 제5장에서는 연구의 결론을 도출하고, 향후 연구 방향을 제안한다.

II. Preliminaries

1. Related works

1.1 Preceding Studies on Generative AI Ethics

생성형 AI는 교육 현장에서 혁신적인 가능성을 제공하지만 동시에 역기능에 대한 문제를 고려하지 않으면 부작용을 초래할 수 있다. 학교에서 생성형 AI를 사용할 때 어떤 점들을 고려해야 하는지에 대한 선행 연구들을 살펴본다.

학교에서 생성형 AI를 도입할 때 가장 큰 도전 중 하나는 교사의 AI 활용 역량 부족이다. Brunner et al.(2025)의 연구에서는 AI 도입의 장애물로 학교의 기술적 준비 부족, 학생들의 AI 윤리 의식 결여, 그리고 교사의 AI 활용 역량 부족을 주요 문제로 지적하였다[2].

Hooper와 Lunn(2025) 또한 교사들을 위한 AI 리터러시 교육의 필요성을 강조하며, 교사들이 AI를 윤리적으로 활용할 수 있도록 훈련을 받을 필요가 있다고 주장하였다. 연구에서는 교사들이 AI를 이해하고 효과적으로 사용할 수 있는 능력이 교육 현장에서 필수적으로 요구되며, 특히 AI가 생성하는 콘텐츠의 신뢰성을 평가하는 방법과 학생들이 AI를 올바르게 활용하도록 지도하는 기술을 익히는 것이 중요하다고 언급하였다[3].

Montero et al.(2025)는 공립학교 교사들의 AI 인식과 윤리적 우려를 조사하였다. 연구 결과, 많은 교사들이 AI를 교육에 활용하는 것에 대해 긍정적인 태도를 보이지만, AI가 학생의 창의성을 억제할 가능성, 데이터 프라이버시 문제, 과제 부정행위 증가 등의 윤리적 문제를 심각하게 우려하는 것으로 나타났다. 이에 따라, 교사들이 AI를 효과적으로 활용할 수 있도록 리터러시 교육이 필수적이며, AI 활용 가이드라인을 개발해야 한다고 주장하였다[4].

학교에서 생성형 AI를 활용할 때 중요한 결림돌 중 하나는 윤리적 가이드라인이 부족하다는 것이다. Hooper와 Lunn(2025)의 연구에서는 교육 정책 측면에서 생성형 AI의 도입을 분석하였다. 연구에서는 대학과 초·중등 학교가 AI 도입 시 윤리적 가이드라인을 어떻게 수립해야 하는지를 탐색하며, 데이터 보호, 저작권 문제, 그리고 AI가 학문적 정직성에 미치는 영향을 주요 이슈로 제시하였다[5]. AI의 남용을 방지하기 위한 명확한 정책이 필요하며, 교육자들이 이를 이해하고 실천할 수 있도록 가이드라인이 제공되어야 한다고 하였다.

Nguyen et al.(2025)의 연구에서는 AI 기반 챗봇을 교육에 도입할 때 발생할 수 있는 윤리적 문제를 분석하였다. 챗봇은 학생들에게 즉각적인 피드백을 제공할 수 있는 장점이 있지만, 데이터 프라이버시 문제, 학생과의 상호작용에서 발생하는 윤리적 딜레마, 잘못된 정보 제공 가능성이 단점으로 지적되었다. 연구에서는 AI 챗봇을 설계할 때 윤리적 원칙을 적용하는 것이 필수적이라고 주장하였다[6].

Mishra et al.(2024)는 AI 도구를 연구 및 교육에서 사용할 때의 윤리적 가이드라인을 탐색하였다. 연구에서는 AI가 생성하는 정보의 신뢰성 문제, AI를 통해 생성된 콘텐츠의 소유권 문제, 그리고 연구 부정행위 가능성을 주요 이슈로 다루었다. AI가 학문적 연구를 보조하는 역할을 해야 하며, AI 사용을 명확히 규정하는 윤리적 가이드라인이 필요함을 강조하였다[7].

Zlotnikova et al.(2025)는 대학에서 AI를 윤리적으로 사용하기 위한 가이드라인을 연구하였다. 연구에서는 AI가 학문적 정직성에 미치는 영향을 분석하고, AI의 오용을 방

지하기 위한 정책적 해결책을 제안하였다. 학생들이 AI를 활용할 때 출처 표기 및 AI 사용 기준을 명확히 해야 한다고 주장하며, 이를 위해 대학 차원의 가이드라인이 마련될 필요가 있다고 결론지었다[8].

학술 연구 분야에서도 윤리적 쟁점에 대한 심도 있는 논의가 이루어졌다[9]. 연구자들은 가장 중요한 권고사항으로 ‘AI가 생성한 내용의 편향성 확인’(84.2%)을 꼽았으며, 그 뒤를 이어 ‘진위 여부 판단을 위한 선행 연구 검토’(82.4%), ‘AI 활용 사실의 명시 의무(75.9%)’가 중요하게 언급되었다. 이는 연구자들이 AI 생성 내용의 신뢰성과 책임성 문제를 인식하고 있음을 보여준다[10].

