

<b>과 제 명</b>	<b>국방출판지원단 보유 군 지식정보 전 군 검색 활용을 위한 챗봇 개발</b> * 개발된 챗봇은 출판단에서 실증 후 전 군 확산사업(국방지능형플랫폼 탑재) 추진		
<b>부 대 명</b>	국방출판지원단	<b>담당자 성함</b>	
<b>연 락 처</b>		<b>이 메 일</b>	
<b>예산과제예산</b>	1억 원	<b>예산과제기간</b>	6개월

### 1.과제개요

- 국방출판지원단의 방대한 자료를 쉽게 검색하고 답변 받을 수 있는 챗봇 개발
- 한국어를 이해하고 한국어로 추론할 수 있는 한국형 LLM을 사용
- 국방출판지원단의 방대한 자료를 벡터 DB에 저장하고 RAG를 통하여 검색하고 추론 및 답변은 LLM에서 하는 방식으로 환각없는 챗봇 개발 및 실증

### 2.과제개발배경

- 국방출판지원단은 전 군 출판콘텐츠 지원하는 군 책임운영 기관으로 육군 교육사령부 예하 15개 학교 등 전 군 소요부대에서 연 9,000여건의 출판원고 도입 및 출판. 출판 후 원고 데이터(PDF) DB화 하여 관리 중
- 군 독립망 사용 및 보안의 이유로 인터넷시서비스를 사용할 수 없어 국방망 내 군 자료기반 시챗봇 필요
- 60만 군에 시기술 확산을 위한 다양한 국방 시서비스 사업을 추진 중이며 이를 위해 전 군 지식정보데이터(무기3D모델 등 비정형데이터)를 수집 중

#### • 출판데이터 보유현황

-국방출판지원단에서 인쇄된 국방 전문야 출판데이터 최종본 최근 5개년 누적 47,730개 보유

교재	시사 선전	일반 교범	기술 교범	일반 책자	일반 서식	기타	총 개수
							<b>47,730</b>
22,125	3,918	1,232	1,360	14,348	2,210	2,537	

#### • 문제점

- 군의 방대한 지식정보 데이터에서 원하는 결과를 챗봇으로 쉽게 얻을 수 없음
- 군 특성상 보안을 요하여 ChatGPT와 같은 외부 LLM을 이용할 수 없음, 군 자체 챗봇 구축 필요
- 국방분야 지식정보 데이터의 특성상 챗봇의 환각을 줄일 수 있는 RAG가 필요
- 새로운 데이터가 들어올 때 마다 파인튜닝을 하는 것은 비용이 많이 들어 RAG를 통하여 빠르게 데이터를 최신화할 필요가 있음

### 3.기대효과

#### • 기대효과 및 성과

- AI과학기술강군 국방정책 이행 및 AI기술 군 도입 인식 확산
- 국방망 챗봇 도입으로 군 지식정보DB 검색을 활용한 콘텐츠 생성 실증
- 사용되지 못한 출판데이터 內 군 지식정보(비정형데이터) 활용 기반 마련
- 업무 혹은 교육자료 제작 시 민간과 유사한 AI 활용한 40~50% 시간 단축 기회 제공

#### • 제작 콘텐츠 활용 계획 및 기대효과

- 국방망에서 활용 가능토록 군 정보체계 요건을 만족하는 챗봇 개발
- 군 지식정보 출판데이터 제공 및 벡터DB 구축 타 군 활용(각군 교육사, 국방 AI센터 등)
- 군 교육기관 사용자의 교라.교범.교재 제작 시 AI챗봇을 통한 아티클 생성 활용



< 군 디지털교과서(예) >

- 연간 육해공군 및 국방부국직기관 소요 과목수 약 3,300여 건(추정)

구 분	계	육군	해군	공군	국방부/국직
계(건)	3,289	1,569	1,025	378	317
일반/기술교범(건)	298	73	69	35	121
공통교재(건)	66	34	27	-	5
일반교재(건)	2,925	1,462	929	343	191

< 전 군 교육자료 소요 (예) >

- 육해공군 교육사령부 원격교육 총괄부서 챗봇 활용 시범사용 합의( ID 400개 분배 )
- 육군 교육사령부의 경우 예하 15개교(보병학교, 기계교 등)에 ID 10~20개 교관 발급 (교관 1과목 당 교육생 연 50~300명)
- 군 표준 시디지털교과서 플랫폼 사업 추진 시 챗봇 및 벡터DB 연계

