

과업지시서

용역명	임상시험 통합플랫폼 전용 음성·영상 생성형 인공지능 Private LLM 고도화
주관기관	연세대학교 미래의료산학협력단

2026. 5. 20.

과제	의료기기 임상전문가 양성사업			
담당	의료기기CRO센터	과업 관련	tel. 033-741-0442	iays94@naver.com
	구매 관재팀	입찰제도 관련	tel. 033-741-0428	syj@yonsei.ac.kr



연세대학교
미래의료산학협력단

목 차

I. 용역 개요	3
1. 과업명	3
2. 과업기간	3
3. 계약금액	3
4. 입찰방식	3
II. 과업의 목적 및 필요성	3
1. 과업의 목적	3
2. 필요성	4
3. 기대효과	4
III. 과업 추진방안	5
1. 추진목표	5
2. 기능 고도화 개발 내용	5
3. 추진전략	7
4. 추진체계	8
5. 추진일정	10
IV. 과업 내용	11
1. 과업개요	11
2. 과업내용	11
3. 과업범위	13
4. 세부 요구사항	18
5. 성과품 목록	32

I 용역 개요

1. 과업명 : 임상시험 통합플랫폼 전용 음성·영상 생성형 인공지능 Private LLM 고도화
2. 과업기간 : 착수일(계약일)로부터 4개월
3. 계약금액 : 금 80,000,000원(금 팔천만원 정)(VAT 포함)
4. 입찰방식
 - 가. 계약방법 : 제한경쟁입찰(협상에 의한 계약)
 - 나. 입찰참가 조건 및 자격 : 제안요청서 참고
 - 다. 낙찰자선정방법 : 제안요청서 참고

II 과업의 목적 및 필요성

1. 과업의 목적

- 가. 기존 임상시험 통합플랫폼 전용 Private LLM을 확장하여, 의료기기 절차·설명 자료를 기반으로 음성 및 영상 콘텐츠를 자동 생성할 수 있는 Text-to-Audio, Text-to-Video, Image-to-Video 기능을 고도화하기 위함
- 나. 텍스트와 이미지만으로는 충분히 전달하기 어려웠던 의료기기 사용법, 절차, 주의사항 등을 음성 및 영상 형태로 제공함으로써 교육 콘텐츠의 이해도, 품질, 생산성을 동시에 향상시키고자 함
- 다. Text-to-Audio, Text-to-Video, 립싱크 기반 Image-to-Video 기술을 Private LLM 워크플로우에 통합하여, 임상시험 종사자가 텍스트·이미지·음성·영상을 통합적으로 활용할 수 있는 멀티모달 서비스 환경을 구축하고자 함
- 라. 본 과업의 영상 생성 기능은 기술 안정성 및 결과 일관성을 고려하여 장편 영상이 아닌 Short-duration 영상 중심으로 구현하며, Text-to-Video는 5초 내외, 립싱크 기반 Image-to-Video는 10초 내외를 기준 단위로 구현한다

2. 필요성

- 가. 임상시험 및 의료기기 교육 현장에서 절차 중심 시각자료의 중요성이 증가하고 있으나, 기존 촬영·편집 방식은 높은 비용과 긴 제작 시간이 소요되며, 내용 변경 시 재제작 부담이 큰 한계가 있음
- 나. 최근 Text-to-Audio, Text-to-Video, Image-to-Video 등 생성형 AI 기술의 발전으로 짧은 길이의 교육용 콘텐츠 자동 생성이 가능해졌으며, 특히 Short-duration 영상은 현재 기술 수준과 운영 환경을 고려할 때 가장 현실적인 적용 방식임
- 다. 의료기기 사용적합성 예방과 절차 준수 향상을 위해 설명 스크립트, 내레이션, 시나리오 영상 등 멀티모달 교육자료의 필요성이 증가하고 있음
- 라. 기존 Private LLM의 아키텍처, 보안 체계, 사용자 인증 체계, 클라우드 운영 환경을 재활용하면서 음성·영상 기능만 단계적으로 확장하는 방식이 예산, 기간, 기술 연속성 측면에서 가장 효율적임

3. 기대효과

- 가. 음성 및 영상 자동 생성으로 교육자료 제작 시간과 비용 절감
- 나. 절차 이해도 향상 및 임상시험 현장 사용자 오류 감소
- 다. 제출자료, 교육자료, 설명자료 등 시각 기반 콘텐츠 품질 표준화
- 라. 기존 Private LLM을 텍스트 질의응답 중심 서비스에서 음성·영상까지 포함하는 최신 멀티모달 서비스로 확장
- 마. 의료기기 임상시험 전문인력의 역량 강화 및 업무 효율 향상

1. 추진목표

의료기기 임상시험 전문인력의 효율적 지원을 위한
음성·영상 생성형 멀티모달 Private LLM 고도화

기존 텍스트 및 이미지 생성 기반 Private LLM을 확장하여, 의료기기 절차 및 설명 자료를 기반으로 교육용 음성 및 Short-duration 영상을 자동 생성하는 Text-to-Audio, Text-to-Video, 립싱크 기반 Image-to-Video 기능을 경량 고도화함. 이를 통해 임상시험 교육 콘텐츠의 품질 향상과 생산성 증대를 실현하고, 임상시험 종사자의 절차 이해도와 업무 효율을 높이는 통합 서비스 환경을 구축함.

2. 기능 고도화 개발 내용

- 텍스트 기반 음성 생성(Text-to-Audio) 기능 개발
 - Private LLM이 생성한 절차 설명, 교육 스크립트, 안전지침 등을 자연스러운 음성으로 자동 변환하는 Text-to-Audio 모듈 구축
 - 생성된 음성을 기존 대화형 서비스 흐름과 연계하여 재생 및 다운로드 가능하도록 구성
 - 남성형, 여성형 대표 음성을 기본 제공
- 텍스트 기반 영상 생성(Text-to-Video) 기능 개발
 - 사용자의 텍스트 요청 또는 Private LLM이 생성한 절차 설명 스크립트를 기반으로 교육용 Short-duration 영상을 생성하는 기능 구현
 - 텍스트 기반 생성형 영상(Text-to-Video)은 기술적 안정성과 일관성을 고려하여 1회 생성 단위를 5초 내외로 구현
 - 장편(Long-duration video) 영상 생성은 과업 범위에서 제외하며, 필요 시 생성된 Short-duration 영상을 외부 편집 툴로 결합하여 활용 가능하도록 함
- 립싱크 기반 영상 생성(Image-to-Video) 기능 개발

