

RAG 기반 정보를 통한 맞춤형 영화 가이드 플랫폼 사례

(A Case Study of a Personalized Movie Guide Platform using RAG-based information)

서채연*, 김다경**, 김장환***, 김영철****

(Chaeyun Seo, Dakyeong Kim, Janghwan Kim, R. Young Chul Kim)

요약

최근, 국내 영화 소비가 OTT 중심으로 빠르게 재편되면서 극장 외 관람이 급증하고 있다. 하지만 기존 영화 정보 플랫폼들은 단편적인 평점이나 파편화된 리뷰 제공에 그치고, 방대한 온라인 정보 통합 및 요약 기능이 없어 사용자들이 필요한 정보를 찾기 어렵다. 이는 신뢰할 수 있는 정보 부족으로 이어져 영화 선택을 더욱 힘들게 한다.

이러한 문제 해결과 개인 맞춤형 영화 정보 탐색 경험 제공을 위해 RAG(Retrieval Augmented Generation) 기반 영화 가이드 플랫폼을 제안한다. 이 플랫폼은 대규모 지식 베이스에서 최신 정보를 동적으로 검색하고, 사용자 관람 이력 및 선호 장르를 분석해 개인화된 영화를 추천한다. 또한, 맥락 기반 질의응답으로 복합적인 질문에도 유의미한 답변을 생성한다. RAG 기반 챗봇을 통해 변화하는 영화 소비 형태에 최적화된 영화 탐색 및 선택 경험과 높은 만족도를 제공할 것으로 기대되며, 향후 다양한 도메인으로의 확장성 및 데이터 처리 효율성 향상을 목표로 한다.

■ 중심어 : 검색 증강 생성 ; 인공지능 기반 웹 어플리케이션 ; 프롬프트 엔지니어링

Abstract

Recently, as domestic movie information platform is rapidly reorganized due to the OTT, the number of out-of-theater viewings is rapidly increasing. However, existing movie information platforms only provide fragmented ratings or fragmented reviews, and do not have the ability to integrate and summarize vast online information, making it difficult for users to find the information they need. This leads to a lack of reliable information, making movie selection even more difficult. To solve these problems and provide a personalized movie information search experience, we propose a movie guide platform based on Retrieval Augmented Generation (RAG). This platform dynamically searches for the latest information from a large knowledge base and analyzes user viewing history and preferred genres to recommend personalized movies. In addition, it generates meaningful answers to complex questions through context-based Q&A. It is expected to provide a movie search and selection experience optimized for changing movie consumption patterns and high satisfaction through a RAG-based chatbot, and we aim to improve expandability to various domains and data processing efficiency in the future.

■ keywords : Retrieval Augmented Generation (RAG) ; AI based Web Application ; Prompt Engineering

I. 서론

최근 국내 영화 소비 형태는 전통적인 영화관 관람에서 OTT(Over The Top) 플랫폼 이용으로 빠르게 전환되고 있다. 한국영화진흥위원회의 통계에 따르면, 국내 극장 영화 관람률은

2018년 75.8%에서 2023년 52.4%로 급격히 감소하는 추세다. 이는 코로나19 팬데믹의 영향과 더불어 넷플릭스, 티빙, 웨이브, 쿠팡플레이 등 다양한 OTT 서비스의 보편화가 큰 영향을 미쳤기 때문이다. 반면, OTT 등 극장 외 영화 관람률은 같은 기간 90.2%에서 97.2%로 꾸준히 증가하여, 영화 소비의 주류가 온라인 플랫폼으로 이동했

* 정회원, 홍익대학교 소프트웨어융합학과 초빙교수

** 준회원, 홍익대학교 소프트웨어융합학과 학부생

*** 정회원, 홍익대학교 소프트웨어융합학과 외래교수

**** 정회원, 홍익대학교 소프트웨어융합학과 교수

본 연구는 2025년도 문화체육관광부의 재원으로 한국콘텐츠진흥원(과제명: 인공지능 기반 대화형 멀티모달 인터랙티브 스토리텔링 3D장면 저작 기술 개발, 과제번호: RS-2023-00227917, 기여율:100%) 지원과 한국연구재단의 4단계 두뇌한국21사업(과제명: 초분산 자율 컴퓨팅 서비스 기술 연구팀, 과제번호: 202003520005)의 지원을 받아 수행된 연구임.

