

RAG 적용 VR 지상 무인체 훈련 어시스턴트 GPT 시스템 개발

우동현*, 유선진**

Development of a GPT System for VR Ground Unmanned Vehicle Training Assistants with RAG

Donghyun Woo*, Sunjin Yu**

본 논문은 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다.(2021RIS-003) 또한, 2023년도 국방기술품질원의 재원으로 방산혁신클러스터 지원(DCL2020L, 2020년 방산혁신클러스터 방산 소재 부품연구실 사업)을 받아 수행된 연구의 일부임

요약

본 연구는 VR 시뮬레이션에 GPT 모델과 RAG(Retrieval-Augmented Generation)를 접목해 무인 지상 차량 훈련의 정확성과 신뢰도를 높이는 방안을 제시한다. 유니티(Unity) 기반 VR 환경에 Whisper 음성 인식 및 TTS를 연동한 대화형 어시스턴트를 구현했으며, GPT 환각 문제를 줄이기 위해 외부 DB 스니펫을 참조하는 구조를 도입했다. 운전 경험이 있는 25명을 대상으로 한 실험에서, RAG 적용 모드의 SUS(System Usability Scale) 점수가 약 14% 높게 나타났고 환각도 유의미하게 감소했다. 문서 검색으로 인한 응답 지연이 평균 0.6 초 발생했으나, 참가자들은 이를 수용 가능한 수준으로 평가했다. 본 연구는 VR 기반 군사훈련 분야에서 RAG를 적용한 GPT 시스템이 높은 활용 가능성을 지님을 시사한다. 향후에는 보안·성능 최적화와 더불어, 대규모 DB 관리 및 네트워크 환경에서 발생할 수 있는 추가 지연을 최소화하기 위한 후속 연구가 필수적이다.

Abstract

This study enhances the accuracy and reliability of unmanned ground vehicle training by integrating a GPT model with Retrieval-Augmented Generation(RAG) in a Virtual Reality(VR) simulation. A voice-interactive assistant was developed using Unity, Whisper speech recognition, and text-to-speech, referencing external database snippets to reduce hallucinations. In tests with 25 participants, the RAG-based scenario achieved a 14% higher System Usability Scale(SUS) score while significantly lowering hallucination rates. Although a 0.6-second increase in response time was observed, users considered it acceptable for improved information accuracy. These findings highlight the potential of GPT-RAG systems in VR-based military training, emphasizing future research on security, performance optimization, and large-scale database management.

Keywords

virtual reality, military training simulation, GPT, hallucination, RAG

* 국립창원대학교 첨단방위공학과정 박사과정
- ORCID: <https://orcid.org/0000-0002-5562-4158>
** 국립창원대학교 문화테크노학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-9292-4099>

• Received: Jan. 17, 2025, Revised: Feb. 12, 2025, Accepted: Feb. 15, 2025
• Corresponding Author: Sunjin Yu
Dept. of Culture Technology, Changwon National University, 20
Changwondaechak-ro, Uichang-gu, Changwon-si, Gyeongsangnam-do, 51140, Korea
Tel.: +82-55-213-3098, Email: sjyu@changwon.ac.kr

I. 서론

최근 국방 분야에서는 첨단 기술을 활용하여 미래 전장 환경을 효과적으로 모사하고, 안전하고 반복적인 방식으로 전술 능력을 극대화하려는 노력이 활발히 진행되고 있다[1]. 가상현실(VR, Virtual Reality)은 높은 몰입감, 자유도, 그리고 물리적 위험 감소라는 특성을 바탕으로, 군사훈련 시뮬레이션의 핵심 도구로 자리 잡고 있다[2]. VR 시뮬레이션은 차량 조작, 항공기 이착륙, 대규모 전술 시나리오 등의 훈련을 비교적 저비용으로 반복 숙달할 수 있는 장점을 제공한다. 최근에는 무인 지상 차량(UGV, Unmanned Ground Vehicle)과 같은 무인체계 운용이 군사훈련에서 점점 더 중요해짐에 따라, VR 시뮬레이션을 활용하여 이러한 전술 및 운용 절차를 사전에 실습하려는 연구가 증가하고 있다[3].

그러나 VR 시뮬레이션은 물리적 환경 재현과 기본적인 상호작용 제공에서는 뛰어난 성능을 보이지만, 복잡한 전술 결정이나 상황별 질의응답을 처리하는 데는 한계를 보인다. 기존 VR 시스템은 고정된 시나리오와 사전 정의된 스크립트에 따라 진행되므로, 동적이고 예측 불가능한 군사 상황에서 사용자 질문에 실시간으로 대응하기 어렵다. 이러한 문제를 해결하기 위해 최근에는 대규모 언어 모델(LLM, Large Language Model)을 VR에 결합하여, 훈련 중 발생하는 다양한 질의응답을 처리하고 전술적 조언을 제공하려는 시도가 이루어지고 있다. 특히 GPT 계열 모델은 대규모 텍스트 데이터를 학습하여 높은 언어 생성 능력을 보유하고 있으며, 군사 관련 질의에서도 일정 수준 의미 있는 답변을 제공할 수 있다는 가능성을 제시하고 있다[4][5].