- 이미지 입력 또는 시스템에서 생성한 가상 설명 인물 이미지를 기반으로, 음성과 입 모양을 동기화한 립싱크 영상을 생성하는 Image-to-Video 기능 구현
- 립싱크 기반 영상은 설명 전달력을 고려하여 1회 생성 단위를 10초 내외로 구현
- Image + Audio + Lip-sync Video 형태로 최종 결과물을 생성함

○ 기존 Private LLM과의 연계 · 통합

- 생성된 Text와 이미지 입력을 기반으로 음성과 영상 생성 연계 기능을 개발 적용
- 기존에 구현된 대화형 화면을 통해서 연계 처리

○ 임상시험 통합플랫폼 전용 기존 멀티모달 Private LLM과의 연계 · 통합

- 텍스트 질의응답(Text-to-Text), 이미지 생성(Text-to-Image), 음성 생성(Text-to-Audio), 영상 생성(Text-to-Video, Image-to-Video)을 하나의 서비스 흐름으로 통합
- 오픈소스 기반 음성 및 영상 생성 멀티모달 모델과 기존 Private LLM을 연계하여 일원화된 서비스 구조를 구현
- 모든 요청 · 응답 이력은 통합 대화형 UI 및 관리자 기능과 연계하여 관리함

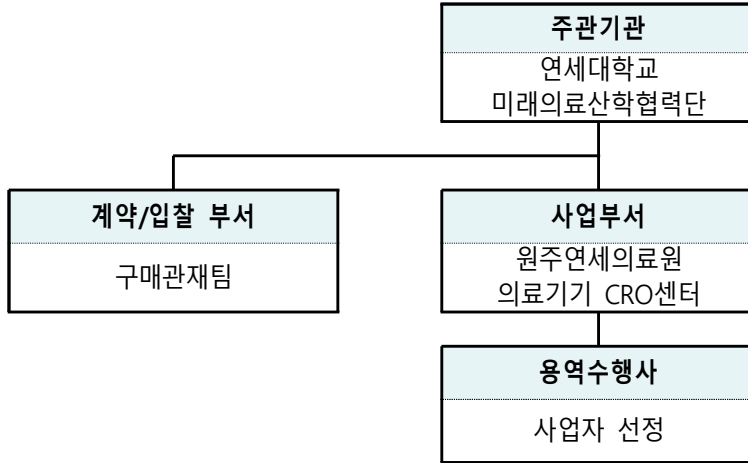
3. 추진전략

- 기존 Private LLM 아키텍처와 운영 환경을 유지하면서 확장
- Private LLM에서 생성된 설명자료를 Text-to-Audio로 연계
- Private LLM에서 생성된 텍스트 및 이미지 자료를 Text-to-Video, Image-to-Video로 연계
- 영상 생성 기능은 기술적 안정성을 고려하여 Short-duration 영상을 중심으로 적용
- Text-to-Video는 5초 내외, 립싱크 기반 Image-to-Video는 10초 내외 기준으로 최적화
- Long-duration 영상은 현 단계 과업 범위에서 제외하고, Short-duration video 생성 결과의 활용성을 우선 확보

4. 추진체계

가. 사업 추진체계

○ 추진조직



○ 조직별 역할

조 직	역 할
사업부서	<ul style="list-style-type: none"> ○ 사업 총괄관리 ○ 유관기관 협력 ○ 이슈사항 의사결정 ○ 사업계획 수립 ○ 제안요청서 작성 및 발주 요청 ○ 사업자 선정 및 계약 체결 ○ 요구사항 및 의견 개선 ○ 사업 수행관리 및 결과물 검사·검수
계약/입찰 부서	<ul style="list-style-type: none"> ○ 계약 추진(공고, 개찰, 계약 등) ○ 사업비 지급
용역 수행사	<ul style="list-style-type: none"> ○ 사업 수행 및 사업관리, 보고 수행 ○ 사업결과물 납품

5. 추진일정

구분	1M	2M	3M	4M
○ 기본 요건 분석/정의				
○ 착수 보고	★			
○ Text-to-Audio 모듈 구축				
○ Text-to-Video 기능 구현				
○ 립싱크 영상 기능 구현 - 이미지·대본·음성·Image-to-Video 접목				
○ 고도화 멀티모달 통합 모델 최적화				
○ 통합테스트 및 보완				
○ 오픈 및 안정화				
○ 검수/완료 보고(오픈 후 별도 일정)				★

※ 상기 추진 일정은 주관기관과 용역수행사의 상호협의를 통해 변경될 수 있음

※ 용역 계약 후, 7일 이내 개발 내용 및 계획에 대한 착수 보고

※ 성과물 납품 후, 한 달 이내 결과보고회 개최(개최 일정은 발주사와 협의 가능)

1. 과업개요

- 본 사업은 기존 임상시험 통합플랫폼 전용 Private LLM (Text-to-Text 기반)을 중심으로, 음성 및 영상 생성 기능을 확장하여 의료기기 임상시험 관련 절차와 설명 자료를 멀티모달 방식으로 제공하기 위한 사업임
- 본 사업의 핵심은 기존 Private LLM 시스템의 데이터, 보안, 인증, 서비스 체계를 그대로 유지하면서 Text-to-Audio, Text-to-Video, 립싱크 기반 Image-to-Video 기능을 경량 확장하는 데 있음

2. 과업내용

가. 멀티모달 기반 클라우드 플랫폼 서비스 제공

- 현재 활용 중인 Private LLM 클라우드 서비스 확장 및 연장
- 기존 Text-to-Text, Text-to-Image 시스템을 기반으로 음성 및 영상 기능을 확장한 인공지능 기반 대화형 서비스를 제공
- 교육 및 설명 스크립트 자동 생성(Text-to-Text) : 기존 기능 연계
- 교육 및 설명 용도 음성 생성(Text-to-Audio) : 신규 기능
- 교육 및 설명 용도 영상 생성(Text-to-Video) : 신규 기능, 5초 내외
- 설명 인물 이미지 기반 립싱크 영상 생성(Image-to-Video) : 신규 기능, 10초 내외
- 사용자가 텍스트, 이미지, 음성, 영상을 하나의 대화형 화면에서 자연스럽게 요청·열람할 수 있도록 구성

나. Private LLM 시스템 내 음성 생성, 동영상 생성 기능 통합 제공

- 기 구축된 Private LLM 시스템 내에서 텍스트 질의응답과 음성·영상 생성 기능이 통합된 일원화된 사용자 인터페이스(UI) 제공
- 멀티모달 요청(Text, Image, Audio, Video)과 기존 텍스트 질의응답을 하나의 통합된 대화형 화면에서 지원
- 사용자가 요청한 모든 텍스트/이미지/음성/영상 요청 및 응답 내역