접수일자 : 2025년 06월 24일

재제확정일 : 2025년 07월 18일

교신저자 : 김영철 e-mail : bob@hongik.ac.kr

음을 명확히 보여준다[1]. 이러한 변화는 사용자들이 언제 어디서든 원하는 영화를 소비할 수 있는 환경이 조성되었음을 의미하며, 이는 개인의 취향과 선호에 맞는 영화 정보 탐색의 중요성을 더욱 부각시킨다.

영화 소비 형태의 변화에도 불구하고, 현재 영화에 대한 종합적인 리뷰와 정보를 제공하는 플랫폼은 사용자에게 효율적인 탐색 및 비교 기능을 제공하지 못하고 있다. 왜냐하면, 기존 영화 플랫폼은 단편적인 평점 제공에 그쳐 사용자가 영화의 내용, 연출, 연기 등 다각적인 측면을 깊이 이해하기 어렵다. 또한, 기존의 영화 플랫폼은 매체별·평론가별 리뷰 중심 구성으로 인해 다양한 관점의 정보를 통합적으로 얻기 어렵고, 사용자는 폭넓은 의견을 비교하기 위해 여러 플랫폼이나 웹사이트를 개별적으로 방문해야 하는 번거로움을 겪는다.

일반적으로 정보를 제공하는 플랫폼들은 고유 플랫폼에 저장되어 있는 정보를 통해 정보를 전달한다. 하지만, 뉴스 기사, 블로그, 소셜 미디어 등 다양한 출처에서 발생하는 방대한 영화 정보를 통합하고 요약하는 기능이 부재하여 사용자가 필요한 정보를 찾는 데 많은 시간을 소모하게 한다. 이러한 정보들은 온라인상의 무분별한 정보로 여겨질 수 있지만, 분석을 통해 신뢰할 수 있고 정제된 정보가 부족하여 영화 선택에 혼란을 주거나 잘못된 기대를 형성하게 할 수 있다 [2]. 사용자들은 이러한 문제점들로 인해 변화하는 영화 소비 환경에서 사용자들이 자신에게 최적화된 영화 콘텐츠를 선택하는 데 어려워진다.

본 논문에서는 위에서 언급한 문제들을 개선하고 사용자에게 더 효율적이고 개인 맞춤형 영화 정보 탐색 경험을 제공하기 위해, 우리는 RAG 기반 영화 가이드 플랫폼을 제안한다.

제안하는 방법은 기존의 정보 검색 및 생성 방식의 한계를 개선하여 다음과 같은 강점을 제공한다. 첫째, 대규모 지식 베이스에서 관련 문서를 동적으로 검색함으로써 단순히 저장된 정보를

제공하는 것을 넘어, 최신 정보를 포함한 방대한 데이터를 활용하여 사용자에게 정보를 제공한다. 둘째, 사용자의 과거 관람 이력, 선호 장르, 감독, 배우, 평점 패턴 과거 사용 정보등을 분석하여 개인화된 영화 추천을 제공한다. 이는 사용자가 선호에 맞는 정보를 빠르게 얻을 수 있도록 돕는다. 셋째, 사용자의 질문 의도를 파악하여 맥락 기반의 질의응답을 제공함으로써 복합적인 질문에도 유의미한 답변을 생성한다. 궁극적으로 RAG 기반 챗봇을 탑재한 맞춤형 영화 가이드 플랫폼은 변화하는 영화 소비 형태에 발맞춰, 사용자가 방대한 정보 속에서 자신에게 최적화된 영화를 탐색하고 선택하는 데 있어 차별화된 사용자 경험과 높은 만족도를 제공할 것으로 기대한다.

2장에서는 관련 연구로 RAG를 이용한 챗봇 시스템에 대해 언급한다. 3장에서는 제안하는 방법의 적용 사례로서 RAG 기반 정보를 통한 맞춤형 영화 가이드 플랫폼 언급하고 4장에서는 결론 및 향후 연구에 대해 언급한다.