하지만 GPT 모델은 환각(Hallucination) 문제라는 고질적 한계를 지니고 있다. 이는 모델이 학습 데이터에 존재하지 않는 정보나 문맥적으로 부합하지 않는 내용을 사실인 것처럼 생성하는 현상으로, 군사훈련처럼 고정확성을 요구하는 맥락에서는 치명적인 위험 요소가 된다[6]. 예를 들어, 전술 매뉴얼에 없는 무기체계를 제시하거나 잘못된 적군 배치 정보를 생성할 경우, 훈련자는 이를 사실로 받아들여 잘못된 판단을 내릴 위험이 크다. 이를 해결하기

위해 RAG(Retrieval-Augmented Generation) 기법을 도입하고 있다. RAG는 LLM이 답변을 생성하기 전에 외부 데이터베이스(DB)를 검색하고, 문서 스니펫을 참조함으로써 환각 문제를 완화할 수 있는 기법이다. 이를 통해 LLM의 신뢰도를 향상시키고, 군사훈련과 같이 도메인 지식이 중요한 분야에서 높은 활용 가능성을 제공할 수 있다[7].

본 연구는 VR 시뮬레이션에서 GPT 모델과 RAG 기법을 결합하여, 지상 무인체 운용 훈련에서의 정확성과 실효성을 검증하는 데 목적을 둔다. 구체적으로, GPT 4o 모델을 기반으로 무인 지상 차량이 정찰 및 대응 임무를 수행하는 시나리오를 구현하고, RAG를 적용한 경우와 적용하지 않은 경우를 비교 분석한다. 이를 통해 사용자 만족도, 정보 신뢰도에서 나타나는 차이를 정량적으로 평가한다. 또한 Whisper 기반 STT(Speech-To-Text) 및 TTS(Text-To-Speech) 기술을 적용하여 음성 질의응답 시스템을 구현함으로써, 사용자가 대화형 조언을 받을 수 있는 몰입형 훈련 환경을 제공한다.

본 논문의 구성은 다음과 같다. II장에서는 VR 시뮬레이션 개발과 군사훈련 적용, GPT 계열 모델과 환각 문제, 그리고 RAG 기법과 관련된 선행연구에 대하여 검토 및 비교분석한다. III장에서는 제안하는 VR 시뮬레이션 시스템과 GPT 기반 지상 무인체 어시스턴트의 설계 및 구현 방안에 대하여 논의한다. IV장에서는 사용자 설문 등을 중심으로 RAG 적용 효과에 대하여 분석한다. 마지막 V장에서는 본 연구가 가지는 의의와 한계, 그리고 향후 연구 방향을 제시한다.

II. 배경 이론

2.1 VR 시뮬레이션 개발과 군사훈련 적용

가상현실은 물리적 현실을 가상공간에서 재현하여 사용자가 실제와 유사한 경험을 안전하고 비용 효율적으로 반복적으로 수행할 수 있게 하는 기술이다. 군사훈련에서는 전술 숙달, 무기체계 운용, 위험지역 정찰 등 다양한 시나리오를 현실감 있게 구현하기 위해 VR 기술이 적극적으로 도입되고 있다.

특히, 무인 지상 차량의 조작, 정찰, 교전 절차는 실제 환경에서 높은 위험성을 수반하므로 반복 훈련이 어려운 경우가 많다. 그러나 VR 환경에서는 이러한 위험 없이 동일한 상황을 여러 차례 시뮬레이션함으로써 작전 효율성과 수행 능력을 향상시킬 수 있다. 이러한 장점으로 국내외에서 VR 기술과 시뮬레이터를 결합하여 개인 훈련에서 전술 훈련에 이르기까지 다양한 활용 범위가 확장되고 있다[8].

VR 시뮬레이션이 시각적 및 청각적 몰입감을 제공하는 데에는 강점이 있으나, 훈련 중 발생할 수 있는 복잡한 질의나 전술적 의사결정 지원 기능은 제한적인 경우가 많다. 예를 들어, 전차 시뮬레이션이나 항공 시뮬레이션은 고정된 시나리오와 스크립트에 기반하여 진행되는 경우가 대부분이기 때문에, 사용자가 훈련 중 상호작용형 질문을 제기했을 때 실시간으로 대응하는 데 한계가 있다. 이러한 문제를 해결하기 위해, 최근에는 군사훈련용 VR 시스템에 인공지능 기반 대화형 기능을 접목하려는 연구가 활발히 진행되고 있다. 사용자의 자유로운 질의 응답을 지원하고, 전술적 조언을 제공함으로써 기존 VR 훈련 시스템의 한계를 보완하는 데 기여할 수 있을 것으로 기대된다.

2.2 GPT 계열 모델과 환각 문제

대규모 언어 모델은 방대한 텍스트 데이터를 학습하여 기존의 규칙 기반 챗봇이나 AI 시스템에 비해 자연스러운 응답을 생성할 수 있다. 특히, GPT-4와 같은 고도화된 LLM은 정교한 맥락 이해와 추론 능력을 바탕으로 다양한 분야에서 활용 가능성이 높으며, 군사적 응용 분야에서도 사용자 질의에 대한 전술적 인사이트를 제공할 수 있는 잠재력을 보유하고 있다. 예를 들어, 무인체 운용 시나리오에서 GPT 계열 모델은 학습된 데이터를 기반으로 사용자 요구에 적합한 분석과 추론을 수행할 수 있다.

그러나 GPT 계열 모델 및 대규모 언어 모델 전반에는 환각 문제라는 중요한 한계가 존재한다. 이는 모델이 학습 데이터에 존재하지 않거나 문맥적으로 부합하지 않는 정보를 생성하여 마치 사실인 것처럼 사용자에게 제시하는 현상을 말한다. 특히

군사 훈련 및 시뮬레이션 분야에서는 이러한 환각 문제가 심각한 위험 요소가 될 수 있다. 모델이 생성한 부정확한 정보는 훈련 효과를 저하시킬 뿐만 아니라 훈련자에게 잘못된 전술적 판단이나 비효율적인 학습 습관을 형성할 우려가 크다[9]. 따라서, 환각 문제를 최소화하거나 이를 완화하기 위한 모델 설계 및 활용 방안에 대한 연구가 필수적이다.