(히스토리)을 통합 관리 및 열람 가능하도록 구현

- 대화 기반으로 반복 요청, 후속 질의, 수정 요청이 가능해야 함
- 사용자는 기존과 동일한 로그인 계정 체계를 이용해 접근 가능하며, 별도 회원가입은 불필요함
- 향후 필요 시 관리자 모드에서 응답 모니터링, 품질 관리 기능 확장이 가능하여야 함

다. 멀티모달 데이터에 대한 정보 보안 강화

- 기 구축된 Private LLM 시스템의 보안 및 데이터 관리 체계를 유지해야 함.
- 멀티모달 요청·응답 데이터(텍스트, 이미지, 음성, 영상 포함)는 모든 과정을 기존 Private LLM 시스템 내부 클라우드 환경에서만 처리해야 함.
- 사용자가 요청한 음성 및 영상 생성 데이터는 기존에 운영 중인 제한된 클라우드 저장소에서만 관리해야 함.
- 금칙어와 같은 민감 단어가 포함된 요청의 경우, 자동 필터링 및 관리자 사전 검토 기능을 통해 정보 노출 위험 차단
- 기존 Private LLM 시스템의 보안 관리 모듈을 확장하여 멀티모달 전용 요청·응답 관리 대시보드에서 관리자 기록, 로그 분석 기능까지 통합 제공

3. 과업범위

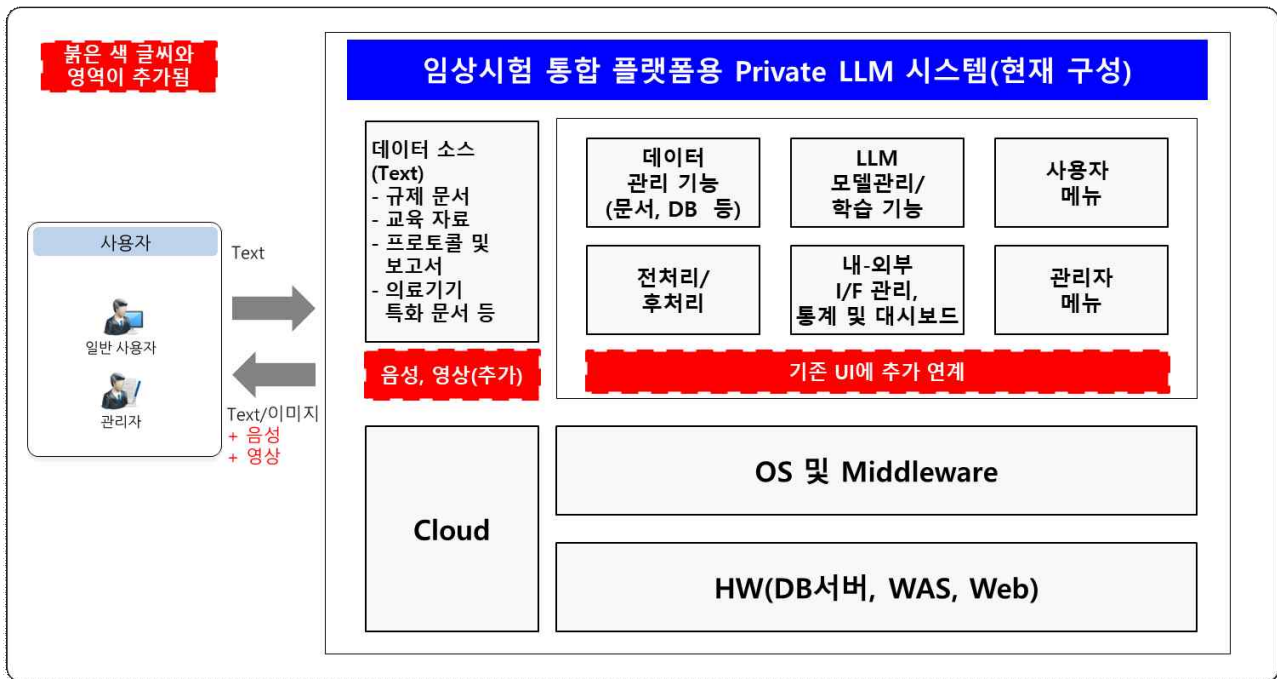
가. 클라우드 환경 기반 멀티모달 영상-음성 생성 통합 환경 구축

- 기존 Private LLM 시스템의 클라우드 환경을 활용하여 Text-to-Audio, Text-to-Video, Image-to-Video 기능을 확장하는 운영 환경 구축
- 멀티모달 서비스가 기존 Private LLM 플랫폼과 완전히 통합되어 하나의 서비스 환경 내에서 동작하도록 설계
- 대규모 병렬 연산이 가능한 GPU 서버 환경을 기반으로 연산 자원 최적화 및 확장 적용
- 기존 Private LLM 시스템의 텍스트 데이터, 보안 체계, 사용자 계정, 대화형 서비스 흐름과 일관된 플랫폼 구조로 멀티모달 확장
- 모든 멀티모달 데이터와 요청/응답 처리는 기존 Private LLM 클라우드 환경에서 수행

나. 음성·영상 멀티모달 기능 고도화

- 기존 텍스트 기반 Private LLM 서비스에 Text-to-Audio, Text-to-Video, Image-to-Video 기능을 통합 확장하여 임상시험 종사자를 위한 영상 기반의 시각적·직관적 정보 제공 서비스를 고도화
- LLM이 생성한 절차 설명 스크립트를 기반으로 교육용 음성 내레이션 자동 생성
- 텍스트 기반 영상 생성(Text-to-Video)은 5초 내외 Short-duration video로 구현
- 생성 인물 또는 입력 이미지 기반 립싱크 영상 생성(Image-to-Video)은 10초 내외로 구현
- 사용자는 이미지 생성, 스크립트 생성, 음성 생성, 최종 영상 생성 요청을 하나의 대화형 창에서 통합 지원받아야 함
- 사용자는 기존 Private LLM 계정을 그대로 활용하여 멀티모달 기능 사용 (별도의 회원가입/계정 필요 없음)
- 모든 텍스트·이미지·음성·영상 요청/응답 이력을 통합 관리하여 사용자가 대화형으로 후속 질의, 반복 요청 가능

다. 시스템 구성도



[음성, 영상 기능 통합 멀티모달 Private LLM 구성도]

