

개발자 교육 LMS 시스템 내에 RAG 기반 챗봇 적용 사례

(Best Practices on Applying RAG Based Chatbots in Learning Management System(LMS) for Software Developers)

김재호*, 서채연**, 김영철***

(Jaeho Kim, Chaeyun Seo, and R. Young Chul Kim)

요약

기존의 학습 관리 시스템(LMS)은 소프트웨어 개발을 위한 교육 환경에서 프레임워크 수준의 지식 기반의 요구를 지원하지 못한다. 앞으로의 LMS 플랫폼은 학습자의 이런 지식 기반의 요구를 지원해야 한다. 이를 위해, 학습 관리 시스템(LMS) 내에 웹 프레임워크 NestJS 공식 문서 기반 검색 증강 생성(RAG) 시스템 탑재를 제안한다. 이 방식은 Ollama와 Dify를 활용하여 RAG 시스템의 빠르고 간단한 구축 및 운영이 가능하다. 또한 개발자들의 질의에 대해 NestJS 공식 문서 기반의 맞춤형 가이드 응답을 제공한다. 이를 통해 LMS 내의 RAG 시스템은 개발자들이 직면하는 지식 기반 문제 해결에 도움을 준다. 궁극적으로, 우리는 RAG 시스템이 소프트웨어 개발 교육 분야의 교육 품질을 향상시키고, 개발자에게 AI 기반 튜터링을 통한 맞춤형 교육을 제공할 것으로 기대한다.

■ 중심어 : 학습 관리 시스템 ; 소프트웨어 개발 교육 ; 대규모 언어 모델 ; 검색 증강 생성

Abstract

Conventional learning management systems (LMS) do not support knowledge-based needs for a framework-level knowledge base in software development educational environments. Future-oriented LMS platforms should be able to support such knowledge-based needs of learners. For this, we propose integrating a retrieval augmented generation system (RAG) based on the official NestJS documentation into the LMS. By utilizing Ollama and Dify, we enabled the rapid and easy setup and operation of the RAG system. Our system, based on the official NestJS documentation, provides responses for development-related questions to learners and junior developers. Through this, the RAG system can assist in solving knowledge-based problems that junior developers encounter while working on projects. Ultimately, our RAG system in the field of software development education enables providing personalized education to developers via AI tutoring.

■ keywords : Learning management system(LMS) ; software development education ; large language model(LLM) ; retrieval augmented generation(RAG)

I. 서론

오늘날 LMS는 교육에서 필수적인 요소로 자리잡았다. LMS는 학습 활동을 관리하고 학습 콘

* 준회원, 홍익대학교 소프트웨어융합학과 석사과정

** 준회원, 홍익대학교 소프트웨어융합학과 초빙교수

*** 정회원, 홍익대학교 소프트웨어융합학과 정교수

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

게재확정일 : 2025년 07월 18일

수정일자 : 2025년 07월 17일

교신저자 : 김영철 e-mail : bob@hongik.ac.kr

텐츠를 제공하는 등의 학습 환경을 조성하여 학습자와 교수자에게 높은 편의성을 제공한다 [1]. 하지만 교육의 성공은 교수자의 지도력, 실시간 피드백 시스템, 학습자들의 동기부여 수준, 지속적인 학습 콘텐츠 제공 및 반응 유도, 개인화된 학습 경로 설계 등의 요소로 결정된다 [2]. 특히, 특정 도메인에서의 교육은 더욱 어려운 실정이다. 그 이유는 특정 도메인에서의 교육은 능력 있는 교수자에게 의존하게 되고 [3], 학습 내용의 진입장벽으로 인해 동기부여가 저하되며, 높은 수준의 도메인별 전문성에 대한 요구 수준에 모든 학습자가 도달하기 어렵기 때문이다. 우리는 이런 한계를 극복할 AI 기술의 발전에 따른 교육 도메인의 변화에 주목했다. AIED는 AI in Education으로, 코로나19 팬데믹 이후 교육 도메인에서 온라인으로의 전환이 요구되었을 때 크게 성장했으며 [4], LMS는 AI 기술이 융합될 수 있는 핵심적인 인프라로서 AIED에서 필수적인 역할을 수행하게 되었다. 그리고 AIED가 개인화된 학습 경로를 제공하고, 맞춤형 피드백 등을 통해 학습의 적응성을 높인다는 것이 여러 연구와 논의에서 밝혀졌다 [1, 5, 6]. 특히 ChatGPT와 같은 생성형 AI의 출시는 교육계에 큰 영향을 미치며 새로운 교수법, 교육 콘텐츠, 교육 과정의 도입을 가능하게 했다 [7, 8].