2.3 RAG 기법

RAG 기법은 대규모 언어 모델이 내재된 확률 기반 추론만으로 답변을 생성하기 전에, 외부 지식베이스에서 관련 문서를 검색하고 이를 모델 입력에 포함함으로써 환각 문제를 완화하는 방식이다. RAG 기법은 모델이 무작위로 잘못된 정보를 생성하는 대신, DB에 저장된 신뢰할 수 있는 근거(전술 매뉴얼, 기술 보고서 등)를 우선적으로 참고하도록 하여 답변의 정확성과 신뢰도를 향상시키는 데 기여한다.

특히 군사훈련과 같은 도메인 지식이 풍부하게 축적된 분야에서 RAG 기법은 의미 있는 대안으로 주목받고 있다. 예를 들어, 실시간 전술적 의사결정을 지원하거나 복잡한 무기체계의 작동 원리를 설명하는 데 있어 RAG는 기존 모델이 제공할 수 없던 신뢰도 높은 응답을 생성할 수 있다. 또한, RAG를 활용하면 사용자 질문에 대한 응답이 특정 매뉴얼이나 문헌에 기반하므로, 사용자에게 더욱 설득력 있는 정보를 제공할 수 있다.

그러나 RAG 기법의 실질적 적용에는 몇 가지 제약 조건이 따른다. 첫째, 검색 프로세스에 소요되는 시간이 실시간 응답 성능에 영향을 줄 수 있다. 둘째, DB의 품질과 최신성은 RAG의 성능에 직접적인 영향을 미친다. 셋째, 군사 도메인에서는 보안이 중요한 이슈로 부각되며, 민감한 데이터를 포함한 DB 관리는 철저히 이루어져야 한다.

또한, RAG 기법을 실제 시스템에 적용할 경우 검색 성능과 시스템 부하에 대한 고려가 필수적이다. 군사훈련 시뮬레이션은 실시간 상호작용이 요구되는 환경이므로, 매 질의마다 외부 DB를 검색하고 스니펫을 추출하는 과정에서 발생하는 응답 지연이 훈련 몰입도를 떨어뜨릴 수 있다.

따라서 대규모 문서를 신속하게 검색할 수 있는 고성능 벡터 인덱싱 및 병렬 검색 구조를 설계해야 하며, 캐싱(Caching) 및 사전 로드(Preloading) 기법 등을 통해 검색 지연을 최소화해야 한다[10]. 더불어 군사 도메인에서는 문서가 매우 방대한 동시에 보안 문제가 중요하므로, RAG에 활용되는 DB의 최신성·무결성·접근 권한을 지속적으로 관리해야 한다. 문서 업데이트 시 임베딩을 재생성하는 작업, 검색 정확도 향상을 위한 메타데이터 관리 등은 추가적인 처리 부하를 야기한다. 이처럼 RAG 기법은 환각 문제를 완화하는 효과적 수단이지만, 이를 효과적으로 운용하기 위해서는 시스템 성능과 데이터 보안 을 함께 고려하는 종합적인 접근이 요구된다[11].

III. 시뮬레이션 설계 및 구현

VR 환경에서 지상 무인체(UGV)를 운용하며, 사용자가 음성으로 GPT 4o 모델에게 질의하면 조언을 제공하는 ‘대화형 어시스턴트’를 설계·구현하는 데 있다. 해당 어시스턴트는 정찰 임무, 적 조우 시 대응 여부, 돌발 장애물 회피 같은 군사적 의사결정을 실시간으로 보조하도록 설계되었다. 다만 GPT 계열 모델 자체는 환각 문제로 인해 잘못된 정보를 제시할 우려가 있어, 이를 최소화하기 위해 RAG 기법을 도입하여 군사 매뉴얼·무인체 운용 보고서 등 DB 근거를 활용하는 구조를 구현한다.

3.1 Unity 기반 VR 시뮬레이션 개발 환경

본 연구에서는 유니티(Unity) 2021 버전을 활용하여 VR 시뮬레이션을 개발하였다. 유니티는 3D 콘텐츠 구현의 용이성, 다양한 에셋 및 플러그인 지원, 주요 VR 헤드셋과의 높은 호환성으로 VR 콘텐츠 개발에 적합하다. 본 시뮬레이션은 군사 지형 데이터를 기반으로, 무인 지상 차량이 주행할 수 있는 도로, 장애물, 위험 요소 등을 그림 1과 같이 가상 환경에 구현하였다. 또한, 라이더(LiDAR) 센서가 UGV에 탑재된 것으로 가정하고, 유니티의 C# 스크립트를 활용하여 센서의 탐지 범위 내에서 적 오브젝트를 인식하고 경고하도록 설계하였다. 이 과정에서 센서의 탐지 거리, 감지 정확도 등의 주요 파라미터를 사용자가 제어할 수 있도록 구현하였다.



그림 1. 유니티기반 VR 시뮬레이션 환경
Fig. 1. Unity-based VR simulation environment

UGV는 차량형 모델로 설정되었으며, 유니티의 내장 물리엔진을 활용하여 바퀴 수, 차량 길이, 중량, 회전 반경 등 주요 물리적 속성을 조정하였다. 이를 통해 사용자는 실제 차량 조작과 유사한 주행 경험을 체감할 수 있도록 하였다. 또한, 건물, 나무, 바위와 같은 장애물 오브젝트에 충돌 판정용 콜라이더를 배치하여, 무인체의 주행 경로가 잘못 설정될 경우 물리적 충돌이 발생하도록 설계하였다. 사용자는 VR HMD(Head-Mounted Display)와 드라이빙 휠(핸들 및 페달)을 사용하여 시뮬레이션을 조작할 수 있다.

