연합학습 기반 신약개발 가속화 프로젝트 사업

연합학습 플랫폼 구축 및 개발

착수보고회

evidnet KETI Korea Electronics Technology Institute 연세대학교 YONSEI UNIVERSITY 코어시큐리티(주)

연구 개발 목표

제약 관련 기관들의 **민감 데이터 및 연구비밀을 보호**하고, **이질적인 데이터를 통합**하여,
연합학습 기반 고신뢰 AI 모델 서비스를 제공하는 신약 개발 가속화 플랫폼

플랫폼 보안 강화


보안 문제 해결


보안 위협 식별


개인정보보호 대응


보안 레드팀 운영

연합학습 현안 해결


시스템 이질성 문제 해결


통계적 이질성 문제 해결


프라이버시 문제 해결


실제 운용 가능한 알고리즘

국내 최초 신약 개발 분야 연합 학습 플랫폼 확보

Private and Secure data | Local AI Model

Private and Secure data | Local AI Model

Private and Secure data | Local AI Model

Federated Drug Discovery Platform


이종 FAM 지원


친화적 인터페이스


클라우드 기반 관리


사업화

인센티브 매커니즘


클라이언트 기여도 평가

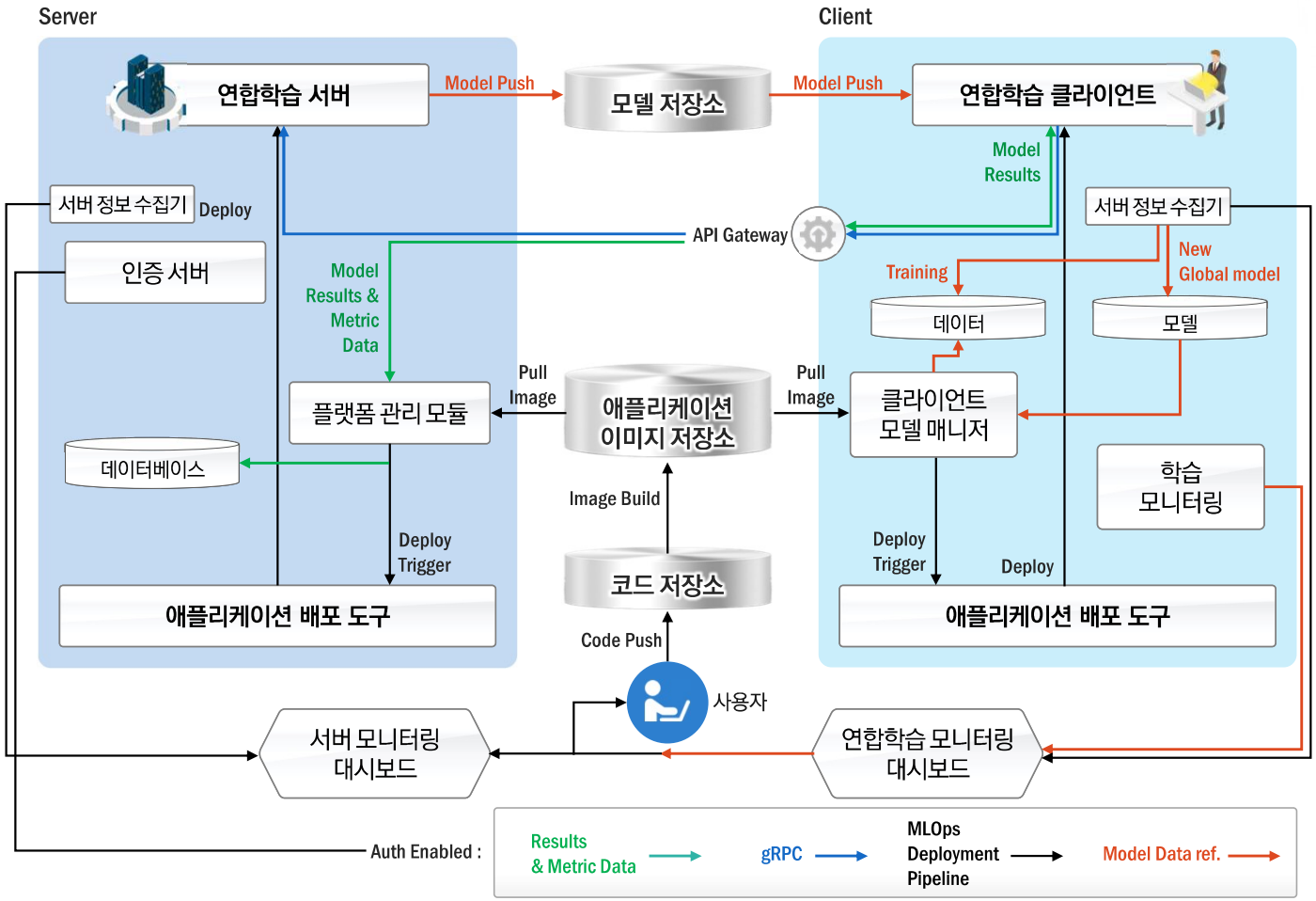

플랫폼 효용성 평가

Local AI Model 지원


모델 개인화


전처리 도구 개발 지원/배포

연합학습 플랫폼 아키텍쳐



MLOps 세부 내용

01 자동화된 CI/CD 파이프라인

- 사용자는 **코드 저장소** 및 **모델 저장소** 기반으로 자동화 된 CI/CD 파이프라인을 통해 항상 최신 연합학습 서버 및 클라이언트 이미지와 모델을 공유
- 자동화된 **애플리케이션 배포 도구**로 최신 이미지를 배포하여 학습

02 서버 상태 및 연합학습 모니터링

- **서버 메트릭 정보** 및 **연합학습 트레이닝 정보** 모니터링 시각화 대시보드
- Weights & Biases 등 모니터링에 특화된 오픈소스를 사용하여 효과적인 연합학습 결과 추적과 모니터링 수행

학습용 샘플 데이터 포털

세부 2의 20개 기관의 500개의 샘플 데이터 저장 및 배포
세부 3 기관이 데이터를 다운로드

주요 기능

- 세부 2 20개의 기관이 제공하는 500개의 샘플 데이터 저장 및 배포