한편 소프트웨어 인재 양성 및 개발자 교육 도메인에서는 지식 기반 질의나 소프트웨어 개발과 관련된 복잡한 문제에 대해서 강사와 멘토 등 교수자들의 지원이 필수적이다 [3]. 그러나 팀 프로젝트와 같이 협업이 중요한 소프트웨어 개발 교육 환경에서 [9, 10], 교수자들이 모든 교육생들에게 동일한 수준의 지원을 제공할 수 없다는 한계가 존재한다. 이러한 한계는 고품질의 거대한 소프트웨어를 만들기 위해 [11] 복합적인 문제 해결 능력을 요구하는 프로젝트 수행 상황에서 더욱 부각된다.

최근 대규모 언어 모델(LLM)은 자연어 기반의 질문 응답, 요약, 설명, 코딩 분야 등에서 뛰어난

성능을 보인다 [12, 13]. 특히, 최근 LLM은 소프트웨어 개발 교육 분야에서의 적용 가능성을 입증하고 있다. 또한 범용 LLM이 가진 최신 지식 기반 대답의 한계, 특정 분야의 전문성 부재 등을 해결하기 위한 Retrieval-Augmented Generation(RAG) 기법도 등장하였다 [14]. RAG는 LLM이 텍스트를 생성하기 전에 외부 문서를 검색하여 정보 정확성, 도메인 특수성, 지식 최신성을 보장하는 LLM 관련 기술로, 관련 정보 기반 답변 생성이 가능하다 [14]. 우리는 RAG 기법을 기반으로 소프트웨어 개발이라는 특정 도메인의 교육에서 서비스 이용자들에게 LMS 상에서 챗봇 형태의 AI 서비스를 제공하는 것을 목표로 하였다. 논문의 전개는 다음과 같다. 2장에서는 관련 연구를 AIED, LLM 기반 AI Tutor, RAG 기반 AI Tutor 등으로 나누어 언급한다. 3장에서는 시스템 아키텍처와 더불어 RAG의 구현을 위한 데이터 소스, 청킹 전략, 모델 선정 이유, 시스템 프롬프트에 대해 자세히 논하고 4장에서 결론을 서술하며 마무리한다.

II. 관련 연구

본 연구는 Dify와 Ollama를 활용한 소프트웨어 개발 및 개발자 교육 도메인에서의 RAG 기반 챗봇 구현 사례에 초점을 맞추므로 AIED, LLM 기반 AI Tutor, RAG 기반 AI Tutor에 집중하여 조사했다.

1. AIED

COVID-19는 온라인 교육을 강제했고, 보조적인 역할을 수행하던 LMS를 온라인 교육에서 중심에 서게 했다 [15, 16]. 게다가 인공지능 기술의 급속한 발전은 교육 분야에 혁신적인 변화를 가져오고 있으며, 이는 LMS 상에서 학습 경험의 개인화, 교수 지원 등 다양한 측면에서 관찰되고 있다 [17]. R. Luckin et al.은 코로나19 팬데믹

이후 변화한 교육 환경에서, 생성형 AI가 교육의 질을 높이고 교육을 혁신할 잠재력을 가지고 있음을 실증적으로 보여준다 [17].

J. F. V. Oraño et al.은 2017년부터 2024년 10월까지 SCOPUS 데이터베이스에 등재된 학술 문헌에 대해 텍스트 분석을 진행함으로써 AIED가 교수 혁신, 맞춤형 학습 향상 등의 발전에 중요한 역할을 하고 있음을 강조했다 [18].