라. 멀티모달 활용 서비스 시나리오 구현(Text-to-Audio 예시)

- 텍스트 형태의 의료기기 매뉴얼 및 임상시험 절차를 자연스러운 의료 교육용 음성 콘텐츠로 자동 변환하는 서비스 흐름을 구현함
- 기존의 정적인 텍스트 지침을 '청각적 가이드'로 전환함으로써 교육 대상자의 이해도를 제고하고, 현장 내 사용자 오류(Use Error) 예방을 위한 실질적 활용 기반을 마련함

- [적용 예시] 의료기기 표준 사용법 음성 안내 자동 생성

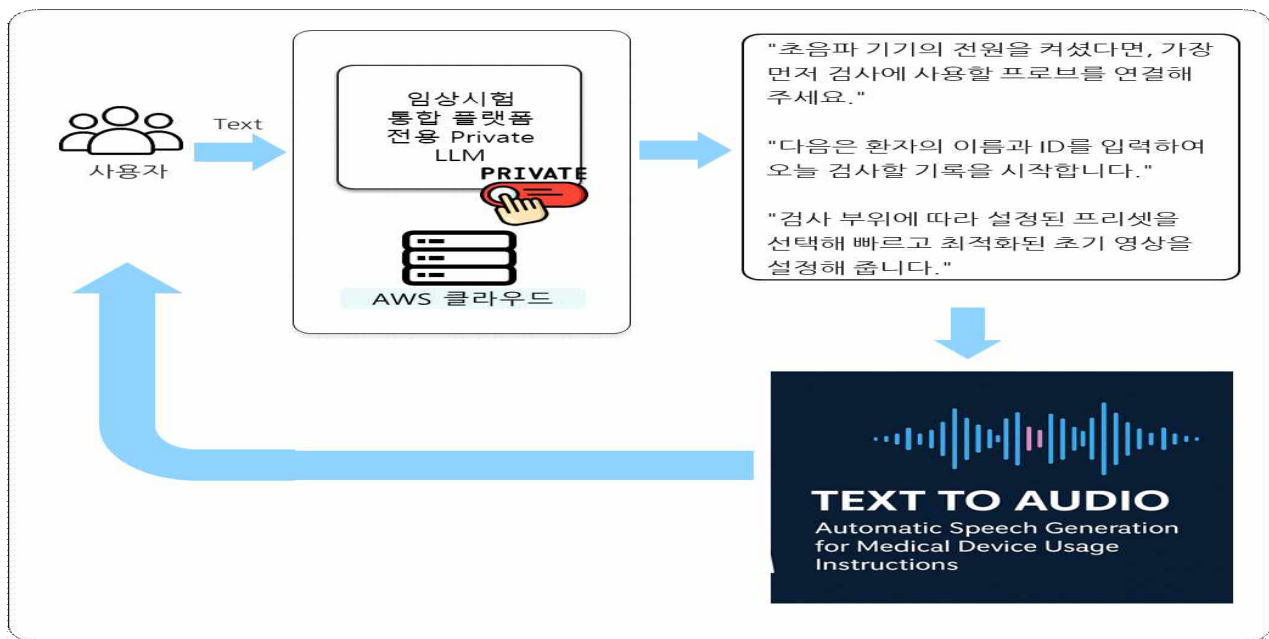
(사용자 요청) “초음파 영상기기의 기본 사용법에 대한 음성 안내를 생성해 줘.”

(시스템 처리) 1. Private LLM이 최신 매뉴얼을 기반으로 교육용 스크립트(Script)를 자동 구성. 구성된 스크립트를 Text-to-Audio 엔진을 통해 자연스러운 내레이션으로 변환

(출력 스크립트 예시)

“초음파 기기의 전원을 켜신 후, 검사에 적합한 프로브를 연결해 주십시오. 이어서 환자의 식별 정보를 입력하여 검사 기록을 생성합니다. 마지막으로 검사 부위에 맞는 프리셋(Preset)을 선택하여 최적화된 초기 영상을 설정하십시오.“

(최종 결과물) 대화형 UI 내에서 즉시 재생 및 다운로드가 가능한 고품질 음성 파일(Audio) 제공



마. 멀티모달 활용 서비스 시나리오 구현(Image-to-Video 예시. 립싱크 기능)

- Image-to-Video 기능은 사용자가 입력한 텍스트만으로 가상의 인물 이미지를 생성하고, 이를 음성과 동기화(Lip-sync)하여 마치 실제 사람이 설명하는 듯한 시청각 교육 영상을 자동 제작
- 임상 절차나 기기 조작법을 시각적 브리핑 형태로 제공함으로써, 교육 몰입도와 정보 전달력을 극대화 함
- [적용 예시] 가상 인물을 활용한 의료기기 조작법 브리핑 영상 자동 생성 (사용자 요청) 텍스트 프롬프트 입력 (예: “초음파 기기의 기본 세팅 방법에 대해 설명하는 안내 영상을 만들어 줘.”)