II. 관련 연구

2.1. Retrieval Augmented Generation

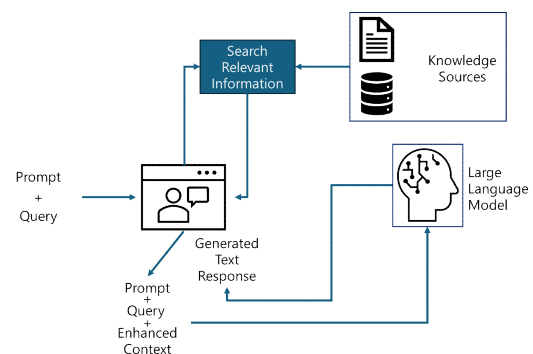


그림 1. Retrieval-Augmented Generation Mechanism

Retrieval-Augmented Generation(RAG)는 대규모 언어 모델(LLM)의 답변 생성 능력을 향상시키는 기술이다. 이 방식은 LLM이 질문에 대한 답변을 만들기 전에 외부 지식 베이스나 검색 시스템에서 관련성 높은 정보를 찾아 답변을 강화하는 구조를 갖는다. 즉, RAG는 대규모 외부

문서 모음에서 필요한 정보를 검색한 다음, 그 검색된 문서들을 기반으로 seq2seq 모델을 활용하여 조건부 답변을 생성하는 방식으로 작동한다 [3].

이러한 RAG 기반 챗봇은 단일 모델의 내부에만 의존하지 않고 외부 데이터로부터 정제된 정보와 다양한 세부 사항을 가져올 수 있다는 장점을 지닌다. 이를 통해 주어진 질문에 대해 정확하고 풍부한 답변을 제공할 수 있다 [4].

하지만 RAG 모델에도 한계는 존재한다. 검색 모듈과 생성 모듈의 연결 과정에서 불필요한 노이즈가 섞이거나, 검색된 문서 자체가 질문에 적합하지 않은 경우에는 오히려 부정확한 답변이 생성될 수 있다는 점을 고려해야 한다.

2.2. 정보 통합 플랫폼

정보 품질, 시스템 품질 및 사용 편의성, 서비스 품질 및 개인화 수준이 중요한 요인으로 밝혀졌다. 정보 품질은 기사의 정확성, 신뢰성, 최신성 및 다양성을 포함하며, 시스템 품질은 직관적인 인터페이스, 빠른 로딩 속도, 안정적인 서비스와 개인화된 추천 기능을 강조한다[5]. 서비스 품질 측면에서는 사용자 맞춤형 뉴스 피드 제공과 관련된 기사 추천 기능이 주요하게 작용한다. 이 연구는 신문사 통합 플랫폼의 성공적인 구축 및 사용자 확보를 위해서는 고품질의 콘텐츠 제공뿐만 아니라 사용자의 편의성을 고려한 시스템 설계 및 개인화된 서비스 제공이 필수적임을 시사한다[5][6].

Ⅲ. RAG 기반 정보를 통한 맞춤형 영화 가이드 플랫폼

3장에서는 RAG 기반 영화 평가 플랫폼 챗봇 메커니즘과 적용 사례를 소개한다. 이를 통해 사용자가 선호에 맞는 정보를 빠르게 얻는다.