1.2 RAG (Retrieval-Augmented Generation) and LangChain: Concepts and Principles

최근 대규모 언어 모델(Large Language Model, LLM)의 발전과 함께 자연어 처리(Natural Language Processing, NLP) 기반 애플리케이션의 활용이 증가하고 있다. LLM을 효과적으로 활용하기 위해서는 단순한 모델 호출을 넘어, 데이터 검색, 프롬프트 최적화, 컨텍스트 관리 등의 기능이 필요하다. 이러한 기능을 효과적으로 구현할 수 있도록 돋는 프레임워크 중 하나가 LangChain이다. LangChain은 LLM을 기반으로 한 애플리케이션 개발을 용이하게 하기 위해 설계된 프레임워크이다. 이는 단순한 텍스트 생성 기능을 넘어, 외부 데이터 소스와의 통합, 연속적인 프롬프트 체인 구성, 사용자 맞춤형 질의응답 시스템 개발 등의 기능을 포함한다[11]. LangChain의 기본 개념은 프롬프트 엔지니어링, 체인 구조, 메모리 관리, 검색 증강 생성(Retrieval-Augmented Generation, RAG) 통합 등의 기능으로 정리될 수 있다.

RAG는 LLM이 가지고 있는 한계, 즉 실시간 정보 검색이 어렵거나 문맥이 부족해 생기는 환각 증상을 해결하기 위한 기법으로 등장했다. RAG는 입력된 데이터베이스에서 관련 정보를 검색한 후, 이를 LLM과 결합하여 보다 신뢰성 높은 응답을 생성하는 기술이다[12]. RAG는 기존의 LLM 기반 텍스트 생성 방식에 정보 검색 기능을 추가한 하이브리드 모델이다. 일반적인 LLM이 사전에 훈련된 데이터만을 활용하는 반면, RAG는 실시간으로 외부 데이터를 검색하여 보다 정확하고 최신 정보를 포함한 응답을 생성할 수 있다.

RAG는 크게 두 가지 주요 과정으로 구성된다: 검색 단계(Retrieval)와 생성 단계(Generation)이다. 검색 단계에서는 사용자의 입력(질문)이 들어오면, 관련 정보를 찾기 위해 외부 데이터 소스(예: 데이터베이스, 문서 저장소)에

서 관련 문서를 검색한다[13]. 검색된 문서는 텍스트 임베딩(Embedding) 과정을 거쳐 LLM이 이해할 수 있는 형태로 변환된다. 생성 단계에서는 검색된 문서를 입력 데이터로 활용하여 LLM이 응답을 생성한다. 일반적인 LLM과 달리, RAG는 검색된 문서의 내용을 반영하여 보다 정확하고 문맥적으로 풍부한 답변을 제공한다[14, 15]. RAG는 학술, 법률, 의료, 고객 서비스 등 다양한 분야에서 활용되고 있으며, 향후 더 정교한 검색 알고리즘과 결합하여 더욱 발전할 것으로 기대된다. 특히, LangChain과 같은 프레임워크와 함께 사용될 경우, 보다 유연하고 효율적인 AI 기반 애플리케이션 개발이 가능하다.

1.3 Recent Research Cases on LangChain and RAG

Byun et al.(2024)은 개인화 데이터베이스에 저장된 문서를 활용하여 사용자별 맞춤 검색 및 Q&A 시스템을 구현하였다. 연구에서는 SQL 데이터베이스와 벡터 데이터베이스, 그리고 MongoDB를 활용해 사용자 문서와 대화 기록을 관리하고, LangChain을 통해 데이터를 전처리 및 임베딩하여 LLM(GPT-3.5-Turbo 등)으로부터 정확한 답변을 생성하였다. 또한, 다양한 모델 조합에 대한 정량적 평가(정확도, 응답 신뢰성, 문맥 정밀도 등)를 수행하여 시스템 성능을 면밀히 분석하였다[16].

Iaroshev et al.(2024)는 금융 보고서를 대상으로 RAG 기법을 적용해 투자자들이 은행의 반기 및 분기 보고서를 보다 정확하게 분석할 수 있도록 지원하는 Q&A 시스템을 개발하였다. 연구에서는 금융 도메인에 특화된 데이터 전처리 및 평가 체계를 마련하고, 여러 모델 구성(OpenAI ADA, GPT-4o, Google Gemini 1.5 Pro 등)의 비교 평가를 통해 금융 보고서 분석에 최적화된 RAG 시스템의 성능과 한계를 심도 있게 분석하였다[17].

Jeong(2024)은 생성형 AI 서비스 구현의 전반적인 기술 스택과 아키텍처에 초점을 두었다. 연구에서는 파운데이션 모델, LLM, 프롬프트 엔지니어링, 그리고 오피스트레이션 계층을 포함한 생성형 AI 서비스의 구성 요소를 상세하게 분석하였다. 또한, 다양한 생성형 AI 모델과 프레임워크를 비교·분석하여 실제 비즈니스 환경에서의 적용 가능성과 비용 효율성을 논의하였다[18].

Choi et al.(2024)는 한국어 챗봇에 QA Pair Passage 기법을 적용하여, 기존 Zero-shot 방식 대비 GPT-3.5와 GPT-4 모델에서 각각 평균 23.3%, 25%의 성능 향상을 확인하였다[19]. 또한, Park(2024)는 치과 상담 분야에 특화된 챗봇을 개발하기 위해, 국내 치과 대학병원 웹 게시판의 데이터를 활용해 상담 데이터를 구축하고,

LangChain 기반의 RAG 시스템을 통해 사용자 질의와 가장 유사한 상담 내용을 효과적으로 검색·출력하는 방안을 제시하였다[20].

본 연구에서는, 이미지, 표 등이 포함된 PDF 형식의 한글 문서에 대해 Markdown 언어로 파싱하는 최신 기법을 활용하여 문서의 내용을 손실 없이 분석하였고, RAGAS를 통한 객관적인 성능 평가와 최신 언어 모델과의 비교 실험을 수행하였다.