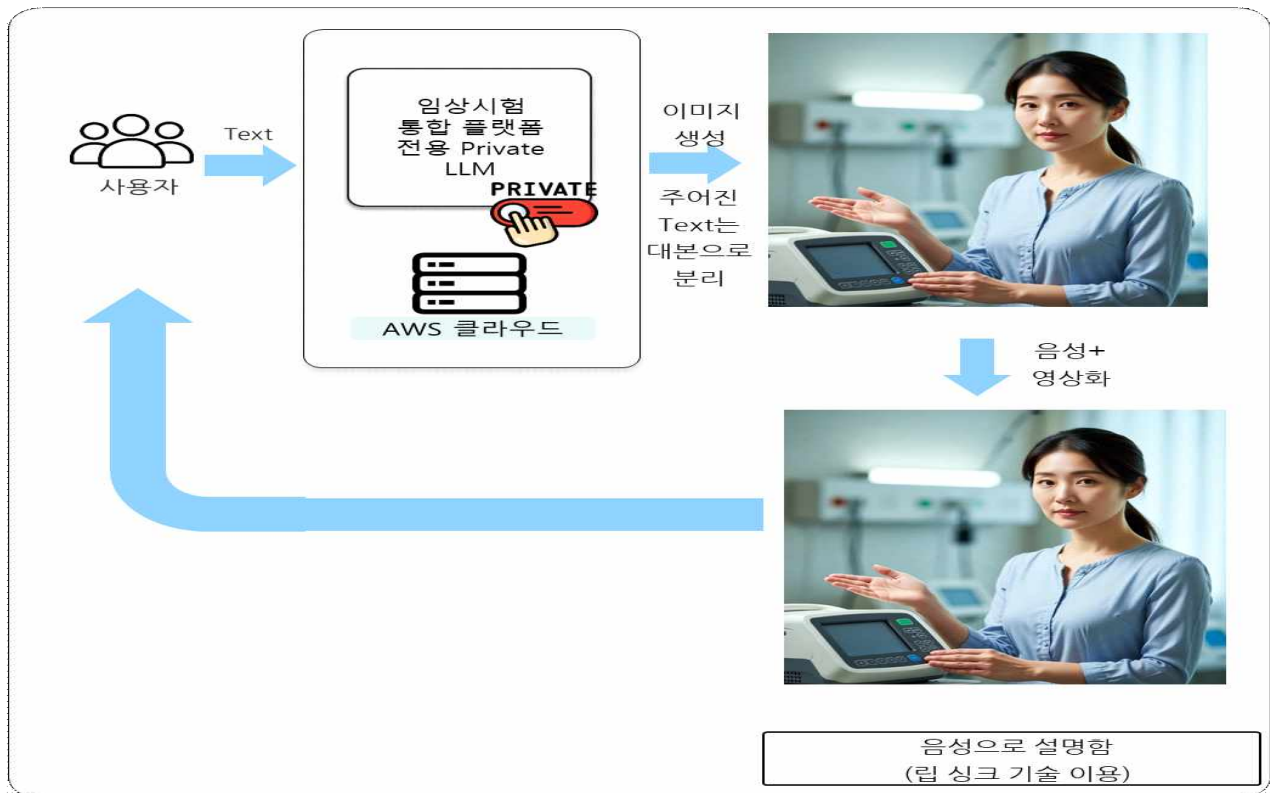
(시스템 처리)

[텍스트 분석 및 분리] AWS 클라우드 기반의 임상시험 통합플랫폼 전용 Private LLM이 사용자의 입력 텍스트를 분석하여, 영상 구성을 위한 '이미지 생성 지시어'와 '설명용 대본(Script)'으로 분리함.

[이미지 자동 생성] 분석된 지시어를 바탕으로, 설명 내용에 부합하는 가상의 안내자 인물 및 배경 이미지(예: 초음파 영상기기 옆에 대기 중인 임상 전문가)를 생성함.

[멀티모달 융합 (음성+영상화)] 정제된 대본을 자연스러운 음성으로 변환함과 동시에, 생성된 정적 인물 이미지에 립싱크(Lip-sync) 기술을 적용함. 대본의 음성 흐름에 맞춰 가상 인물의 입모양이 자연스럽게 움직이도록 구현

(최종 결과물) 기존 대화형 UI를 통해 사용자에게 회신되며, 재생 및 활용이 가능한 10초 내외의 '가상 인물 설명형(Talking Head) Short-duration 영상' 제공



4. 세부 요구사항

○ 요구사항 총괄표

구분	설명	요구사항 수
시스템 구성 요구사항 (Equipment Composition Requirement)	본 사업에서 도입 또는 활용하는 인프라 및 SW 등에 대한 요구사항 기술	1
기능 요구사항 (System Function Requirement)	본 사업에서 구축하는 시스템의 기능 요구사항 기술	5
인터페이스 요구사항 (System Interface Requirement)	본 사업에서 구축하는 시스템과 외부 시스템을 연결하는 인터페이스에 대한 요건을 기술	1
성능 요구사항 (Performance Requirement)	본 사업에서 구축하는 시스템의 성능에 대한 요구사항을 기술	1
데이터 요구사항 (Data Requirement)	본 사업에서 구축하는 시스템에 필요한 초기자료 구축 및 DB 설계 등 데이터와 관련된 필요 요건 기술	1
테스트 요구사항 (Test Requirement)	본 사업에서 구축하는 시스템의 테스트 및 점검을 위한 요건 기술	1
보안 요구사항 (Security Requirement)	정보 자산의 기밀성과 무결성을 위한 구축 시스템의 보안 요구사항 기술	2
품질 요구사항 (Quality Requirement)	본 사업의 원활한 수행 및 운영을 위해 관리가 필요한 품질 항목, 품질평가 대상 및 목표에 대한 요구사항 기술	1
제약사항 (Constraint Requirement)	시스템의 구축을 위해 본 사업에서 필요한 제약 및 요건을 기술	1
프로젝트 관리 요구사항 (Project Mgmt. Requirement)	프로젝트의 원활한 수행을 위한 관리방법 및 추진단계별 수행방안에 대한 요구사항을 기술	2
프로젝트 지원 요구사항 (Project Support Requirement)	프로젝트 수행 및 향후 지원을 위해 필요한 요구사항 기술	3

○ 요구사항 목록

구분	번호	요구사항 명칭	비고
시스템 구성 요구사항 (Equipment Composition Requirement)	ECR-001	클라우드 인프라 활용	
기능 요구사항 (System Function Requirement)	SFR-001	일반 및 서비스 제공 요구사항	
	SFR-002	Text-to-Audio 서비스 제공	
	SFR-003	Text-to-Video 서비스 제공	
	SFR-004	립싱크용 Image-to-Video 서비스 제공	
	SFR-005	시스템 및 서비스 운영/관리 기능	
인터페이스 요구사항 (System Interface Requirement)	SIR-001	사용자 인터페이스	
성능 요구사항 (Performance Requirement)	PER-001	성능요건 일반사항	
데이터 요구사항 (Data Requirement)	DAR-001	음성, 영상 생성 멀티모달 서비스 고도화를 위한 데이터 요구사항	
테스트 요구사항 (Test Requirement)	TER-001	테스트 일반	
보안 요구사항 (Security Requirement)	SER-001	보안지침 준수	
	SER-002	음성, 영상 멀티모달 서비스 보안	
품질 요구사항 (Quality Requirement)	QUR-001	장애 및 품질 보증 관리	
제약사항 (Constraint Requirement)	COR-001	음성, 영상 멀티모달 Private LLM 고도화 제약사항	
프로젝트 관리 요구사항 (Project Mgmt. Requirement)	PMR-001	사업 수행 관리	
	PMR-002	업무보고	
프로젝트 지원 요구사항 (Project Support Requirement)	PSR-001	프로젝트 지원 일반사항	
	PSR-002	교육훈련	
	PSR-003	유지보수	

○ 세부 요구사항

가. 시스템 구성 요구사항(Equipment Composition Requirement)