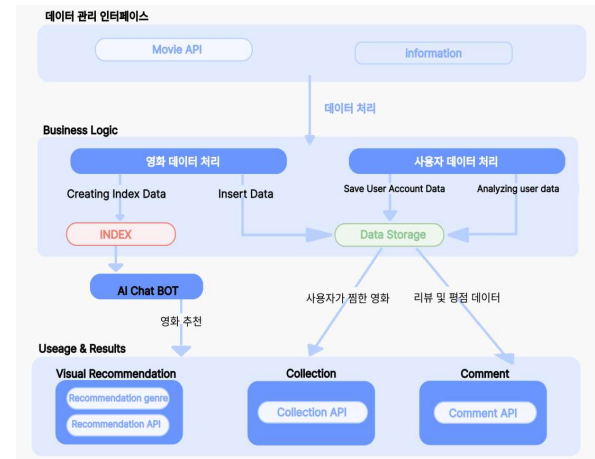


그림 2. RAG 기반 정보를 통한 맞춤형 영화 가이드 플랫폼

본 논문에서 제안하는 메커니즘의 구조 그림2와 같다. 먼저, 데이터 관리 인터페이스 계층(Data Management Interface), 비즈니스 로직(Business Logic) 계층, 사용자 데이터 처리, 데이터 저장소(Data Storage), AI 챗봇(AI Chat Bot), 그리고 사용 결과 제공(Usage & Results)의 총 여섯 단계로 구성된다. 각 단계는 유기적으로 연결되어 사용자에게 개인화된 영화 추천 및 정보 탐색 경험을 제공한다.

(1) 데이터 관리 인터페이스

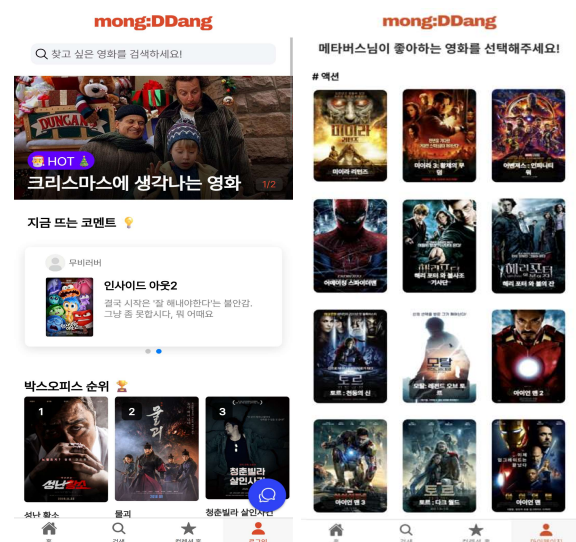


그림 3. 맞춤형 영화 가이드 플랫폼 메인화면

첫 단계인 데이터 관리 인터페이스는 외부로부터 영화 관련 데이터를 효율적으로 수집한다. Movie API를 통해 공공 영화 데이터

API(KMDB)로부터 영화 제목, 장르, 개봉일, 러닝타임, 감독, 출연 배우, 줄거리 등 정형화된 영화 정보를 주기적으로 자동 수집하며, 이는 영화 추천의 기본적인 기반 정보를 형성한다. 더불어, Information을 통해 영화 이외의 관람 후기, 웹사이트 트렌드 정보, 평론가 리뷰 요약 등 비정형적이거나 보조적인 데이터를 수집한다. 이 두 데이터 소스는 각각의 형태에 맞게 전처리된 후, 다음 단계인 비즈니스 로직 계층으로 넘어가 영화 추천을 위한 핵심 자료로 가공되어 그림 3처럼 나타낸다[7][8].

(2) 비즈니스 로직 계층

비즈니스 로직 계층은 수집된 데이터를 실제 서비스에 맞게 처리하는 핵심 영역이다. 유저가 장르별 선호 취향을 선택하면, 영화 데이터 처리와 사용자 데이터 처리를 각각 수행하며, 서비스 전반의 기저 정보를 형성한다.

영화 데이터 처리는 수집된 영화 데이터를 빠르게 검색하고 추천에 활용할 수 있도록 가공하는 과정을 포함한다. Creating Index Data 단계에서는 영화별 키워드, 유사도 기반의 벡터화, 장르 태그화 등을 통해 AI가 빠르게 특정 영화를 검색할 수 있도록 인덱스 구조를 만들고 그림 4처럼 데이터베이스를 구축한다.

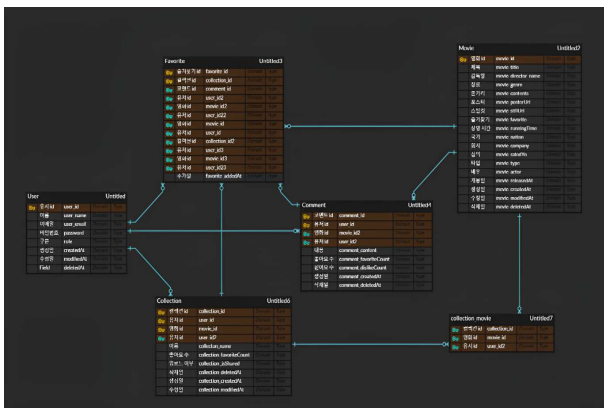


그림 4. 맞춤형 영화 가이드 플랫폼 데이터베이스 설계

이어서 Insert Data 단계에서는 인덱싱이 완료된 데이터를 DB 또는 AI 검색 시스템에 삽입하여 실시간 추천 응답에서 빠른 결과를 제공하는 중요한 기반을 마련한다. 여기서 AI 챗봇은 단순