III. The Proposed Scheme

본 연구에서 제안하는 RAG 기반 Q&A 챗봇 시스템은 문서 내 정보를 효과적으로 검색하고, 생성형 AI 모델과 결합하여 정확한 답변을 제공하도록 설계되었다. 시스템의 전체적인 흐름은 데이터 로드, 데이터 파싱, 텍스트 분할, 임베딩 생성, 벡터 저장소 구축, 검색 및 프롬프트 생성, 최종 응답 생성의 7단계로 구성된다. 시스템의 전체적인 설계도는 Fig. 1에 제시되어 있다.

본 시스템은 PDF 문서를 입력 데이터로 사용한다. 사용자가 특정 문서를 업로드하면, 해당 문서가 분석 및 검색 가능한 형태로 변환된다. 본 연구에서는 “생성형 AI 윤리 가이드북”PDF를 기본 데이터로 활용하였다.

“생성형 AI 윤리 가이드북[21]”은 방송통신위원회와 한국지능정보사회진흥원이 2023년 12월 공동제작한 것으로 74페이지 분량이다. 저작권, 책임성, 허위조작정보, 개인정보와 인격권, 오남용에 관한 주제로 질문과 답변 형식으로 작성되었다.

본 연구에서는 다이어그램, 이미지로 된 표, 그래프 등 비정형 데이터가 포함된 문서의 처리 방식에 있어서 단순한 텍스트 파싱을 넘어 시각적 구조를 보존한 정보 추출이 핵심 과제라고 판단하였다. 따라서 PDF 내 다이어그램이나 그래프를 이해하기 위해, LlamaParse를 이용하여 markdown으로 변환하는 방식을 실험을 통해 구현하였다. 본 연구에서 선택한 LlamaParse는 일반적인 OCR(Text-in-Image) 기술을 넘어서 이미지 내 텍스트 추출과 문서 구조의 마크다운 변환 기능을 동시에 제공한다. 즉, LlamaParse는 PDF 내 텍스트를 추출하고 문서의 구조를 유지하면서 변환하는 역할을 수행한다. LlamaParse는 주로 LlamaIndex와 함께 사용되며, RAG 기반 시스템에서 정보를 사전 처리하는 부분을 담당한다 [22, 23].

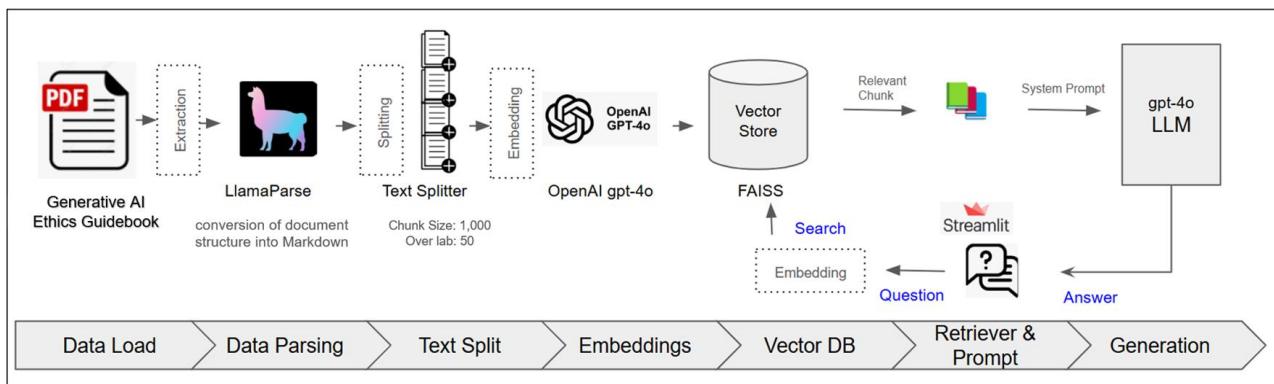


Fig. 1. System Architecture

텍스트는 RecursiveCharacterTextSplitter를 이용하여 작은 청크(Chunk)로 분할된다. Chunk는 긴 문서를 효과적으로 처리하기 위해 문서를 작은 조각(Chunk)로 나누는 것을 의미한다. 긴 문서를 분할할 때 일반적으로 500~1500자의 Chunk 크기가 문맥 보존과 검색 효율성 사이의 균형을 유지하는 데 적합한 범위로 보고되고 있다 [24]. 본 연구에서는 Chunk 크기를 1000, Overlap 크기를 50으로 설정하여 문맥을 유지하면서 데이터 검색이 가능하도록 구성하였다. Overlap은 각 Chunk간 50개의 문자가 겹치도록 설정하는 것으로 이전 Chunk의 일부 문맥을 다음 Chunk에 포함시킬 수 있는 기법이다. Overlap 크기가 50일 경우 의미 손실없이 문맥이 이어지는 것으로 실험을 통해 확인하였다.

Chunk와 Overlap의 크기는 실험을 통해 속도와 문맥 유지 간의 균형을 갖춘 적절한 값을 선택할 수 있다.

분할된 텍스트 청크는 OpenAI GPT-4o의 임베딩 모델로 벡터화되어, 문서의 의미 정보를 벡터 공간상에 표현함으로써 검색 정확성을 향상시킨다. 생성된 벡터는 FAISS(Facebook AI Similarity Search)를 이용하여 저장된다.

FAISS는 대규모 벡터 데이터베이스에서 빠르고 효율적인 근사 최근접 이웃(Approximate Nearest Neighbor, ANN) 검색을 수행하는 라이브러리로서, 본 시스템에서는 유사도 검색을 수행하는 핵심 요소로 활용된다. 최근 연구에서, FAISS를 활용하여 대규모 PDF 문서를 효율적으로 검색하는 시스템을 개발하거나 이미지 검색 및 분류 시스템을 구축한 사례들이 있다[24, 25].