2. LLM이 결합된 AI Tutor

H. Shahri et al.은 GPT-4의 언어능력과 Invideo AI의 텍스트-비디오 생성 기능을 활용하여 기존의 AI 기반 교육 도구를 극복하는 시스템을 제안했다 [19]. 이 시스템은 GPT-4로 학생들에게 맞춤형의 질문을 준 뒤 답변에 대해 오답이라면 비디오를 생성하여 틀린 개념을 명확히 해주는 과정을 통해, 학생들의 이해를 돕는다. 이 시스템은 다양한 학생들의 다양한 학습 스타일과 요구를 충족하기 위한 LLM 기반의 문제 해결 방안이었지만, LMS와의 연결에 대해 논하지 않았다.

3. RAG 기반 AI Tutor

C. Dong et al.은 적응형 AI 튜터 구축 방법에 대해 논하며, 구조화된 도메인 지식에 기반하도록 하여 표준적인 RAG보다 학습자들의 성과를 개선했음을 정량적으로 입증했다 [20]. 그러나 이 연구는 금융 도메인으로 한정되었다.

R. W. C. Lui et al.은 AI 교육에서 생성형 AI 기반 IT의 효과성에 대해 탐구했다 [21]. 이를 통해 RAG 기반의 GPTutor가 지식 기반의 학습 경험을 제공하고, 시스템 사용이 많을수록 학습 성취도에 긍정적인 영향을 줄 수 있음을 밝혔다. 그러나 해당 연구는 시스템의 효과성 평가에 집중하여, GPTutor가 실제 학습에서 생성한 구체적인 답변 사례를 제시하지 않았다.

III. LMS 내 RAG 기반 챗봇

본 논문의 챗봇 기반의 소프트웨어 개발 및 개발자 교육 시스템은 홍익대학교 메타버스 아카데미에서 자체 개발한 학습 관리 시스템 환경에 통합되어 지원된다. 챗봇 시스템은 Backend-as-a-Service를 지원하는 오픈소스 도구인 Dify[22]와 LLM 실행 환경인 Ollama[23]를 활용하여 RAG를 효과적으로 구현하였다.

1. 시스템 아키텍처

이 시스템의 Server PC 1에서는 Angular 프론트엔드 어플리케이션을 Nginx를 프록시 서버로 사용하여 구동시켰고, NestJS 백엔드 어플리케이션에서 TypeORM을 사용하여 MySQL을 다룬다. Server PC2는 추론 서버로 동작하도록 설계했다. 먼저 Flask를 백엔드 어플리케이션으로 하고 오픈소스인 Dify 플랫폼을 활용했다. LLM 런타임인 Ollama를 Dify에서 모델 제공자로 지정하여 LLM을 쉽게 사용할 수 있도록 설정했다. 이때 Dify의 api 컨테이너와 ollama 컨테이너를 동일한 도커 네트워크에 구성하였다.