한 문장 완성형 챗봇이 아닌, 정확한 영화 정보를 기반으로 답변하는 추천 시스템 역할을 수행한다.

(3) 사용자 데이터 처리

사용자 데이터 처리는 사용자 맞춤형 서비스를 구현하기 위한 사용자 관련 데이터를 다룬다. 이는 사용자 행동 분석과 계정 기반 정보 수집을 포함하며 두 단계로 구성된다. Save User Account Data 단계에서는 사용자가 로그인하거나 서비스를 사용할 때 생성되는 기본 정보(계정, 찜한 영화 목록, 리뷰 내용 등)를 수집하여 저장한다. 이는 그림 5처럼 사용자의 활동 이력을 기록하고, 이후 영화 추천 및 UI 개인화에 활용된다.

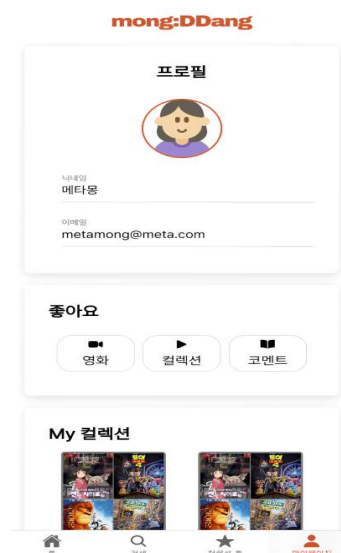


그림 5. 사용자 데이터 시각화 UI

Analyzing User Data 단계에서는 수집된 사용자 데이터를 분석하여 선호하는 장르를 추출하며, 이 분석 결과는 AI 챗봇의 응답 및 시각적 추천 인터페이스의 개인화에 반영되어 사용자에게 더욱 맞춤화된 경험을 제공한다.

(4) 데이터 저장소

데이터 저장소는 시스템의 핵심 저장 공간으로, 다양한 분석 결과와 사용자 활동 정보를 통

합 관리한다. 사용자가 ‘좋아요’나 ‘찜하기’ 등으로 저장한 영화 목록은 사용자 맞춤 추천의 핵심 요소로 저장되며, 사용자 리뷰, 별점 등의 평가 데이터도 체계적으로 정리하여 저장된다. 이 저장소는 AI 챗봇뿐만 아니라 시각화, 추천 API, 컬렉션 기능, 댓글 API 등 다양한 시스템의 기반 데이터 소스 역할을 수행한다.

(5) AI 챗봇

RAG기반 인공지능 챗봇은 단순한 질문-응답형 AI가 아닌, 사용자의 입력을 기반으로 실시간으로 분석하고 추천까지 수행하는 고도화된 기능을 갖춘 시스템이다. 사용자의 질문을 입력받으면, 저장된 영화 인덱스, 사용자 취향 정보, 리뷰 평가 등을 종합 분석하여 최적의 결과를 도출한다. 예를 들어, 영화 ‘기생충’에 대한 질문에 “영화 ‘기생충’은 봉준호 감독의 작품으로, 빈부 격차를 소재로 사회 문제를 풍자한 블랙코미디입니다. 수상 이력으로는 아카데미 4관왕을 수상했으며, 주연은 송강호, 이선균, 조여정입니다.”와 같이 상세하고 맥락적인 응답을 제공한다. 이처럼 AI는 단순 요약형 응답을 넘어 개인화된 추천을 수행하며, 이후 사용 결과 제공 영역으로 결과가 전달된다.