사용자의 질의가 입력되면, 시스템은 FAISS에서 가장 관련성이 높은 텍스트 청크를 검색(Search) 하여 반환한다. 이후, 검색된 정보와 사용자 질의를 조합하여 프롬프트(Prompt)를 생성한다. 시스템 프롬프트는 Fig. 2와 같이 설정하였다.

```

prompt = PromptTemplate.from_template(
    """You are an assistant for question-answering tasks. Use the
following pieces of retrieved context to answer the question. If you
don't know the answer, just say that you don't know.
Answer in Korean.

#Question:
{question}
#Context:
{context}
#Answer:""")

#Question:
{question}
#Context:
{context}
#Answer:"""

```

Fig. 2. System Prompt

생성된 프롬프트는 GPT-4o LLM으로 전달되며, 모델은 문맥을 기반으로 최적의 응답을 생성한다.

최종적으로, 파이썬 웹앱 구축 라이브러리인 Streamlit을 활용한 챗봇 인터페이스를 통해 사용자에게 답변을 반환한다.

IV. Results

1. Comparative Performance Analysis

RAG를 구현하는 과정에서 문서 내 다양한 형태의 데이터를 효율적으로 처리하는 것은 중요한 과제 중 하나이다. PDF와 같은 문서에는 텍스트뿐만 아니라 표, 다이어그램, 이미지 내 텍스트 등의 비정형 데이터가 포함되어 있으며, 이를 효과적으로 구조화하는 기술이 필요하다. LlamaParse는 문서 내 이미지 기반 텍스트를 자동으로 추출하고, Markdown 포맷으로 변환하는 기능을 제공하여, AI 기반 검색 및 자연어 처리(NLP) 응용에서 중요한 역할을 한다. PDF 문서 내 그림으로 표현된 텍스트 내용을 기반으로 한 질문들에서 얼마나 정확하게 관련 텍스트를 추출하는지 실험해보았다.

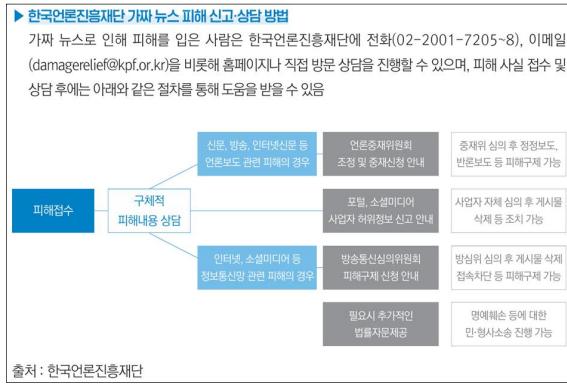


Fig. 3. Sample Text with Diagram from PDF

Fig. 3과 같이 PDF 파일 46페이지에 있는 가짜 뉴스 피해 신고 상담 방법에 대한 설명은 다이어그램 형식으로 되어 있다. 이 부분에서 이미지 내의 텍스트를 정확히 추출하는 것도 중요하지만, 구성 요소 간의 관계성을 파악하는 것이 더욱 중요하다. 피해접수부터 시작해 구체적인 피해 내용이 어떤 것인지에 따라 절차가 달라지기 때문이다. 본 연구에서는 이미지를 마크다운으로 변환하여 ‘피해접수’와 ‘구체적 피해내용 상담’의 관계가 비교적 정확하게 파악될 수 있다.

본 연구에서 구현한 Q&A챗봇에게 “한국언론진흥재단 가짜뉴스 피해신고 방법 중에서 언론보도 관련 피해의 경우 어떻게 해야 해?”라고 물어봤을 때, “언론보도로 인한 피해의 경우, 한국언론진흥재단을 통해 언론중재위원회에 조정 및 중재 신청을 안내받을 수 있고, 정정보도나 반론보도 등을 통해 구제를 받을 수 있습니다”라고 관련 내용을 정확히 찾아서 답변하였다. 이는 이미지 내용을 표형식으로 구조화하여 파악했기 때문인 것으로 판단된다. 또한, Fig. 4와 같은 체크리스트 형식의 윤리 항목 점검표를 인식하는 능력을 테스트한 결과, 1번 ‘투명성’부터 10번 ‘연대성’까지 모두 정확히 답변하였다.

최근 ChatGPT, Gemini, DeepSeek 등의 생성형 AI는 파일을 첨부하면 문서 파일에 기반한 질의, 응답 서비스를 제공하고 있다. 따라서 본 연구에서와 같이 별도의 RAG 시스템을 활용한 챗봇을 구현할 필요가 있을까 하는 의구심이 들 수 있다. 따라서 문서 Parsing과 Split, Embedding, 벡터저장소, 시스템 프롬프트 작성까지 문서의 특성에 기반한 설계가 어떻게 상용 생성형 AI 서비스와 다를 수 있는지 알아보았다. 본 연구에서 구현한 챗봇에게 물어본 질문들을 다른 생성형 AI에게 똑같이 했을 때 결과가 어떠한지 확인해보았다.

▶ 인공지능(AI) 윤리 국가표준(KS) 첫 제정(국기기술표준원, 2023. 6. 14.)

산업통상자원부 국기기술표준원은 ‘AI 윤리 점검 서식’에 대한 국가 표준(KS)을 제정, 발표함.

최근 생성형 AI의 윤리적인 사용 문제가 제기되는 시점에서 AI 제품, 서비스 개발 시에 필요한 윤리적 고려 항목을 제시하고 자체 점검할 수 있는 체크리스트를 공개해 기본적인 가이드라인의 역할을 할 수 있도록 전망됨.