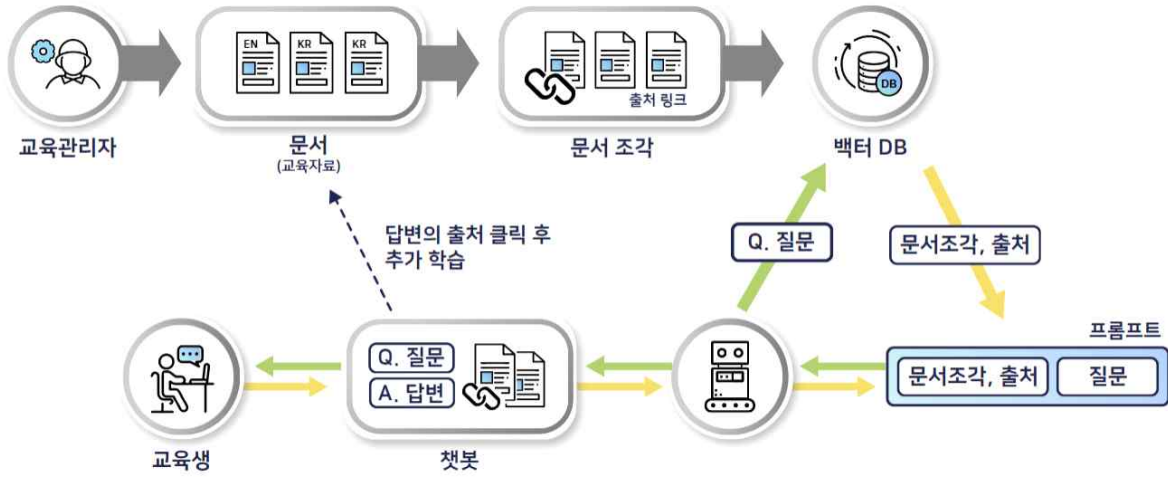
#### • 기타활용

- 공개가능한 벡터데이터의 경우 국방부 개방관리시스템을 통한 공공데이터 개방
- 군 출판데이터로 AI확습된 군 일반자료 민간 배포를 통한 군↔민 데이터 활용

#### 4.국 내 외 기술개발 현황

- ChatGPT와 같은 대형언어모델(LLM)이 이미 매우 성숙되었고 유익한 수준에 도달했음
- LLM 모델만으로는 환각현상이나 외부 최신 데이터를 사용하지 못하는 한계점이 있어 검색증강생성(RAG)가 각광받고 있음
- RAG를 위한 구성요소의 핵심인 Vector DB도 Chroma같은 단순한 구조에서 대용량 처리를 위해서 GPU를 사용할 수 있고 병렬로도 사용할 수 있는 Milvus 같은 Vector DB가 주목 받고 있음
- 또한 Vector DB 검색도 벡터의 유사도에로만 검색하였을 때 정확도가 떨어져, 벡터 유사도와 Full Text Search 등을 같이 수행하는 Hybrid Search가 대두되고 있음
- 다양한 검색 방식을 조합한 Hybrid Search를 할 경우 각 검색 방식에서 검색된 데이터의 Ranking을 다시 조정하는 Rerank 전략이 검색 결과의 품질을 높일 수 있어 Rerank 전략을 포함한 RAG가 최신 흐름으로 자리 잡고 있음
- RAG는 Naive RAG, Advanced RAG, Modular RAG등 이 나오고 있으며, Naive RAG는 단순 벡터 유사도 검색이며, AdvancedRAG는 Hybrid Search, Rerank 전략을 포함하여 검색 품질을 향상시키는 것임.
- Modular RAG는 유연성과 적응성을 높이기 위해 다양한 모듈을 추가하는 등의 RAG를 의미하며, 초기 연구 단계임
- 벡터 검색도 100% 정확하지 않아 새롭게 대두되고 있는 것이 GraphRAG임
- GraphRAG는 데이터를 벡터로 인벤딩한 것이 아니 각 단어별로 나누어 각 단어 간에 그래프로 연결하여 좀 더 정확한 검색이 이루어 질 수 있도록 함
- GraphRAG와 RAG의 실질적 성능(검색의 수준, 속도) 등은 다양한 성능 기준으로 측정을 해 봐야 함
- 군에서 사용하는 특성상 RAG와 GraphRAG의 접근제어 수준을 실험하여 더 좋은 방식을 선택해야 함
- RAG의 성능을 측정하는 것이 주관적이었으나, 최근에 RAG의 성능을 측정하는 프레임워크가 나오고 있음
- RAG 성능 측정 도구를 활용하여 객관적인 성능 측정이 필요

## 5.과제목표



< AI챗봇 군 교육 활용 시 개념도 >

### • 과제목표

- 군에 적합한 환각이 없고 최신 데이터에 기반한 답변을 하여 교재, 교보재, 질의 응답이 가능한 RAG 기반의 챗봇 개발

주요성능 (핵심가치)	개발목표
검색의 정확도	벡터 DB 또는 Graph DB의 검색 정확도 측정, 정확도 90%이상
LLM의 답변 정확도	벡터 DB 또는 Graph DB에서 검색한 데이터에 기반하여 정확한 답변을 하는 LLM 답변의 정확도 측정, 95% 이상
보안성 평가	챗봇의 접근 제어 성능, 100% 목표
응답 시간	대용량 데이터에 대한 응답 속도, 5분 이내,

- 검색의 정확도 세부 시나리오
  - 원본 데이터에서 선별한 20개의 질문과 답변을 선택
  - 질문을 하였을 때 검색된 데이터가 질문과 답변이 있는 데이터 인지 확인
- LLM 답변의 정확도 세부 시나리오
  - 원본 데이터에서 선별한 20개의 질문과 답변을 선택
  - 정확한 데이터가 주어졌을 때 정확한 답변이 나오는지 확인
- 보안성 평가 세부 시나리오
  - 보안 등급이 다른 20개의 데이터를 선별
  - 보안 등급이 다른 인원이 접근하였을 때 접근 가능한 데이터가 검색되는지 확인
- 응답 시간 세부 시나리오
  - RTX 4090 이상의 GPU를 가진 PC에서 RAG 시스템을 설치하여 질문 20개를 준비하고 PC에서 인원이 20번 질문을 하고 답변이 나오는데까지 걸리는 시간을 측정