3.2 GPT 4o 모델 연동 및 프롬프트 설계

유니티와 GPT 4o API를 연동하기 위해, 유니티와 Python(LLM API) 간 통신을 REST 방식으로 구현하였다. 사용자가 질의하면 유니티에서 사용자 음성 질의 텍스트를 GPT 4o API로 요청을 보내고 응답을 받아 다시 유니티로 반환하여 사용자에게 전달한다[12].

GPT 4o 모델이 군사 훈련 맥락에서 무인체 운용 보조 역할을 수행하도록 하기 위해, 시스템 메시지에 다음과 같은 설명을 삽입한다. ‘당신은 지상 무인체 운용 보조 AI이다. 사용자 질문이 오면, 군사 매뉴얼·전술 보고서 등을 참고해 가능한 한 정확하고 현실성 있는 조언을 제시하라.’ 사용자 질의나 RAG 검색 결과를 User Prompt 형태로 전달한다. ‘전방의 건물은 위험요소로 의심되는데, 우회로를 어떻게 설정해야 하는가?’ 같은 질의를 넣으면, GPT 4o은 내부 지식 또는 추가 스니펫 정보를 결합해 최종 답변을 생성한다.

3.3 RAG 기법의 DB 검색 구조와 스니펫 생성

그림 2는 유니티와 GPT 4o를 연동하여 RAG 기법을 적용한 시스템의 전체 흐름을 나타낸다. RAG를 적용하기 위해 군사 자료(UGV 운용 매뉴얼, 전술 보고서, 훈련교범 등)를 수집하였고, PDF·TXT 등 형태를 일괄 텍스트로 변환해 TF-IDF 및 문장 임베딩 인덱스를 구축했다. 수집된 문서는 아래 수식 (1)에 제시된 방식으로 형태소 분석 후 TF-IDF 가중치 기반 인덱싱을 구축한다[13].

$$TF-IDF(t,d) = TF(t,d) \times \log \frac{N}{DF(t)} \quad (1)$$

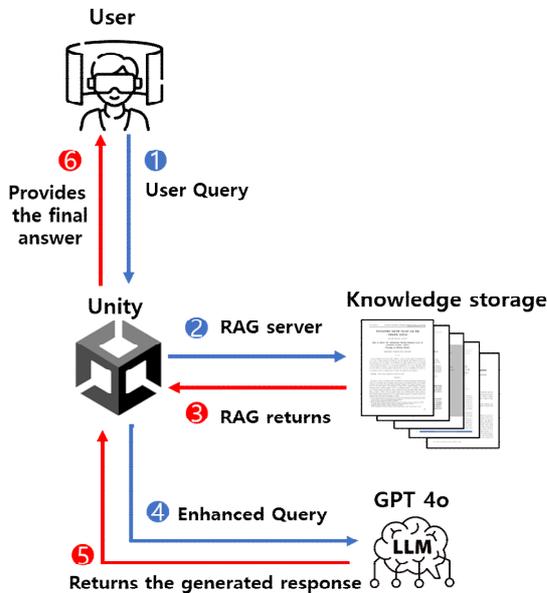


그림 2. 유니티-GPT 연동 시스템 흐름도
Fig. 2. Unity-GPT Integration system flow diagram

위 식에서 $TF(t,d)$ 는 단어 t 가 문서 d 에 등장하는 빈도, N 은 전체 문서 수, $DF(t)$ 는 단어 t 가 등장하는 문서 수를 의미한다. 더불어 문장 의미를 파악하기 위해 Sentence-Transformers 기반 임베딩도 병행하여, 유사도 검색에도 대비하였다. 특정 단어 t 가 문서 d 에서 나타날 때, 아래의 TF-IDF 값을 기준으로 문서를 검색할 수 있도록 준비한다. TF-IDF 인덱스는 키워드 기반 고속 검색에, 임베딩 인덱스는 유사 의미 문장 검색에 유리하다.

그림 3의 흐름도와 같이, RAG 서버가 질의를 받으면, TF-IDF와 임베딩 검색으로 상위 k 개 문서를

찾은 뒤, 각 문서에서 사용자 질문과 가장 관련도 높은 문단(스니펫)을 발췌한다. 문서 스니펫이 결정되면, 이를 포함한 프롬프트를 LLM API에 전달한다. 사용자가 질의를 보내면, TF-IDF 상위 3건, 임베딩 상위 3건을 추출해 중복 제거한 뒤, 가장 연관도 높은 문단 스니펫을 결합해 GPT 4o에 전달한다. 동일한 VR 시나리오에서, RAG를 적용한 시뮬레이션은 DB 검색 과정을 수행하며, RAG 미적용 시뮬레이션은 검색 없이 GPT 4o에 질의만 보낸다. 이를 실험적으로 비교하기 위해, 두 시뮬레이션을 분리해서 실행하도록 설계했다. 프롬프트 말미에는 ‘아래 문서에 없는 내용을 함부로 추측하지 마세요’는 지시문을 넣어 GPT가 스니펫 범위를 벗어난 내용은 최소화하도록 구현한다.