4.9 관련 윤리 항목 점검표			
	서비스 제공자	사용자	개발자
(1) 투명성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(2) 공정성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(3) 우해성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(4) 책임성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(5) 사생활 보호성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(6) 편의성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(7) 자율성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(8) 신뢰성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(9) 지속성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(10) 연대성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

→ 행위 주체별(서비스 제공자, 사용자, 개발자)
체크 항목이나 고려 되어야하는 윤리적인 항목

Fig. 4. Sample Text with Table from PDF

먼저, ChatGPT 4o 모델을 사용하여 PDF 파일을 첨부하고 언론보도 피해 대처방안에 대해 물었을 때, 관련 그림에 있는 내용을 찾아서 답변을 해주었는데, 방송통신심의위원회 피해 구제 신청과 같이 질문과 직접 관련 없는 내용까지 언급하였다.

두 번째 AI윤리 점검 서식에 관련된 질문에서는 관련 내용을 찾을 수 없다고 답변하였다. 그래도 모르면 ‘모른다고 답하라’는 시스템 프롬프트에 근거한 답변이라고 할 수 있다. RAG 시스템에서는 해당 문서에 있는 내용으로만 답변하는 것이 신뢰성을 높이는데 중요하다.

다음으로 구글의 최신 모델인 Gemini 2.0 Flash를 활용해서 같은 질문을 제시하였다.

언론보도 피해 관련 답변에서는, 인터넷신문윤리위원회와 같이 문서에는 존재하지 않는 기관을 언급하는 오류를 보였고 근거가 없는 답변을 길게 늘어놓았다.

또한, AI 윤리 점검서식에 대한 내용을 찾지 못하였고 첨부한 문서에 근거한 답변을 하지 못하였다. 할루시네이션이라고까지 말하긴 어렵지만 국가기술표준원의 윤리항목점검이 아닌, 가이드라인 문서에서 제시한 체크리스트를 가져와서 답변하였다.

마지막으로 최근 화재가 되고 있는 DeepSeek에게도 같은 질문을 하였다. 언론 보도 피해 관련 질문에, 다이어그램 내용을 언급하지 못하였고, 윤리 점검 서식에 관한 질문에서도 관련 내용을 찾지 못하고, 문서에 없는 내용들을 장황하게 대답하였다. AI윤리점검서식에 대한 답변은 국가기술표준원의 해당 문서 내용을 찾지 못하였고 구글 제미나이와 마찬가지로 생성형 AI를 혁명하게 활용하기 위한 체크리스트 페이지의 내용을 답변하였다.

이렇듯 불완전한 답변을 생성하는 이유를 다시 한 번 확인하기 위해, 첨부된 PDF 내의 내용에서만 답변하고 답변의 출처를 명확하게 제시하라는 프롬프트를 추가로 전달하였다. 그 결과, 예를 들어 ChatGPT는 “내가 확인한 조각에서는 ‘절차’의 구체적 단계(예: 피해 구제, 법률 자문 연결, 언론중재위 안내 등)가 잘려 있습니다. 즉, 원문에 이어지는 실제 절차 단계 전체가 다 드러나지 않은 상태입니다”라고 답변하여 이미지를 가져오는데 실패했음을 확인해주었다. 마찬가지로 Gemini도 “요청하신 문서 '생성형_AI윤리_가이드북.pdf'에는 한국언론진흥재단 기자뉴스 피해신고 방법에 대한 내용이 포함되어 있지 않습니다. 따라서 문서 내용을 참고하여 답변드릴 수 없습니다.”고 하였다.

또한, pdf 파일에 특화된 서비스인 ChatPDF.com에서도 같은 방식으로 질문을 하고 확인해보면 “해당 자료에 포함된 표가 이미지로 된 표인 경우, 저는 텍스트 인식을 통해 표의 내용을 직접 읽거나 파악할 수 없습니다.”고 답변하여 상용 서비스의 한계를 보여주었다.

마지막으로 RAG 방식에 특화된 구글의 notebooklm을 통해 확인해본 결과, 첫 번째 질문에 해당하는 Fig. 3 이미지는 제대로 추출하여 정확한 답변을 하였으나, Fig. 4의 이미지는 인식하지 못해 국가기술표준원의 10가지 점검항목을 답변하지 못하였다.

이상의 실험을 통해 3개의 주요 LLM 모델과 본 연구에서 구현한 챗봇의 성능을 Table 1에서와 같이 Context Recall과 Faithfulness 측면에서 0~1 사이의 점수로 나타냈다. Context Recall과 Faithfulness의 의미와 측정방법에 대해서는 다음 장의 RAG 모델 성능평가에서 이어서 설명한다.

2. Performance Evaluation using RAGAS

RAG모델의 성능을 평가하는 방법은 크게 검색 성능 평가, 생성 성능 평가, 출처 기반 응답 평가로 구분할 수 있다.

이를 효과적으로 분석하기 위해 Precision@k, ROUGE, BERTScore, Faithfulness, Semantic Similarity 등의 지표를 활용하며, 최근에는 RAGAS(Retrieval-Augmented Generation Assessment Score) 프레임워크를 이용한 자동 평가 기법이 도입되고 있다. RAG 모델의 성능을 평가할 때, 주요하게 고려해야 할 요소로는, 정보 검색의 정확성, 생성된 응답의 신뢰성, 출처 기반 응답의 정확성, 사용자 경험 및 유용성 등이 있다.

RAGAS는 RAG 시스템의 성능을 평가하기 위한 도구로, 다양한 평가 지표(Metrics)를 제공한다. 이는 RAG 기반 시스템이 문서에서 적절한 정보를 검색하고, 정확하고 신뢰할 수 있는 응답을 생성하는지 평가하는 데 사용된다. RAGAS는 context recall, factual correctness, faithfulness, semantic similarity 네 가지 주요 평가 지표를 사용하여 RAG 모델의 성능을 종합적으로 측정한다[26, 27].