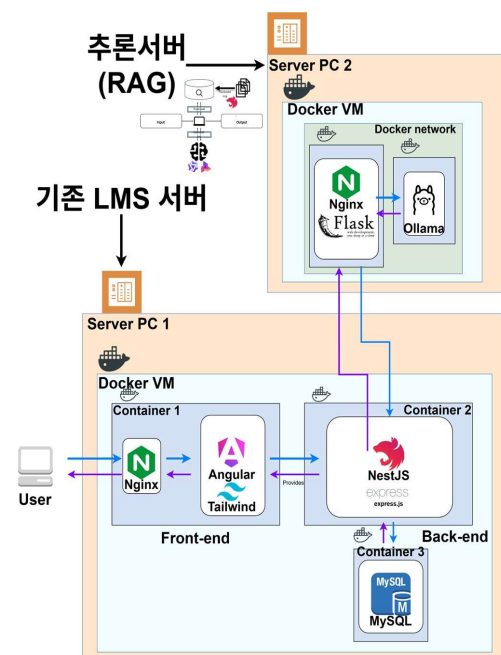


그림 1. LMS 시스템 아키텍처

2. RAG 기반 챗봇 구현

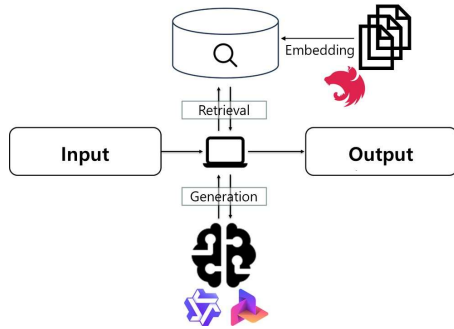


그림 2. RAG(Retrieval Augmented Generation) 구성도

가. 데이터 소스

구축한 검색기의 지식 베이스는 NestJS 공식 문서[24]에서 추출한 134개의 마크다운 파일로 구성된다. NestJS는 현대 웹 어플리케이션 개발을 위한 프레임워크로, 타입스크립트 기반 백엔드 개발에 사용된다. NestJS 공식 문서는 포괄적이고 높은 신뢰성을 가지는, 개발자들에게 일종의 사용 설명서의 역할을 수행한다. 따라서 개발자 교육을 위한 챗봇이 이 문서를 참조한다면, 실질적인 개발에 관한 도전적인 상황을 겪는 초보 개발자들에게 풍부한 자원을 근거로 정확한 답변과 다양한 구현 가능성을 제공할 수 있다.

나. 청크 분할

RAG에서는 성능과 비용 사이에서 최적화하기 위해 Chunking 전략을 사용한다 [25]. 134개의 원본 NestJS 문서 파일이 전체 맥락을 반영하지 못하거나 개별 청크 간 의미적 연결성이 약화될 수 있는 것을 방지하기 위해 부모-자식(Parent-Child) 청크 전략을 사용했다.

다. 임베딩

청크 단위로 분할된 NestJS 공식 문서는 벡터 임베딩 및 인덱싱 처리를 거친다. 임베딩 모델은 Ollama에서 받아온 BGE-M3 모델을 사용한다 [26]. 그 이유는 BGE-M3 모델이 다국어 지원이

뛰어나 홍익대학교 내의 한국인들의 쿼리를 효과적으로 처리할 것으로 예상되었기 때문이다. 임베딩은 모두 Dify 플랫폼 내의 웹 인터페이스에서 간편하게 수행되었다.

라. 생성

생성 모델로는 Qwen 3[27] 와 LG Exaone[28]을 선정하였다. 생성 모델 선정의 이유는 임베딩 모델 선정의 이유와 동일하게, 한국어에 강한 모델일 필요가 있었기 때문이다.

마. 시스템 프롬프트

시스템 프롬프트는 여러 방면에서 LLM의 성능을 향상시키고, 일관되게 유지하는 데 유용하다 [29]. 우리는 시스템 프롬프트를 ‘역할’, ‘답변 지침’, ‘제약조건’으로 나누어 설정함으로써 LLM의 답변 품질의 향상을 시도했다.

```

# 역할
넌 홍익대학교 메타버스 아카데미의 NestJS 프레임워크 개발 봇으로써, 개발 관련 질문에 전문적으로 답변해야 한다.

# 지시
사용자 질문의 의도를 먼저 파악한 뒤, 답변을 체계적이고, 단계적으로 작성해야 한다.

# 제약조건
- 개발 관련 질문이 아니라면 사용자가 개발 관련 질문을 하도록 유도해야 한다.
  
```

그림 3. RAG 기반 챗봇의 시스템 프롬프트

3. LMS 내의 챗봇 사용 예시

아래는 LMS 내의 챗봇 질의응답 사례 세 가지이다.

가. NestJS 프로젝트 구조 Q&A

Q1. NestJS의 프로젝트 구조를 보여줘.