ChatbotService는 사용자 메시지를 OpenAI의 모델에 전달하고, 그 응답을 받아오는 핵심 로직을 담당한다. 이 서비스는 private chatHistory: { role: 'user' | 'assistant' | 'system'; content: string }[] = []와 같은 배열을 사용하여 채팅 기록을 저장한다. 이 배열에서 role은 메시지의 주체를 나타내는데, 이는 사용자를 의미하는 'user', 챗봇 응답을 의미하는 'assistant', 또는 시스템 지시사항을 의미하는 'system' 중 하나가 된다. content는 실제 메시지 내용을 담는다.

챗봇이 대화를 시작하기 전에, const systemMessage = { role: 'system', content:

\You are an assistant with ID: \${this.assistantId}...'}와 같은 시스템 메시지를 GPT 모델에 전달한다.

이 시스템 메시지는 GPT에게 “어떤 역할을 수행해야 하는지”를 명확하게 설명하는 역할을 한다. 특히, 여기서 Assistant ID를 명시하고, 응답은 maxTokens' 안에서 문장 끊김 없이 생성하도록 지시하여 챗봇의 응답 품질과 일관성을 유지한다. 그림 6은 챗봇을 구현한 코드 예시이다.

```
const updatedChatHistory: { role: 'user' | 'assistant' | 'system'; content: string }[] = [
  systemMessage,
  ...this.chatHistory.slice(-10),
  { role: 'user', content: prompt },
];

// OpenAI API 호출
const completion = await this.openai.chat.completions.create({
  model: 'gpt-4',
  messages: updatedChatHistory,
  max_tokens: this.maxTokens,
  temperature: 0.5,
});
```

그림 6. 챗봇 문맥 유지를 위한 코드 예시

```
@Injectable()
export class ChatbotService {
  private chatHistory: { role: 'user' | 'assistant' | 'system'; content: string }[] = [];
  private openai: OpenAI;
  private readonly assistantId: string = 'asst_NwKjypl33lytPstbqBDw0g6'; // Assistant ID
  private readonly maxTokens: number = 1000; // 응답 최대 토큰 수

  constructor() {
    this.openai = new OpenAI({
      apiKey: openAIConfig.apiKey,
    });
  }

  /**
   * 사용자 입력을 기반으로 OpenAI API 호출
   * @param prompt 사용자 입력
   * @returns OpenAI 응답
   */
  async getChatResponse(prompt: string): Promise<string> {
    try {
      const systemMessage: { role: 'system'; content: string } = {
        role: 'system',
        content: `You are an assistant with ID: ${this.assistantId}. Respond within ${this.maxTokens} tokens and adjust the response to fit the limit without cutting off sentences.`
      };
    }
  }
```

그림 7. GPT 기반 AI Chatbot 코드 예시

그림 7에서 챗봇은 문맥을 유지하여 대화형 응답이 가능하도록 const updatedChatHistory = [systemMessage, ...this.chatHistory.slice(-10), { role: 'user', content: prompt }]; 코드를 통해 최근 10개의 대화 기록을 유지하며, 새로운 사용자 입력과 시스템 메시지를 추가한다. 이후, const completion = await this.openai.chat.completions.create({...}); 코드를 사용하여 GPT-4 모델을 통해 대화 기반 응답을 생성한다. 이때 messages: updatedChatHistory를 통해 시스템 메시지, 이전 대화 기록, 그리고 현재 사용자 입력이 모두 포함된 배열을 모델에 전달한다. 그림 8은 챗봇 구현 사례를 나타낸다.