요구사항 분류	시스템 구성 요구사항	
요구사항 ID	ECR-001	
요구사항 명칭	클라우드 인프라 활용	
요구사항 상세내용	정의	기존 Private LLM 시스템에 음성 및 영상 멀티모달 기능을 효율적으로 추가·운영하기 위한 클라우드 기반 시스템 구성 요구사항
	세부 내용	<ul style="list-style-type: none"> ◦ 기존 클라우드 운영 환경에 사설 대형 대화 모델(Private LLM)과 연계하여 구축 및 운영 가능해야 함 ◦ 대규모 병렬 연산에 최적화된 GPU 서버 활용 ◦ 오픈소스 기반 멀티모달 모델 적용 ◦ 고가의 독점 플랫폼 사용을 지양하고 확장·유지보수가 가능한 유연한 구조 구성 ◦ 음성, 영상, 스크립트, 이미지 결과 저장을 위한 안전한 디스크 또는 스토리지 포함 ◦ 보안 가이드, 암호화, 사용자 접근 제한, 민감정보 처리 기준 준수

나. 기능 요구사항(System Function Requirement)

요구사항 분류	기능 요구사항	
요구사항 ID	SFR-001	
요구사항 명칭	일반 및 서비스 제공 요구사항	
요구사항 상세내용	정의	멀티모달 음성·영상 서비스 제공을 위한 기본 기능 및 사용자 경험 요구사항
	세부 내용	<ul style="list-style-type: none"> ◦ 기존 Private LLM 시스템과의 연계 유지 ◦ 챗봇 형태의 대화형 서비스로 음성 및 영상 요청 지원 ◦ 이미지 업로드 기능 구현 ◦ 생성된 음성 및 영상 결과의 재생 및 다운로드 기능 구현 ◦ 웹 표준 및 크로스 브라우징(Chrome, Edge 등) 준수 ◦ 기존 시스템과 통합된 로그인, 권한 기반 서비스 제공 (회원 전용)

요구사항 분류	기능 요구사항	
요구사항 ID	SFR-002	
요구사항 명칭	Text-to-Audio 서비스 제공	
요구사항 상세내용	정의	사용자가 요청한 절차 설명 및 교육 스크립트를 자연스러운 음성으로 제공하는 기능
	세부 내용	<ul style="list-style-type: none"> ◦ 기존 Text-to-Text 채팅창에서 생성된 답변 또는 스크립트를 음성으로 변환 가능해야 함 ◦ 교육용 음성 생성은 복잡하지 않고 명확한 설명 위주로 구현 ◦ 남성형, 여성형 대표 음성 제공 ◦ 관리자 조희가 가능하도록 생성 이력 저장 ◦ 기존 Text-to-Text LLM 모델과 통합 대화가 가능한 구조로 구성

요구사항 분류	기능 요구사항	
요구사항 ID	SFR-003	
요구사항 명칭	Text-to-Video 서비스 제공	
요구사항 상세내용	정의	사용자가 입력한 텍스트 프롬프트 또는 Private LLM이 생성한 절차 설명·교육 스크립트를 기반으로, 짧은 길이의 교육용 영상을 자동 생성하는 Text-to-Video 기능을 제공하기 위한 요구사항
	세부 내용	<ul style="list-style-type: none"> ◦ Text Prompt 또는 LLM 생성 스크립트를 기반으로 영상 생성 요청 처리 기능을 구현하여야 함 ◦ Text-to-Video는 직접 생성형 영상 방식으로 구현하며, 1회 생성 단위는 5초 내외를 기준으로 하여야 함 ◦ 의료기기 사용 절차, 설명 자료, 주의사항 등 교육 목적에 적합한 짧은 영상 생성이 가능하여야 함 ◦ 요청 프롬프트, 생성 스크립트, 생성 영상 결과 간의 연결 데이터를 관리할 수 있어야 함 ◦ 영상 생성 결과에 대한 메타정보(요청 일시, 주제, 사용자, 처리 상태, 생성 결과)를 자동 기록하여야 함 ◦ 기존 Private LLM 대화 흐름과 연계되어 후속 수정 요청, 재생성 요청, 반복 요청이 가능하여야 함 ◦ 텍스트·이미지·음성·영상이 혼합된 멀티모달 응답 흐름을 지원할 수 있어야 함

요구사항 분류	기능 요구사항	
요구사항 ID	SFR-004	
요구사항 명칭	립싱크용 Image-to-Video 서비스 제공	
요구사항 상세내용	정의	사용자가 입력한 이미지 또는 시스템에서 생성한 설명 인물 이미지를 기반으로, 음성과 입 모양을 동기화한 립싱크 영상을 자동 생성하는 Image-to-Video 기능을 제공하기 위한 요구사항
	세부 내용	<ul style="list-style-type: none"> ◦ 이미지 입력 또는 생성 인물 이미지를 활용한 립싱크 영상 생성 기능을 구현하여야 함 ◦ Text-to-Audio 기능과 연계하여 생성된 음성을 영상에 결합할 수 있어야 함 ◦ 립싱크용 영상은 1회 생성 단위는 10초 내외를 기준으로 하여야 함 ◦ 이미지, 음성, 영상 결과 간의 연결 데이터를 통합 관리할 수 있어야 함 ◦ 의료기기 사용법, 절차 설명, 안전지침 등 교육 목적의 설명형 영상 생성에 적합하여야 함 ◦ 영상 생성 메타정보(요청 일시, 주제, 사용자, 처리 상태, 생성 결과)를 자동 기록하여야 함 ◦ 기존 Private LLM 대화 흐름과 연계되어 후속 수정 요청, 재생성 요청, 반복 요청이 가능하여야 함 ◦ 텍스트·이미지·음성·영상이 혼합된 멀티모달 응답 흐름을 지원할 수 있어야 함

요구사항 분류	기능 요구사항	
요구사항 ID	SFR-005	
요구사항 명칭	시스템 및 서비스 운영/관리 기능	
요구사항 상세내용	정의	멀티모달 LLM 시스템 운영 및 성능·보안 관리 기능
	세부 내용	<ul style="list-style-type: none"> ◦ 음성·영상 요청 서비스 통계를 기존 대시보드 화면에 통합 반영 ◦ 요청 건수, 생성 이력, 사용자별 통계, 실패 사례 등 시각화 ◦ 생성된 음성 및 영상, 요청 내역, 오류 이력 조회 기능 제공(관리자용) ◦ 서비스 품질 모니터링(성능, 안정성, 오류율) 기능 제공

다. 인터페이스 요구사항(System Interface Requirement)