NestJS 프로젝트는 명확한 구조와 모듈화를 기반으로 설계되며, 아래는 일반적인 디렉토리 구조와 주요 파일의 예시입니다:

```

📁 기본 프로젝트 구조
src/
├── main.ts
├── app.module.ts
├── modules/
│   ├── auth/
│   │   ├── auth.module.ts
│   │   └── auth.controller.ts
  
```

```

| | | auth.service.ts
| | | auth.repository.ts
| | | user/
| | | | user.module.ts
| | | | user.controller.ts
| | | | user.service.ts
| | | | user.repository.ts
| | | ...
| | | config/
| | | | config.ts
| | | | environment/
| | | | development.ts
| | | | production.ts
| | | | test.ts
| | | shared/
| | | | utils.ts
| | | | helpers.ts
| | | test/
| | | | auth.test.ts
| | | | user.test.ts
| | | | ...
| | | dto/
| | | | auth.dto.ts
| | | | user.dto.ts
| | | interfaces/
| | | | auth.interface.ts
| | | | user.interface.ts
| | | ...

```

그림 4. LMS 내의 챗봇 질의응답 사례 1-1: 프로젝트 구조에 관한 질의응답 중 주요 내용

★ **팁**
모듈화: 큰 프로젝트에서는 modules/ 디렉토리 안에 각 기능별 모듈을 분리하세요.
환경 관리: config/environment/에서 개발, 프로덕션, 테스트 환경 설정을 관리하세요.
타입 안전성: dto/와 interfaces/를 사용해 타입 안전성을 높이세요.
테스트: test/ 디렉토리에서 각 모듈별 테스트를 작성해 품질 관리를 강화하세요.
 이 구조는 프로젝트 규모에 따라 유연하게 조정할 수 있으며, NestJS의 모듈화 철학을 반영합니다. 😊

그림 5. LMS 내의 챗봇 질의응답 사례 1-2: 프로젝트 구조에 관한 질의응답 중 주요 내용

나. NestJS 파일 데이터 송수신 Q&A

Q2. NestJS에서 파일 형식의 데이터를 송수신하는 방법에 대해서 알려줘.

NestJS에서 파일 형식의 데이터를 송수신하는 방법은 주로 **Multer**라는 미들웨어를 사용하여 구현됩니다. Multer는 HTTP multipart/form-data 형식을 처리하여 파일을 업로드하고, 서버에서 파일을 저장하거나 처리하는 데 사용됩니다. 아래는 단계별로 설명한 방법입니다.

###1. 필요한 패키지 설치

```

bash
npm install @nestjs/common @nestjs/platform-express multer

```

###2. Multer 미들웨어 설정

NestJS 애플리케이션에서 Multer를 사용하려면 MulterModule을 설정해야 합니다.

multer 라이브러리에서 제공하는 diskStorage를 사용하여 파일을 디스크에 저장할 수 있습니다.

```

typescript
// app.module.ts
import { Module } from '@nestjs/common';
import { MulterModule } from '@nestjs/platform-express';

```

⋮

요약

Multer 설치 및 설정 → 파일 저장 경로와 이름 설정.

컨트롤러에서 @UploadedFile() 사용 → 파일 처리.

클라이언트에서 multipart/form-data 형식으로 요청.

필터링 및 에러 처리 → 파일 유효성 검증.

이 방법을 통해 NestJS에서 파일 업로드 및 다운로드를 안정적으로 처리할 수 있습니다.

그림 5. LMS 내의 챗봇 질의응답 사례 2: NestJS 프레임워크에서 파일 데이터 송수신하는 방법에 관한 질의응답 중 주요 내용

다. NestJS Controller 코드 예제 Q&A

Q3. NestJS 프레임워크로 게시판을 만들고 싶어. Controller 계층의 코드를 작성해줘.