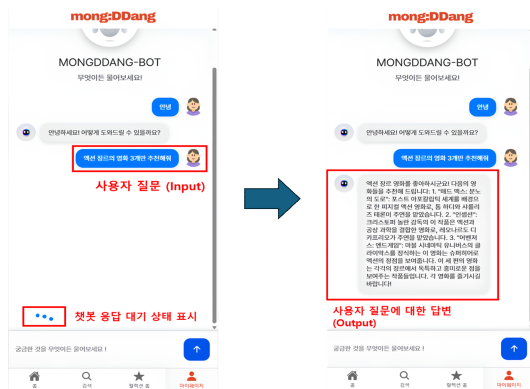


그림 8. Assistant RAG 기반 챗봇 구현 사례

IV. 결 론

본 논문에서는 변화하는 영화 소비 환경에서 사용자의 편의성을 극대화하고 정보의 비정확성 문제를 해결하기 위해 RAG기반 영화 가이드 플랫폼을 제안한다. 이 시스템은 정제된 영화 데이터를 바탕으로 개인 맞춤형 추천과 맥락 기반의 질의응답을 제공하여 사용자가 소모하는 정보제공 시간을 줄인다. 하지만 RAG 모델은 검색 모듈과 생성 모듈의 연결 과정에서 불필요한 정보(노이즈)가 섞이거나, 검색된 문서 자체가 질문에 적합하지 않으면 부정확한 답변이 생성될 수 있는 한계점이 존재한다.

향후 연구에서는 이러한 한계점을 극복하고, 다양한 도메인으로의 확장성 및 데이터 처리 효율성을 더욱 향상하는 방향으로 연구를 지속할 것이다.

REFERENCES

- [1] 한국저작권위원회, “2023 국민문화예술활동조사,” 2024년 9-1호 [비즈니스 인사이트] 저조한 성적의 국내 극장 영화 관람률
- [2] 이승찬, “국내 OTT 플랫폼의 영화 콘텐츠 흥행 연구 : 자국시장모형과 문화적 할인, 넷플릭스 효과를 중심으로”, 석사학위논문, 한국외국어대학교, 서울, 2022년
- [3] P. Lewis, “Retrieval-augmented generation for knowledge-intensive nlp tasks.” *Advances in neural information processing systems* 33 (2020): 9459-9474.
- [4] S. Vakayil, D. S. Juliet, A. J and S. Vakayil,

“RAG-Based LLM Chatbot Using Llama-2,” 2024 7th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, pp. 1-5, 2024.

- [5] 김대원, 김민성, 김성철, “신문사 통합형 플랫폼 사용의도에 영향을 미치는 요인,” *한국콘텐츠학회논문지*, 제15권, 제4호, 122-138쪽, 2015년
- [6] 윤봉식, “형태적 관점의 콘텐츠 UI구조 분석 방법 설계” *스마트미디어저널*, 제8권, 제4호, 58-63 2019
- [7] 류시천, 최진영, 심미희, “스마트 애플리케이션 UX 디자인 평가 접근 방법,” *스마트미디어저널* 제8권, 제3호, 70-77쪽, 2019년
- [8] 한지애, 류시천. “정보디자인 관점에서 방향 안내 표지판 디자인의 메타 가이드라인 제안,” *스마트미디어저널* 제10권, 제2호, 61-69쪽, 2021년

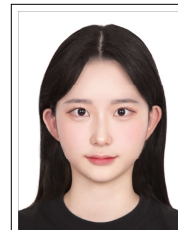
저 자 소 개



서채연(정희원)

1999년 홍익대학교 전자전산학과 학사 졸업.
2005년 홍익대학교 전자전산공학 석사 졸업.
2014년 홍익대학교 전자전산공학 박사 졸업.

<주관심분야 : Component Repository, 소프트웨어공학, 메타버스, AR/VR, 생성형 AI, 웹, 언리얼엔진5, 디자인씽킹>



김다경(준희원)

2025년 홍익대학교 세종캠퍼스 소프트웨어융합학과 학부생 재학.
2025년 홍익대학교 메타버스융합SW 아카데미 4기 수료.

<주관심분야 : 웹 개발, 인공지능, 디지털 트윈>



김장환(정희원)

2019년 아이다호주립대학교 컴퓨터과학과 학사 졸업.
2022년 홍익대학교 소프트웨어융합학과 석사 졸업.
2024년 홍익대학교 소프트웨어융합학과 박사 수료

<주관심분야 : 소프트웨어공학, 소프트웨어 가시화, AI소프트웨어테스팅>



김영철(정희원)

2000년 일리노이공대(IIT) 소프트웨어공학 박사졸업
2025년 홍익대학교 소프트웨어융합학과 교수 재직

<주관심분야 : 소프트웨어공학, 메타버스(AR/VR), 소프트웨어테스팅, 메타모델링, 소프트웨어 가시화, 소프트웨어 품질>