Context Recall(문맥 회수율)은 검색된 문서가 질문과 관련된 정보를 얼마나 잘 포함하고 있는지를 측정하는 지표이다. 값이 높을수록 검색 시스템이 질문에 필요한 정보를 충분히 포함하고 있음을 의미한다.

Factual Correctness(사실적 정확성)은 생성된 응답이 문서에서 가져온 정보와 얼마나 일치하는지 평가하는 지표이다. AI 기반 평가 모델(BERTScore, GPT-4 등)을 사용하여 답변이 문서 내 정보와 얼마나 사실적으로 일치하는지를 측정한다.

Faithfulness(신뢰성)은 생성된 답변이 검색된 문서의 내용에 기반하여 생성되었는지 평가하는 지표이다. 답변이 문서에서 가져온 정보와 직접적으로 관련이 있는지 측정하는 메트릭이다. Factual Correctness는 답변의 사실적 정확성을 평가하는 반면, Faithfulness는 답변이 오직 검색된 문서의 정보만을 기반으로 생성되었는지를 평가한다. 문서에 있는 내용만 말해야지 다른 말을 덧붙인다면 그 내

Table 1. Comparative Evaluation of LLM Responses to Document-Based Questions

	ChatGPT 4o	Gemini Flash 2.0	DeepSeek R1	Ours
Q1) A question regarding the damage caused by media reports	Providing an answer including the correct response along with unrelated additional information	Added content and terms not present in the document (e.g., Korea Internet Newspaper Ethics Committee)	The answer is incomplete due to the inability to locate the relevant document (diagram)	Found the flowchart in the figure and answered only the key points
Q2) A question regarding the AI ethics checklist form	Unable to answer due to not finding relevant documents (tables, figures)	Answered with content unrelated to the question and not found in the document	Could not find the relevant document, so answered using a different checklist	Found the table in the figure and answered accurately
Evaluation Score by Authors	Context Recall 0.5 Faithfulness 0.3	Context Recall 0.3 Faithfulness 0.3	Context Recall 0.0 Faithfulness 0.0	Context Recall 1.0 Faithfulness 1.0

용이 사실이더라도 신뢰성 점수는 낮게 측정될 수 있다.

Semantic Similarity(의미적 유사성)은 모델이 생성한 답변과 문서 간의 의미적 일치도를 측정하는 지표이다. BERTScore, Sentence-BERT(SBERT) 등 임베딩 기반 모델을 사용하여 답변과 문서 간의 의미적 유사도를 비교한다.

RAG의 성능을 평가하기 위해서는 테스트 데이터셋이 필요하다. 최근 연구에서는 가장 최신 모델들을 사용하여 데이터셋을 구축하는 사례가 보고 되고 있다[29]. 연구자들은 GPT-4를 사용하여 메타 템플릿을 자동으로 생성하고, 이를 기반으로 수백만 개의 수학 문제와 해답을 합성한 대규모 수학 문제 데이터셋을 구축하거나[30], GPT-3를 이용해 자연어 추론(NLI) 예시들을 생성한 뒤, 사람 크라우드워커들이 이를 필터링, 수정, 라벨링하는 방식으로 데이터셋을 구축한 사례도 있다[31]. 이러한 절차는 자동 생성의 효율성이 있지만 사람의 검증 과정도 반드시 필요 한 과정이다. 본 연구에서도 데이터셋은 ChatGPT 4-o 모델을 사용하여 만들었다. 데이터셋 형식은 csv 파일로 하였고, 질문, 컨텍스트, 정답, 검색된 문서 정보를 생성하도록 하였다. 연구자가 생성된 데이터셋의 내용과 형식에 문제가 없음을 확인하였다.

또한 문서에 질문과 답변쌍을 만들 때 질문과 연관된 컨텍스트와 검색된 문서의 페이지 번호와 제목을 함께 기입하여, 생성된 데이터의 근거를 쉽게 찾아보고 판단할 수 있도록 하였다.

15개 테스트데이터로 성능을 평가한 결과 <Fig. 5>에서와 같이 평균 성능은 각각 Context Precision은 1.0, Faithfulness는 0.6, Answer Relevancy는 0.7, Context Recall은 1.0을 나타냈다.

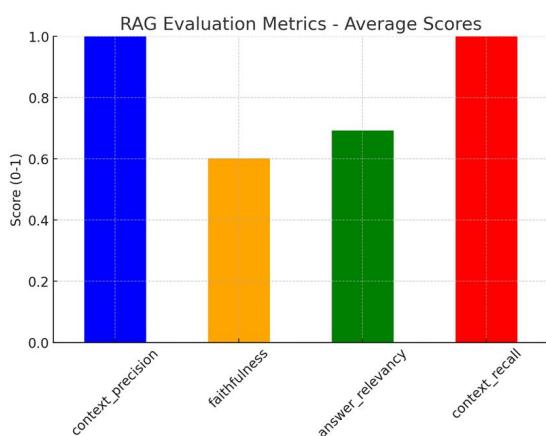


Fig. 5. RAGAS Evaluation Metrics

Context Precision과 Context Recall은 모든 데이터에 대해 1.0으로 매우 높게 나타나, 검색된 문서들이 질의와 관련된 정보를 충분히 포함하고 있음을 확인할 수 있었다. 이는 RAG 시스템이 적절한 컨텍스트를 효과적으로 검색하고 제공하는 능력이 우수함을 시사한다.

반면, Faithfulness 점수가 상대적으로 낮게 측정된 부분이 존재하는데, 이는 LLM이 검색된 문서에서 제공되지 않은 정보를 보완적으로 생성했을 가능성을 의미한다. 그러나 본 연구 참여자들이 책본의 답변을 면밀히 검토한 결과, Faithfulness 점수가 0점으로 측정된 답변들 중에서도 실제로 문서의 내용을 벗어난 근거 없는 정보(Hallucination)가 포함되지 않은 경우도 확인되었다.