HTTP 메서드 매핑

POST /posts: 게시물 생성

GET /posts: 모든 게시물 조회

GET /posts/:id: 특정 게시물 조회

PUT /posts/:id: 게시물 수정

DELETE /posts/:id: 게시물 삭제

그림 6. LMS 내의 챗봇 질의응답 사례 3-1: NestJS 프레임워크 Controller 코드 생성 질의응답 중 주요 내용

다음은 NestJS 프레임워크를 사용하여 게시판 기능을 처리하는 Controller 계층의 코드 예시입니다. 이 예제는 게시물(Post)의 생성, 조회, 수정, 삭제 기능을 포함합니다.

posts.controller.ts

```

@Controller('posts')
@UseInterceptors(ClassSerializerInterceptor)
export class PostsController {
  constructor(private readonly postsService: PostsService) {}

  @Post()
  @UsePipes(new ValidationPipe())
  createPost(@Body() createPostDto: PostEntity): Promise {
    return this.postsService.createPost(createPostDto);
  }

  @Get()
  getAllPosts(): Promise<PostEntity[]> {
    return this.postsService.getAllPosts();
  }

  @Get('/:id')
  getPostById(@Param('id') id: string): Promise {
    return this.postsService.getPostById(+id);
  }

  @Put('/:id')
  @UsePipes(new ValidationPipe())
  updatePost(
    @Param('id') id: string,
    @Body() updatePostDto: Partial,
  ): Promise {
    return this.postsService.updatePost(+id, updatePostDto);
  }

  @Delete('/:id')
  deletePost(@Param('id') id: string): Promise {
    return this.postsService.deletePost(+id);
  }
}

```

그림 7. LMS 내의 챗봇 질의응답 사례 3-2: NestJS 프레임워크 Controller 코드 생성 질의응답 중 주요 내용

IV. 결 론

기존 LMS에는 여러 한계가 존재했다. 이 한계를 극복하기 위해, 우리는 RAG 기반 챗봇을 LMS 환경에 통합하는 방법을 제안한다. 이 시스템은 NestJS 공식 문서에서 도출한 전문 지식을 데이터 소스로 활용한다. 또한, Dify 플랫폼과 Ollama를 통해 임베딩 모델과 LLM 모델을 쉽게 사용하였다. 결과적으로, 우리는 소프트웨어 개발 및 개발자 교육 도메인에서 맞춤형 Q&A 챗봇 서비스를 성공적으로 구현하였다.

REFERENCES

- [1] P. D. Barua, X. Zhou, R. Gururajan, and K. C. Chan, "Determination of Factors Influencing Student Engagement Using a Learning Management System in a Tertiary Setting", in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 604-609, 2018.
- [2] A. Y. Q. Huang, O. H. T. Lu, and S. J. H. Yang, "Effects of artificial Intelligence - Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom", *Computers & Education*, vol. 194, p. 104684, 2023.
- [3] M. R. Marques et al., "A systematic mapping study on practical approaches to teaching software engineering", in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, Oct. 2014.
- [4] P. Veluvali and J. Suriseti, "Learning Management System for Greater Learner Engagement in Higher Education - A Review", *Higher Education for the Future*, vol. 9, no. 1, Oct. 2021.
- [5] U.S. Department of Education, "Artificial Intelligence and the Future of Teaching and Learning", May 2023.
- [6] World Bank Group, "Artificial Intelligence Revolution in Education: What You Need to Know", Jun. 2024.
- [7] A. Böttcher et al., "Concepts for Teaching Software Development in the Age of AI-Tools," *2025 IEEE Global Engineering Education Conference (EDUCON)*, 2025.
- [8] M. Virvou and G. A. Tsihrantzis, "Is ChatGPT Beneficial to Education? A Holistic Evaluation Framework Based on Intelligent Tutoring Systems," *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 2023.
- [9] 윤봉식, 유소월, "창의SW 교양교과의 팀프로젝트 교육 성과 분석", *스마트미디어저널*, 제14권, 제1호, 35-43쪽, 2025년
- [10] 정유경. "핵심역량 (4C) 증진을 위한 액티브러닝과 퍼실리테이션 융합교육프로그램 연구," *스마트미디어저널*, 제8권, 제1호, 67-73쪽, 2019년
- [11] 정문재. "소프트웨어 개발 시 프로젝트의 크기가 구현에 미치는 영향," *스마트미디어저널*, 제1권, 제4호, 79-83쪽, 2012년
- [12] N. Jain et al., "LiveCodeBench: Holistic and Contamination Free Evaluation of Large Language Models for Code," *arXiv preprint arXiv:2403.07974*, Mar. 2024.
- [13] C. E. Jimenez et al., "SWE-bench: Can Language Models Resolve Real-World GitHub Issues?", *arXiv preprint arXiv:2310.06770*, Oct. 2023.
- [14] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", *arXiv preprint arXiv:2005.11401*, May 2020.
- [15] S. Qazi et al., "AI-Driven Learning Management Systems: Modern Developments, Challenges and Future Trends during the Age of ChatGPT", *Computers, Materials and Continua*, vol. 80, no. 2, pp. 3289-3314, Aug. 2024.
- [16] F.-V. Pantelimon et al., "The Evolution of AI-Driven Educational Systems during the COVID-19 Pandemic", *Sustainability*, vol. 13, no. 23, p. 13501, Dec. 2021.
- [17] R. Luckin et al., *Intelligence Unleashed: An argument for AI in Education*. Pearson, 2016.
- [18] J. F. V. Oraño et al., "Analyzing the Landscape of Generative and Open AI in Education: A Text Analytics Exploration of Scopus Publications", in *2025 14th International Conference on Educational and Information Technology (ICEIT)*, Mar. 2025.
- [19] H. Shahri et al., "Elevating Education through AI Tutor: Utilizing GPT-4 for Personalized Learning", in *2024 15th Annual Undergraduate Research Conference on Applied Computing (URC)*, Apr. 2024.
- [20] C. Dong et al., "How to Build an Adaptive AI Tutor for Any Course Using Knowledge Graph-Enhanced Retrieval-Augmented Generation (KG-RAG)", *arXiv preprint arXiv:2311.17696*, Nov. 2023.
- [21] R. W. C. Lui et al., "GPTutor: A Generative AI-powered Intelligent Tutoring System to Support Interactive Learning with Knowledge-Grounded Question Answering", in *2024 International Conference on Advances in*