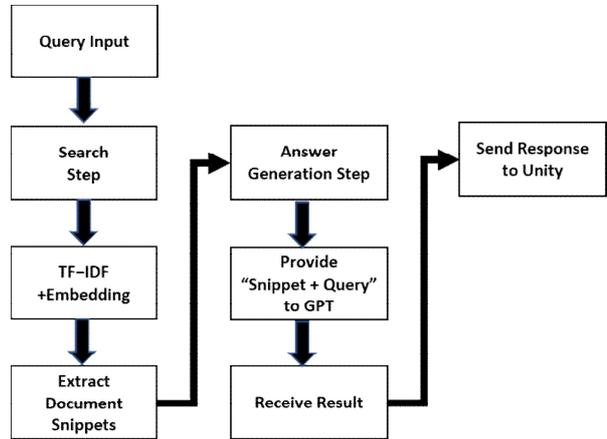


그림 3. RAG 모듈 세부 흐름도
Fig. 3. Detailed flow diagram of the RAG module

본 연구에서는 문서 청킹(Chunking)과 벡터 임베딩 기반 검색을 다음과 같은 절차로 구축하였다. 우선, 군사 매뉴얼·전술 보고서·훈련교범 등에서 추출한 텍스트를 약 600토큰 단위로 분할하고, 앞뒤로 50토큰씩 오버랩(Overlap)하여 각 청크를 생성하였다. 이렇게 분할된 청크마다 Sentence-Transformers 기반 임베딩을 계산한 뒤, 라이브러리를 활용한 벡터 데이터베이스에 저장한다. 검색 과정에서는 다음과 같이 진행된다. 첫째, 사용자의 질의가 들어오면 TF-IDF로 상위 3개 문서를 우선 필터링하고, 동시에 벡터 검색(top-3) 결과를 조회한다. 둘째, TF-IDF와 벡터 검색 결과를 취합하여 가장 유사도가 높은 청크를 최대 5개까지 후보로 선정한다.

셋째, 선정된 청크에서 질의와 직접적으로 연관된 문장을 발췌하여 스니펫을 구성한다. 이 스니펫은 GPT 4o 모델에 추가 프롬프트로 주어지므로, 모델이 실제 군사 문헌 정보를 기반으로 답변을 생성하도록 유도할 수 있다. 청킹 단위와 오버랩 범위, 검색 알고리즘 등은 정확도와 응답 속도 사이에서 최적의 균형점을 찾기 위해 여러 차례 실험을 통해 조정하였으며, 추후 문서가 대폭 증가하거나 다른 언어 모델을 사용하게 될 경우에는 이러한 파라미터를 재조정해줄 필요가 있다.

3.4 Whisper STT와 TTS 연동

본 연구에서는 VR 시뮬레이션에서 음성 질의 및 응답 시스템을 구현하기 위해 Whisper 기반의 STT와 TTS 모듈을 연동하였다. 사용자는 VR 환경에서 마이크 버튼을 눌러 질의하며, 유니티는 음성 데이터를 WAV 형식으로 캡처하고, 일정 길이로 분할하여 전송한다. 전송된 음성 데이터를 전처리한 뒤 Whisper 모델에 입력하여 텍스트로 변환한다[14]. Whisper는 잡음 환경에서도 비교적 높은 인식률을 보이지만, 군사 현장 수준의 소음과 사투리, 구어체 같은 변이형 언어 처리에는 한계가 있다.

텍스트로 변환된 질의는 GPT 4.0 API에 전달되어 응답이 생성되며, 이 응답은 TTS 엔진(Google TTS API)을 활용해 음성 파일로 변환된다. 변환된 음성 파일은 다시 유니티로 전송되어 AudioClip 형태로 로드되며, VR 헤드셋 스피커를 통해 사용자에게 재생된다. 이를 통해 사용자는 운전 중 손을 떼지 않고도 GPT 응답을 청취할 수 있어 몰입감 높은 훈련 환경이 조성된다.

IV. 실험 및 결과 분석

4.1 연구 목적 및 실험 설계

본 연구는 VR 시뮬레이션에서 RAG 기법의 효과를 평가하기 위해, RAG 적용 여부에 따른 사용자 경험의 차이를 분석하였다. 이를 위해 동일한 시나리오를 체험하는 실험을 설계하였으며, 실험 참가자

는 운전 경험과 VR 사용 경험이 있는 성인 25명으로 구성되었다. 실험에 앞서 참가자들은 5분간의 튜토리얼을 통해 컨트롤러 조작법과 음성 질의 방식을 학습하였다. 실험은 두 가지 시나리오로 구성되었다. 첫 번째로, 정찰 시나리오에서는 무인 지상 차량이 의심 지역으로 이동하여 적군 존재 여부를 탐지하도록 설계되었다. 두 번째로, 대응 시나리오에서는 적군과 조우하거나 장애물이 등장했을 때 회피 경로를 결정하도록 구성하였다. 참가자들은 먼저 RAG 미적용 모드에서 시나리오를 체험한 뒤, 같은 시나리오를 RAG 적용 모드에서 다시 체험하도록 하였다. 이러한 실험 설계를 통해 동일한 상황에서 RAG가 사용자 경험에 미치는 영향을 정량적으로 비교·분석하였다.

한편, RAG 미적용 시나리오에서는 군사 매뉴얼·전술 보고서 등 해당 분야 문서를 GPT 4o 모델에 제공하지 않고, ‘당신은 지상 무인체 운용 보조 AI이다’라는 최소한의 시스템 메시지만 주어진 상태에서 질의응답을 진행하였다. 즉, RAG 미적용 모드에서는 실제 군사 자료를 별도로 참조하지 않은 채 모델이 일반적 사전 학습 지식만으로 답변하게끔 설정하였다. 반면, RAG 적용 시나리오에서는 3장에서 언급한 청킹과 벡터 DB 기반 검색 절차를 통해 추출된 스니펫을 GPT 4o에 추가로 제공함으로써, 실제 군사 문헌 정보를 적극 반영하도록 구현하였다. 이를 통해 동일한 VR 시나리오에서 문서 기반 답변과 비문서 기반 답변의 차이를 비교함으로써, RAG 기법의 효과성을 평가하고자 하였다.