요구사항 분류	인터페이스 요구사항	
요구사항 ID	SIR-001	
요구사항 명칭	사용자 인터페이스	
요구사항 상세내용	정의	멀티모달 음성·영상 요청을 지원하는 직관적이고 일관성 있는 사용자 인터페이스 제공 요구사항
	세부 내용	<ul style="list-style-type: none"> ◦ 정보 접근성, 가독성, 편의성을 고려하여 멀티모달 요청이 직관적으로 수행되도록 구성 ◦ 사용자가 텍스트 요청, 이미지 입력, 음성 생성, 영상 생성을 한 화면에서 손쉽게 선택할 수 있도록 기존 채팅 UI에 적용 ◦ 모든 화면 및 기능에 대해 웹 표준 규격 적용 ◦ 오류 발생 시 명확한 오류 메시지와 해결 방안 안내 제공 ◦ 기존 사용자 인터페이스와 일관성 유지

라. 성능 요구사항(Performance Requirement)

요구사항 분류	성능 요구사항	
요구사항 ID	PER-001	
요구사항 명칭	성능요건 일반사항	
요구사항 상세내용	정의	음성, 영상 서비스 제공 시 사용자 요구에 적합한 응답 성능과 안정적인 시스템 운영을 위한 일반 성능요건
	세부 내용	<ul style="list-style-type: none"> ◦ 현재 시스템 성능을 고려한 개발 방안을 제시해야 함 ◦ 응답시간은 현재 하드웨어 구성과 GPU 환경을 고려해 최적화해야 하며, 처리 시간이 필요한 경우 진행 상태를 사용자에게 표시해야 함 ◦ Text-to-Video와 립싱크 기반 Image-to-Video는 생성 결과를 안정적으로 제공할 수 있도록 최적화해야 함 ◦ GPU 대기 시간 발생시 GPU 할당에 대한 표시 필수. 처리 시간은 GPU를 최대한 활용하여 최적화 할 것 ◦ 텍스트, 이미지, 음성, 영상 요청·응답 간 데이터 손실 및 오류가 발생하지 않도록 안정적인 서비스 제공 ◦ 데이터 형식 오류 시 오류 메시지 제공

마. 데이터 요구사항(Data Requirement)

요구사항 분류	데이터 요구사항	
요구사항 ID	DAR-001	
요구사항 명칭	음성, 영상 생성 멀티모달 서비스 고도화를 위한 데이터 요구사항	
요구사항 상세내용	정의	멀티모달(Text-to-Audio, Text-to-Video, Image-to-Video) 데이터 요구사항
	세부 내용	<ul style="list-style-type: none"> ◦ Text-to-Audio 요청 예제를 위한 최소한의 스크립트 예제 데이터 확보 ◦ Text-to-Video 요청 처리를 위한 프롬프트 및 시나리오 예제 데이터 확보 ◦ 립싱크 기반 Image-to-Video 처리를 위한 의료기기 설명 인물 이미지 및 음성 샘플 데이터 확보 ◦ 기존 텍스트 데이터와 연계 가능한 멀티모달 샘플 구축 ◦ 기존 Private LLM 데이터와 통합 활용 가능한 형식으로 관리

바. 테스트 요구사항(Test Requirement)

요구사항 분류	테스트 요구사항	
요구사항 ID	TER-001	
요구사항 명칭	테스트 일반	
요구사항 상세내용	정의	음성, 영상 멀티모달 서비스에 대한 테스트 계획 수립 및 실시 요구사항
	세부 내용	<ul style="list-style-type: none"> ◦ Text-to-Audio 서비스 테스트 : 스크립트 기반 음성이 적절히 생성되는지 테스트 ◦ Text-to-Video 서비스 테스트 : 사용자의 텍스트 요청에 대해 5초 내외의 적절한 영상이 생성되는지 테스트 ◦ 립싱크 기반 Image-to-Video 테스트 : 입력 이미지와 음성이 연계된 10초 내외 영상이 적절히 생성되는지 테스트 ◦ 일반 사용자 관점 사용성 테스트 : 멀티모달 요청이 편리하고 직관적으로 수행 가능한지 확인 ◦ 오류 처리 테스트 : 비정상 요청, 실패 시 오류 메시지 및 안내 제공 여부 확인 ◦ 관리자 대시보드 및 로그 검증 : 요청/응답 기록, 실패 사례 기록, 관리자 검수 기능 정상 작동 여부 확인

사. 보안 요구사항(Security Requirement)

요구사항 분류	보안 요구사항	
요구사항 ID	SER-001	
요구사항 명칭	보안지침 준수	
요구사항 상세내용	정의	보안지침 준수
	세부 내용	<p>물리적, 관리적, 기술적 보안대책 등 보안 관리 운영 규정 마련 투입 인원의 출입통제, 자료 누설 및 외부 유출 방지 대책 수립</p> <p>본 사업 참여 인력 전원의 보안서약서 제출</p> <p>시스템 설치 장소와 시스템 접근은 인가받은 담당자에 한하여 허용</p> <p>사업 수행 중 취득한 시스템 구성, 데이터, 음성, 영상, 스크립트 정보 등에 대한 보안 유지 책임 부담</p> <p>작업장소 보안 및 저장매체 반출입 통제 준수</p>

요구사항 분류	보안 요구사항	
요구사항 ID	SER-002	
요구사항 명칭	음성, 영상 멀티모달 서비스 보안	
요구사항 상세내용	정의	멀티모달 서비스 구축에 따른 기존 Private LLM 보안 체계 유지
	세부내용	<ul style="list-style-type: none"> ◦ 기존 Private LLM을 통한 사용자 인증 수행 ◦ 공공기관 보안 가이드라인을 준수하여 추가 개발 ◦ 개인정보보호 관련 법령 및 지침 준수 ◦ 추가 개발 관련 정보의 접근, 열람, 저장, 생성 이력에 대한 로그 저장 관리

아. 품질 요구사항(Quality Requirement)

요구사항 분류	품질 요구사항	
요구사항 ID	QUR-001	
요구사항 명칭	장애 및 품질 보증 관리	
요구사항 상세내용	정의	시스템의 원활한 운영 및 관리를 위한 품질 요구사항
	세부내용	<ul style="list-style-type: none"> ◦ 시간 제약없이 운영이 가능해야 함 ◦ 장애 발생 시 신속한 복구 절차 마련 ◦ 신속한 장애복구를 위한 장애 상황별 대응방안을 제시하고, 지원체계를 마련 ◦ 사용자 요청 처리 후 품질 피드백 시스템 제공

자. 제약사항(Constraint Requirement)