이는 현재의 자동 평가 방식이 AI 응답의 신뢰성을 완벽하게 측정하는 데 한계가 있을 수 있으며, 생성형 AI의 응답 품질을 평가할 때 휴먼 전문가(Human-in-the-loop)의 정성적 검토가 함께 이루어져야 함을 시사한다. 따라서, 향후 연구에서는 Faithfulness 평가 방식을 보다 정교화하고, AI 응답의 신뢰성을 보다 정확하게 반영할 수 있는 평가 체계를 추가적으로 도입하는 방안이 필요하다.

V. Conclusions

본 연구에서는 검색 증강 생성(RAG, Retrieval-Augmented Generation) 기법을 적용한 AI 윤리 Q&A 책본 시스템을 설계하고 구현하였다. AI 윤리 교육 및 정보 제공의 신뢰성을 높이기 위해 Vector DB를 활용한 문서 검색과 LLM(대규모 언어 모델)의 응답 생성 기법을 결합하였으며, 기존의 RAG 시스템이 가지는 한계를 극복하는 새로운 접근 방식을 제안하였다. 기존의 RAG 연구와 차별화되는 요소를 다음과 같이 정리할 수 있다. 기존 RAG 시스템은 일반적인 텍스트 문서(위키, 논문, 웹 문서 등)에서 검색하여 답변을 생성하는 방식이었으나, 본 연구에서는 표, 다이어그램, 이미지 내 정보까지 분석할 수 있도록 설계하였다. 특히, 기존 상용 AI 모델에서는 문서 내용을 벗어난 정보를 생성하거나(할루시네이션) 문서 내 필요한 정보를 충분히 반영하지 못하는 경우가 발생했으나, 본 연구에서 구현한 시스템은 문서에 근거한 응답 생성을 통해 AI 답변의 신뢰성을 높였다.

본 연구는 생성형 AI 윤리 가이드북을 기반으로 RAG 시스템을 구축하였으나, 향후 다양한 분야의 윤리적 이슈를 포괄하는 데이터 확장과, 윤리적 개념을 활용한 논리적 판단 능력 평가가 필요하다. 즉 기존 문서에서 직접적인

답을 찾을 수 없는 윤리적 판단을 요구하는 질문을 추가하여, RAG 시스템이 단순한 검색 결과 전달이 아닌 윤리적 원칙을 해석하고 적용하는 능력을 평가할 필요가 있다. 또한, 실시간으로 업데이트되는 최신 윤리 가이드라인과 정책 문서를 반영할 수 있는 자동화된 데이터 수집 및 업데이트 기능을 추가하는 것이 연구의 다음 과제이다.

마지막으로, 본 연구에서 적용한 RAGAS 기반의 정량적인 평가 방식 이외에도 정성적인 방식으로 휴먼 피드백 (Human-in-the-loop) 시스템을 추가하여, 실제 사용자 경험을 반영한 AI 응답 평가 기법을 도입하는 것을 중요한 후속 연구과제로 도출하였다.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A01090926)

REFERENCES

- [1] D. T. K. Ng, E. K. C. Chan, and C. K. Lo, "Opportunities, Challenges and School Strategies for Integrating Generative AI in Education," *Computers and Education: Artificial Intelligence*, p. 100373, 2025. DOI:10.1016/j.caai.2025.100373
- [2] B. Brunner, S. Schon, and M. Ebner, "From Gretel to Strudelcity: Empowering Teachers Regarding Generative AI for Enhanced AI Literacy with CollectiveGPT," *Education Sciences*, vol. 15, no. 2, p. 206, 2025. DOI: 10.3390/educsci15020206
- [3] K. Hooper and S. J. Lunn, "Traversing New Horizons: An Exploration of Educational Policies on Generative AI," in *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2*, 2025, pp. 1479-1480. DOI: 10.1145/3641555.3705272
- [4] M. A. A. Montero, G. C. B. Serafin, J. A. A. Lucaylucay, and P. A. M. Gapol, "Exploring the Knowledge, Willingness, and Concerns of Generative AI Adoption: A Quantitative Study Among MAPEH Teachers."
- [5] K. Hooper and S. J. Lunn, "Traversing New Horizons: An Exploration of Educational Policies on Generative AI," in *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2*, 2025, pp. 1479-1480. DOI: 10.1145/3641555.3705272
- [6] H. Nguyen, V. Nguyen, S. Ludovise, and R. Santagata, "Value-sensitive design of chatbots in environmental education: Supporting identity, connectedness, well-being and sustainability" *British Journal of Educational Technology*, 2025. DOI: 10.1111/bjet.13568
- [7] T. Mishra, E. Sutanto, R. Rossanti et al., "Use of large language models as artificial intelligence tools in academic research and publishing among global clinical researchers," *Scientific Reports*, vol. 14, no. 1, Art. 31672, 2024. DOI: 10.1038/s41598-024-81370-6.
- [8] I. Zlotnikova, H. Hlomani, T. Mokgetse, and K. Bagai, "Establishing ethical standards for GenAI in university education: a roadmap for academic integrity and fairness," *Journal of Information, Communication and Ethics in Society*, vol. 23, no. 2, pp. 188–216, 2025. DOI: 10.1108/JICES-07-2024-0104.
- [9] D. Schlagwein and L. Willcocks, "'ChatGPT et al.': The ethics of using (generative) artificial intelligence in research and science," *Journal of Information Technology*, vol. 38, no. 3, pp. 232–238, 2023. DOI: 10.1177/02683962231200411.
- [10] H. J. Jungh, "NRF Hompage", <https://webzine.nrf.re.kr/magazine/2412/sub3.php>(December, 2024)
- [11] X. Liu, Y. Shen, H. Cheng, H. Zhang, T. He, and Y. Liang, "Prompting frameworks for large language models: A survey," *arXiv preprint arXiv:2311.12785*, 2023.
- [12] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020. DOI: 10.48550/arXiv.2005.11401
- [13] V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, "Dense Passage Retrieval for Open-Domain Question Answering," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, Nov. 2020, pp. 6769–6781. DOI: 10.18653/v1/2020.emnlp-main.550.
- [14] G. Izacard and E. Grave, "Leveraging Passage Retrieval with Generative Models for Open-Domain Question Answering," *arXiv preprint arXiv:2007.01282*, 2020. DOI: 10.48550/arXiv.2007.01282.
- [15] K. Vaishnav and G. O. Nathan, "Need of retrieval augmented generation for conversational assistants for government social welfare scheme popularization," *International Journal of Science and Research Archive*, vol. 13, no. 1, pp. 1299–1312, Sep. 2024. DOI: 10.30574/ijrsa.2024.13.1.1747.
- [16] J. Byun, B. Kim, K. Cha, and E. Lee, "Design and Implementation of an Interactive Question-Answering System with Retrieval-Augmented Generation for Personalized Databases," *Applied Sciences*, vol. 14, no. 17, Art. 7995, 2024. DOI: 10.3390/app14177995