4.2 측정 지표

실험의 측정 지표로는 SUS(System Usability Scale) 설문과 본 연구에 특화된 자체 추가 문항을 활용하였다. SUS는 학습 난이도, 조작성, 안정성 등 사용성을 평가하는 10개 문항으로 구성되며, 응답은 5점 척도(1점=전혀 아니다 ~ 5점=매우 그렇다)로 수집되었다. 일부 문항은 복잡성, 오류 발생 빈도 등을 평가하는 역문항으로 구성하여 사용자가 부정적 경험을 표현할 수 있도록 하였다. 본 연구에서는 표 1과 같이 원문 SUS 설문 문항을 VR 시뮬레이션 맥락에 맞게 수정하였다.

표 1. SUS 설문 문항

Table 1. SUS questionnaire items

item	content	remarks
SUS-1	I found it easy to learn and use the system	Usability evaluation
SUS-2	I found the simulation interface to be overly complex. (Reverse item)	UI complexity
SUS-3	I think the necessary features are well integrated	Integration
SUS-4	I felt there were too many unnecessary steps in the simulation process. (Reverse item)	Step efficiency
SUS-5	I was able to perform the desired tasks (e.g., driving, querying) easily	Operability
SUS-6	I felt the system frequently caused errors. (Reverse item)	Stability
SUS-7	I could quickly learn how to operate the system	Ease of learning
SUS-8	I felt the system operated consistently overall	Consistency
SUS-9	I felt a high cognitive load when using the simulation. (Reverse item)	Cognitive load
SUS-10	I would be willing to use this system frequently in the future	Satisfaction

4.3 실험 결과 및 분석

앞서 RAG 미적용 시나리오와 RAG 적용 시나리오를 동일 시나리오로 각각 체험한 후, 25명 참가자의 SUS 설문 응답을 하였다. 표 2는 RAG 미적용 모드, 표 3은 RAG 적용 모드에서 각 항목별 평균 점수와 표준편차 결과이다. SUS 10개의 문항에서 SUS-1, SUS-3, SUS-5, SUS-7, SUS-8, SUS-10 문항은 일반 문항이며, SUS-2, SUS-4, SUS-6, SUS-9 문항은 역문항이다.

표 2. RAG 미적용 모드 SUS 설문 문항 결과

Table 2. SUS questionnaire results without RAG mode

item	score	standard deviation	item	score	standard deviation
SUS-1	3.8	0.4	SUS-6	3.0	0.4
SUS-2	2.9	0.3	SUS-7	3.7	0.5
SUS-3	3.9	0.5	SUS-8	4.0	0.4
SUS-4	2.7	0.4	SUS-9	2.9	0.3
SUS-5	3.9	0.3	SUS-10	3.8	0.4

표 2에 따르면, RAG 미적용 모드 결과에서는 SUS-1(학습 난이도) 점수가 3.8(±0.4)로 비교적 긍정적인 평가를 받았으며, 이는 사용자가 VR 시뮬레이션을 배우는 데 큰 어려움을 느끼지 않았음을 나타낸다. SUS-2(복잡성, 역문항)는 2.9(±0.3)로, 인터페이스가 과도하게 복잡하지 않다는 평가를 보여준다. 또한 SUS-3(기능 통합성)과 SUS-5(작업 수행 용이성)는 각각 3.9(±0.5)와 3.9(±0.3)로 나타나, 기능 조작과 관련된 사용성 측면에서 긍정적 평가를 받았다. 반면, SUS-6(오류 발생 빈도, 역문항)은 3.0(±0.4)으로 중립적인 평가를 기록하였으며, 이는 일부 사용자가 시스템의 안정성에 대해 의문을 제기했음을 시사한다.

표 3의 RAG 적용 모드에서는 대부분의 항목에서 점수가 상승하였다. SUS-1(학습 난이도)은 4.0(±0.3)으로 증가하였고, SUS-2(복잡성, 역문항)는 2.7(±0.4)로 감소하여 시스템이 더 단순하게 느껴졌다는 점을 보여준다. SUS-3(기능 통합성)은 4.1(±0.47)로 향상되었으며, SUS-5(작업 수행 용이성) 역시 4.1(±0.5)로 나타나 RAG를 활용한 정보 제공이 사용자 의사결정을 지원했음을 시사한다. 또한 SUS-8(시스템 일관성)은 4.2(±0.45)로 미적용 모드 대비 상승하였으며, 시스템에 대한 사용자 신뢰도가 전반적으로 향상되었음을 보여준다.

표 3. RAG 적용 모드 SUS 설문 문항 결과

Table 3. SUS questionnaire results with RAG mode

item	score	standard deviation	item	score	standard deviation
SUS-1	4.0	0.3	SUS-6	2.9	0.61
SUS-2	2.7	0.4	SUS-7	3.9	0.48
SUS-3	4.1	0.47	SUS-8	4.2	0.45
SUS-4	2.8	0.55	SUS-9	2.8	0.59
SUS-5	4.1	0.5	SUS-10	4.0	0.51

일반 문항의 평균 점수는 RAG 미적용 모드에서 3.5, 적용 모드에서는 4.0으로 나타나 RAG의 적용이 사용자 경험에 긍정적인 영향을 미쳤음을 알 수 있다. 역문항 점수는 적용 모드에서 더 낮게 나타나, 불필요한 복잡성과 오류 빈도가 감소한 것으로 분석된다.

다만 SUS-6(오류 발생 빈도, 역문항)의 경우, 미 적용 모드 3.0(± 0.4)에서 적용 모드 2.9(± 0.61)로 큰 차이가 없었는데, 이는 DB 검색 지연 또는 STT/TTS 관련 성능의 한계에서 기인한 것으로 보인다. 응답 의향을 묻는 SUS-10(향후 사용 의향) 점수는 미적용 모드 3.8(± 0.4)에서 적용 모드 4.0(± 0.51)로 상승하였으며, 이는 RAG 적용과 VR 시뮬레이션의 결합에 대해 사용자들이 긍정적인 반응을 보였음을 나타낸다.