요구사항 분류	제약사항	
요구사항 ID	COR-001	
요구사항 명칭	음성, 영상 멀티모달 Private LLM 고도화 제약사항	
요구사항 상세내용	정의	기존 Private LLM 시스템 및 클라우드 운영 환경 내에서 확장·통합하기 위한 기술적·운영적 제약사항
	세부내용	<ul style="list-style-type: none"> ◦ 기존 Private LLM 시스템의 보안, 인증, 데이터 체계와 완전한 연계 및 통합 구조로 구축할 것

	<ul style="list-style-type: none"> ◦ 기존 클라우드 환경 내 멀티모달 처리 모듈 추가 설치 및 연동 ◦ 클라우드 기반으로 시간, 장소, 기기 제약 없이 기존 사용자 계정으로 연동 이용 가능해야 함 ◦ 기존 임상시험 정보관리시스템(MDCRO 등) 로그인 계정 체계와 동일하게 통합 연동 ◦ Long-duration video 영상은 본 과업 범위에서 제외하고 Short-duration video 중심으로 구축할 것
--	---

차. 프로젝트 관리 요구사항(Project Mgmt. Requirement)

요구사항 분류	프로젝트 관리 요구사항	
요구사항 ID	PMR-001	
요구사항 명칭	사업 수행 관리	
요구사항 상세내용	정의	사업 수행 요구사항
	세부내용	<ul style="list-style-type: none"> ◦ 계약 후 14일 이내에 본 사업의 수행을 위한 사업 수행계획서 제출 ◦ 일정, 품질, 의사소통, 위험 및 이슈, 인력관리, 형상 관리, 보안 보장 등에 관한 상세 계획 포함 ◦ 본 사업을 수행할 조직 및 팀원의 업무분장 내용을 상세히 제시 ◦ 사업 진행 중 투입인력에 대한 변경이 불가피한 경우에는 사전 공문을 통해 본원의 사전 승인을 받아야 하며, 교체인력은 최소 2주 전에 투입하여 인수인계함 ◦ 투입인력이 부득이한 경우(병가 등)로 1개월 이상 공석이 발생할 경우 동급 이상의 대체인력을 본원의 승인에 따라 투입함

요구사항 분류	프로젝트 관리 요구사항	
요구사항 ID	PMR-002	
요구사항 명칭	업무보고	
요구사항 상세내용	정의	프로젝트 업무보고 요구사항
	세부내용	<ul style="list-style-type: none"> ◦ 프로젝트 착수 시 착수 보고, 완료 시 완료 보고 실시 ◦ 월간 보고를 통한 업무 진행 상황을 정기적으로 보고하고 그 내용을 공유

	<ul style="list-style-type: none"> ◦ 기타 필요한 경우 수시 보고 실시 ◦ 보고 및 회의 진행 시 회의록 작성 및 공유
--	--

카. 프로젝트 지원 요구사항(Project Support Requirement)

요구사항 분류	프로젝트 지원 요구사항	
요구사항 ID	PSR-001	
요구사항 명칭	프로젝트 지원 일반사항	
요구사항 상세내용	정의	프로젝트 지원 일반사항
	세부내용	<ul style="list-style-type: none"> ◦ 제안사는 사업 수행을 위한 작업장소 등은 상호 협의하여 결정함 ◦ 제안사는 시스템 구축 및 운영과 관련된 기술이전 계획을 구체적으로 제시하고 수행하여야 함 ◦ 제안사는 본 사업 관련 분야의 정보기술에 대한 지속적 정보 제공 및 기술자문과 본원이 요구하는 기술인력을 지원하여야 함 ◦ 개발 및 테스트에 필요한 S/W, Tools 및 사무용품 등 개발환경 일체는 용역수행자가 자체적으로 조달하여 구축·운영하여야 함

요구사항 분류	프로젝트 지원 요구사항	
요구사항 ID	PSR-002	
요구사항 명칭	교육훈련	
요구사항 상세내용	정의	교육훈련 일반사항
	세부내용	<ul style="list-style-type: none"> ◦ 시스템을 쉽게 사용할 수 있는 교육훈련 방안을 제시하여야 함 - 멀티모달 기능 구현에 따른 추가 교육에 대해 제시 ◦ 시스템 관리와 사용자 교육은 시스템의 전반적인 이해와 장애 시 신속히 장애대응을 할 수 있도록 기술적인 부분을 포함하여야 함 ◦ 시스템 운용 및 유지관리 등에 대한 충분한 기술 이전 및 교육을 시행하여야 함 ◦ 교육을 위한 자료, 사용자 및 관리자 매뉴얼 제출

--	--	--

요구사항 분류	프로젝트 지원 요구사항	
요구사항 ID	PSR-003	
요구사항 명칭	유지보수	
요구사항 상세내용	정의	유지보수 일반사항
	세부내용	<ul style="list-style-type: none"> 무상 유지보수는 최종검수 후 1년간 보장함 하자보수란 하자보수 기간 중에 발견한 모든 종류의 소프트웨어 결함을 수정하는 것을 말함. 무상 유지보수 기간 중 정기 점검 방안 제시 무상 유지보수 기간 경과 후 발생된 하자에 대해서는 유상 유지보수로 진행

V 과업 보고

가. 착수보고 : 계약일로부터 7일 이내

- 과업수행계획서(분야별 참여인력 및 추진일정 포함) 제출

나. 중간보고 : 과업착수 후 2개월 이내

- 중간결과보고서 제출 및 중간보고

다. 최종보고 : 과업착수 후 4개월 이내

- 최종결과보고서 제출 및 최종보고

라. 성과품 목록

	구 분	납품수량	비 고
1	음성·영상 멀티모달 기능이 적용된 임상시험 통합플랫폼 전용 Private LLM 시스템	1 SET	사용자 매뉴얼 1건 소스 코드 1건
2	최종결과보고서	1부	한글 또는 PDF (원본(USB) 포함)

VI 과업의 일반지침

가. 일반지침

- 예산지급방식 : 선금은 요청 시 계약 총액의 50% 범위 내에서 지급, 잔금은 최종보고서 검수 후 지급

나. 특수지침

- 사후 서비스 기간 : 최종보고 후 1년간
 - 계약 종료 후에도 대규모 추가 비용 및 인력 투입이 발생하지 않는 범위 내에서 적극 대응
- 보안사항
 - 문서보안 관련하여 사업관련 자료 및 문서는 외부에 노출되지 않도록 보관·관리하고, 발주기관의 승인 없이 목적 외 사용이나 제3자에게 제공하지 않음