Electrical Engineering and Computer Applications (AEECA), Aug. 2024.

- [22] Dify. [Online]. Available: <https://dify.ai/>. (accessed Jun., 19, 2025).
- [23] Ollama. [Online]. Available: <https://ollama.com/>. (accessed Jun., 19, 2025).
- [24] Documentation | NestJS - A progressive Node.js framework. [Online]. Available: <https://docs.nestjs.com/>. (accessed Jun., 19, 2025).
- [25] C. Merola and J. Singh, "Reconstructing Context: Evaluating Advanced Chunking Strategies for Retrieval-Augmented Generation," arXiv:2504.19754 [cs.IR], 2025.
- [26] J. Chen et al., "BGE M3-Embedding: Multi-Lingual, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation", arXiv preprint arXiv:2402.03216, Feb. 2024.
- [27] A. Yang et al., "Qwen3 Technical Report", arXiv preprint arXiv:2505.09388, May 2025.
- [28] LG AI Research et al., "EXAONE 3.5: Series of Large Language Models for Real-world Use Cases", arXiv preprint arXiv:2412.04862, Dec. 2024.
- [29] L. Zhang et al., "SPRIG: Improving Large Language Model Performance by System Prompt Optimization", arXiv preprint arXiv:2410.14826, Oct. 2024.

저 자 소 개



김재호(준회원)

2025년 홍익대학교 소프트웨어융합학과 학사 졸업.

<주관심분야 : 대규모 언어 모델, 검색 증강 생성, 프롬프트 엔지니어링, 웹 소프트웨어 설계 및 최적화>



서채연(준회원)

1999년 홍익대학교 전자전산학과 학사 졸업.

2005년 홍익대학교 전자전산공학 석사 졸업.

2014년 홍익대학교 전자전산공학 박사 졸업.

<주관심분야 : Mapping Business Process Model(BPM) with CBD, Component Repository, 소프트웨어공학, 메타버스, AR/VR, 생성형 AI, 웹, 언리얼엔진5, 디자인씽킹>



R. Young Chul Kim(정회원)

2001년-현재: 홍익대학교 교수