실험 결과는 RAG 기법이 VR 시뮬레이션 사용자 경험에 긍정적인 영향을 미친다는 점을 정량적으로 입증하였다. 특히 학습 난이도, 조작성, 기능 통합성, 일관성 등 주요 항목에서 RAG 적용 모드가 미적용 모드에 비해 더 높은 점수를 기록하였다. 이는 RAG가 신뢰도 높은 정보를 제공함으로써 사용자가 VR 환경에서의 의사결정을 보다 효과적으로 수행할 수 있게 했음을 보여준다. 다만 일부 역문항에서 표준편차가 높게 나타난 점은 사용자 간 경험의 편차가 존재함을 시사하며, 이는 네트워크 상태, 음성 인식 정확도 등의 요인에 따라 달라졌을 가능성을 반영한다. 따라서 시스템 성능 최적화와 사용자 맞춤형 설정 제공을 통해 이러한 편차를 줄이는 방안이 필요하다.

결론적으로, 본 연구는 RAG 기법이 VR 시뮬레이션의 신뢰성과 사용성을 향상시키는 데 기여할 수 있음을 보여주었으며, 향후 군사 훈련과 같은 실질적 응용에서 RAG 기법의 잠재적 활용 가능성을 입증하였다.

추가로, RAG 적용 시스템의 응답 시간을 정량적으로 측정하였다. 평균적으로 RAG 미적용 시 약 1.6초, 적용 시 약 2.2초로 측정되어, 검색 및 스니펫 생성 과정으로 인해 약 0.6초가 추가 소요되었다. 이는 VR 훈련 시뮬레이션에서 큰 지연은 아니었지만, 문서량이 많아지거나 네트워크 상태가 열악해지면 지연이 가중될 수 있으므로, 효율적인 인덱싱 및 캐싱 전략을 고려할 필요가 있다.

V. 결 론

본 연구에서는 VR 시뮬레이션 환경에서 GPT 계열 대규모 언어 모델을 적용하여, 지상 무인체 운용 훈련에 요구되는 정확도와 신뢰도 높은 전술 조언을

제공하는 방안을 제시하였다. 이를 위하여 RAG 기법을 결합하고, 음성 인식과 TTS 기술을 통해 대화형 어시스턴트를 구축함으로써 사용자의 몰입도를 높였다. 연구 결과, RAG를 적용하지 않은 시뮬레이션에서는 GPT 4o 모델이 존재하지 않는 무기체계나 지형 정보를 사실처럼 생성해내는 ‘환각’ 사례가 발견되었으며, 이는 군사훈련 상황에서 훈련자의 잘못된 의사결정을 유발할 수 있음을 재확인한 것이다. 반면, 문서 스니펫을 추출해 모델에 제공하는 RAG 기법을 도입하자 환각 발생률이 유의미하게 줄었고, 사용자 설문에서도 더 높은 점수가 확인되었다. 이는 VR 시뮬레이션에 RAG를 결합함으로써 LLM 환각 문제를 상당 부분 억제하면서도 원활한 훈련 경험을 유지할 수 있음을 시사한다. 또한, Whisper 음성 인식과 TTS를 활용해 운전 중 손을 떼지 않고도 AI 어시스턴트의 전술 조언을 청취할 수 있도록 구현함으로써, 군사훈련 시나리오의 실전감이 한층 높아졌다. 한편, RAG 적용 시 평균 0.6초가량 추가 응답 지연이 발생하였으나, 실험 참가자들은 이를 수용 가능한 수준으로 평가하였고, 오히려 문서 기반 정보 제공의 정확성을 더 가치 있게 여기는 경향을 보였다. 그럼에도 본 연구에는 몇 가지 한계가 존재한다. 먼저, 25명의 비군사 전문가를 대상으로 실험을 수행하여, 전문가 시각에서의 전술적 정확도나 실전 적용성을 충분히 평가하지 못했다. 또한, DB에 오픈된 문헌 위주 자료만 사용하여 군사기관의 기밀 자료나 최신 전장 상황 정보가 반영되지 못한 점도 제약으로 남는다. 나아가 정찰 및 대응 두 가지 시나리오만 다루어 복합 상황을 충분히 포괄하지 못했다는 점 역시 한계로 지적할 수 있다.

따라서 향후 연구에서는 군사 전문가와의 협력을 통해 실제 전술 매뉴얼을 확보하고, 다양한 시나리오를 VR 시뮬레이션에 통합하여 GPT 어시스턴트의 적용 가능성을 더욱 정밀하게 검증할 필요가 있다. 또한, 소음 환경이 극심하거나 다양한 억양 언어가 혼재된 실제 훈련 상황에서도 음성 인식 및 합성 품질을 안정적으로 유지하도록 STT 및 TTS 기술 최적화가 진행되어야 한다. 마지막으로, 대규모 문서와 열악한 네트워크 환경에서 지연이 가중될 수 있으므로, 인덱싱 최적화와 캐싱 전략에 관한 후속 연구 역시 필수적이다.

결론적으로, 본 연구는 VR 시뮬레이션에 GPT 기반 대화형 어시스턴트를 도입하고, RAG 기법을 통해 환각 문제를 억제하면서도 고도의 전문 정보를 제공할 수 있는 시스템을 구축·검증했다는 점에서 의의를 지닌다. VR과 AI 융합을 통한 군사훈련 시스템의 발전 가능성을 확인함으로써, 추후 군사 전문가 대상의 정밀 평가와 실전 환경 모의 시나리오를 활용한 다각적인 성능 검증이 이루어지길 기대한다.