- [17] KI. Iaroshev, R. Pillai, L. Vaglietti, and T. Hanne, "Evaluating Retrieval-Augmented Generation Models for Financial Report Question and Answering," *Applied Sciences*, vol. 14, no. 20, Art. 9318, 2024. DOI: 10.3390/app14209318
- [18] I. Jeong, "A Graph-Agent-Based Approach to Enhancing Knowledge-Based QA with Advanced RAG," *Knowledge Management Research*, vol. 25, no. 3, pp. 99–119, Sep. 2024. DOI: 10.15813/kmr.2024.25.3.005
- [19] S. Choi, S. Park, C. Park, J. Wang, J. Kim, and B. Kim, "Enhancing Passage Selection and Answer Generation in FiD Systems Using Relevance Gating," *Journal of the Korean Institute of Information Scientists and Engineers*, vol. 52, no. 5, pp. 385–392, 2025. DOI: 10.5626/JOK.2025.52.5.385.
- [20] J. Park, "Development of dental consultation chatbot using retrieval augmented LLM," *The Journal of the Institute of Internet, Broadcasting and Communication*, vol. 24, no. 2, pp. 87–92, 2024. DOI: 10.7236/JIIBC.2024.24.2.87
- [21] J. Kim, S. Lee, H. Park, and Y. Choi, "Generative AI Ethics Guidebook," *Korea Intelligent Information Society Agency (NIA)*, 2023.
- [22] Z.-F. Shi, K. Liu, S. Bai, Y.-T. Jiang, T. Huo, X. Jing, R.-Z. Li, and X.-J. Ma, "Meta data retrieval for data infrastructure via RAG," in Proc. of the 2024 IEEE International Conference on Web Services (ICWS), Jul. 2024, pp. 100–107. DOI: 10.1109/ICWS62655.2024.00029.
- [23] P. Ingemarsson and D. Persson, "PDF Parsing: Unveiling the Most Efficient Method," *Engineering Degree Project*, Linnaeus University, 2024.
- [24] S. Nurhaliza, F. Atoilah, A. Alimin, R. Selviana, and N. m. Muhimah, "Fruit Segmentation and Identification through Image Processing with K-Means and MobileNet V2," *Journal of Information Technology application in Education, Economy, Health and Agriculture*, vol. 1, no. 3, pp. 63-69, 2024.
- [25] B. Chandra, P. Preethika, S. Challagundla, and Y. Gogireddy, "End-to-End Neural Embedding Pipeline for Large-Scale PDF Document Retrieval Using Distributed FAISS and Sentence Transformer Models," *International Journal of Advanced Research in Computer Science and Engineering*, vol. 1, no. 2, Art. IJARCSE_01_02_001, pp. 1–21, 2024.
- [26] F. Öberg, "Automating the Assessment of Retrieval-augmented Generation Responses," Bachelor's Thesis, Department of Mathematics and Computer Science, Karlstad University, 2025, p. 95.
- [27] P. Thetbanthat, B. Sathanarugsawat, and P. Praneetpolgrang, "Application of Generative Artificial Intelligence Models for Accurate Prescription Label Identification and Information Retrieval for the Elderly in Northern East of Thailand," *Journal of Imaging*, vol. 11, no. 1, Art. 11, 2025. DOI: 10.3390/jimaging11010011I. Introduction

Authors



Wan-Je Gil received the B.A. degree in German Language and Literature from Yonsei University, Korea, in 1995, the M.Ed. degree in Robot Education from Gongju National University of Education, Korea, and the Ph.D.

degree in Computer Engineering from Kongju National University, Korea, in 2023. Since 2020, he has been serving as an adjunct professor in the Graduate School of Education at Dongguk University, majoring in AI Convergence Education. His research interests include AI-integrated education, educational data science, and generative AI.



In-Soo Shin received the B.A. degree in Education from Yonsei University, Seoul, Korea, and the M.S. and Ph.D. degrees in Educational Measurement and Statistics from Florida State University, Florida, USA.

Dr. Shin joined the faculty of the Graduate School of Education at Dongguk University, Seoul, Korea, where he is currently a Professor in the Department of AI Convergence Education. He also serves as a faculty member in the Department of Educational Data Science at the Graduate School of Educational Service Science. His research interests include meta-analysis, AI-integrated education, and educational data science.