References

- [1] B. C. Arjun and K. R. Sanjay, "Enhancing Military Training Through VR Applications", *International Scientific Journal of Engineering and Management*, Vol. 3, No. 5, May 2024. <https://doi.org/10.55041/ISJEM01739>.
- [2] D. J. Harris, T. Arthur, J. Kearse, M. Olonilua, E. K. Hassan, T. C. D. Burgh, M. R. Wilson, and S. J. Vine, "Exploring the role of virtual reality in military decision training", *Frontiers in Virtual Reality*, Vol. 4, Mar. 2023. <https://doi.org/10.3389/frvir.2023.1165030>
- [3] C. A. Andersson, et al., "Effectiveness of an Expendable Unmanned Ground Vehicle Stalling a Mechanized Infantry Company's Primary Combat Units—A Virtual Simulation Experiment", *Journal of Field Robotics*, pp. 1-18. Sep. 2024. <https://doi.org/10.1002/rob.22442>.
- [4] F. Sufi, "An innovative GPT-based open-source intelligence using historical cyber incident reports", *Natural Language Processing Journal*, Vol. 7, pp. 100074, Jun. 2024. <https://doi.org/10.1016/j.nlp.2024.100074>.
- [5] Y. Zhang, P. Shao, Y. Sang, Pa. Wang, Pe. Wang and S. Niu, "A Strategy for Enhancing Modal Fusion Based on LLaVA towards Military Field", 2024 6th International Conference on Electronic Engineering and Informatics (EEI), Chongqing, China, pp. 1482-1487, Jun. 2024. <https://doi.org/10.1109/EEI63073.2024.10696382>.
- [6] D. H. Park and H. J. Lee, "Literature Review of AI Hallucination Research Since the Advent of ChatGPT: Focusing on Papers from arXiv", *Informatization Policy*, Vol. 31, No. 2, PP. 3-38, Jun. 2024. <https://doi.org/10.22693/NIAIP.2024.31.2.003>.
- [7] J. H. Shin, S. R. Park, H. R. Kim, and J. H. Lee, "Search-based Generation Techniques for Enhancing LLM Responses: A Comparative Study of GPT3.5 and GPT4 in Zero-shot and RAG", *Proc. of the Korea Information and Communication Society General Conference*, Anseong, Korea, Vol. 27, No. 2, pp. 350-352, Oct. 2023.
- [8] M. H. Pa, S. S. Lee, K. S. Jeon, and H. J. Seol, "A Study on the Development Direction of Education and Training System based on AR/VR Technology", *Journal of the KIMST*, Vol. 22, No. 4, pp. 545-554. Aug. 2019. <https://doi.org/10.9766/KIMST.2019.22.4.545>.
- [9] L. Xue, L. Jie, Z. Peipei, and X. Tao, "MilChat: A Large Language Model and Application for Military Equipment", 2024 7th International Conference on Machine Learning and Natural Language Processing (MLNLP), Chengdu, China, pp. 1-5, Oct. 2024. <https://doi.org/10.1109/MLNLP.63328.2024.10800008>.
- [10] P. Lewis, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", *Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver BC Canada, Vol. 33, pp. 9459-9474, Dec. 2020.
- [11] V. Karpukhin, et al., "Dense Passage Retrieval for Open-Domain Question Answering", *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 6769-6781, Nov. 2020. <https://doi.org/10.18653/v1/2020.emnlp-main.550>.
- [12] L. J. Jose, G. Z. Lucia, E. M. Maria, A. Mariano, and M. M. Javier, "Developing conversational

Virtual Humans for social emotion elicitation based on large language models", Expert Systems with Applications, Vol. 246, pp. 123261, Jul. 2024. <https://doi.org/10.1016/j.eswa.2024.123261>.

[13] D. D. Thinh, N. C. Thanh, L. V. Diep, N. D. Vuong, L. D. Ha, and P. N. Hai, "A voice search engine for military symbols to enhance the drafting of operational plan documents on digital map", Journal of Military Science and Technology, Vol. 87, pp. 40-49, May 2023. <https://doi.org/10.54939/1859-1043.j.mst.87.2023.40-49>.

[14] C. W. Yoon, S. G. Yang, J. S. Park, J. W. Si, Y. C. Jung, and S. Y. Kim, "Metaverse Virtual Interview Platform Leveraging Generative AI and Speech Recognition", Journal of KIIT, Vol. 22, No. 6, pp. 163-173, Jun. 2024. <http://doi.org/10.14801/jkiit.2024.22.6.163>.

저자소개

우 동 현 (Donghyun Woo)



2012년 2월 : 배재대학교
전자공학과(공학사)
2018년 8월 : 중국 동북전력대학교
정보및통신공학(공학석사)
2021년 3월 ~ 현재 : 창원대학교
첨단방위공학과정 박사과정
관심분야 : 컴퓨터비전,
증강/가상현실

유 선 진 (Sunjin Yu)



2003년 8월 : 고려대학교
전자정보공학(공학사)
2006년 2월 : 연세대학교
생체인식공학(공학석사)
2011년 2월 : 연세대학교
전기전자공학(공학박사)
2011년 1월 ~ 2012년 5월 :
LG전자기술원 미래IT융합연구소 선임연구원
2012년 5월 ~ 2013년 2월 : 연세대학교 전기전자공학과
연구교수
2013년 3월 ~ 2016년 8월 : 제주한라대학교 방송영상학과
조교수
2016년 9월 ~ 2019년 8월 : 동명대학교 디지털미디어공학부
부교수
2019년 9월 ~ 2024년 9월 : 국립창원대학교
문화테크노학과 부교수
2024년 10월 ~ 현재 : 국립창원대학교 문화테크노학과
정교수
관심분야 : 컴퓨터비전, 증강/가상현실, HCI