- 메타 데이터 설명(데이터 컬럼, 타입 등)
- 데이터 제공기관명 표시(기관 요청 시 익명 처리)
- 편리한 데이터 검색 및 분류
- 데이터 이용 통계 (사용자, 기관별 통계 등)
- 사용자 유형별 기능 분리
- Logging & Audit (사용 로그 관리 및 보안 감사)
- 이중 인증 등 사용자 보안 장치 강화

연구개발 전략 – 연차별 개발 계획

1단계 2단계

2024년	2025년	2026년	2027년	2028년
연합학습 기본 기능	플랫폼 기본 기능	플랫폼 안정화	플랫폼 고도화	플랫폼 사업화
연합학습 구성 (서버, 클라이언트 등록)	사용자 인터페이스 (관리자, 데이터 소유자, AI 모델 공급자)	보안 강화 기술 적용 (클라우드, 통신, 보안, 위협 등을 보호하기 위한 보안 강화 기술 적용 확인)	모델 개인화 (글로벌 모델을 개별 기관 보유 데이터에 맞도록 미세조정)	솔루션 서비스화 또는 모듈화
전처리 도구 및 모델 탑재	실시간 학습 상황 모니터링	시스템 견고성 (클라이언트 오프라인, 데이터 이질성 등)	인센티브 매커니즘 (참여 혜택 배분)	추가 기관 연합학습 참여 (확장성)
서버-클라이언트 통신 기능				
연합학습 기능 (초기화, 로컬학습, 모델통합, 모델평가				

연구개발 최종 결과물 및 산출물

구분	연구개발 최종 결과물	산출물
FDD 플랫폼	FDD 플랫폼 프론트엔드 애플리케이션	서버탑재형 SW
	FDD 플랫폼 백엔드 애플리케이션	플랫폼 설계서
	학습용 샘플 데이터 포털	클라우드 인프라 구축 및 활용 계획서
	FDD 플랫폼 F/L 서버	플랫폼 개발 보고서
	FDD 플랫폼 F/L Client	플랫폼 사용자 매뉴얼
	플랫폼 위협 모델링 기반 보안 모델	기술 문서
	플랫폼 사이버 보안 취약점 점검	점검 결과 보고서
	플랫폼 레드팀 기반 시나리오 보안 침투 테스트	점검 결과 보고서
인센티브 매커니즘 및 기여도 평가 기술	연합학습 훈련 참여자의 FAM 모델 훈련 기여도 산정 기술	서버탑재형 SW
	이용자/훈련 참여자의 플랫폼 기여도 기반 인센티브 제공 기술	서버탑재형 SW
연합학습 주요 현안에 대한 FDD 플랫폼 제어 기술	참여자 시스템 이질성 기반 서버-클라이언트 제어 기술	서버탑재형 SW
	참여자 통계적 이질성 기반 서버 모델 훈련 제어 기술	서버탑재형 SW
	시나리오 기반 FAM 모델 보안 훈련 및 배포 기술	서버탑재형 SW
FDD 플랫폼 이용자에 대한 모델 개인화 지원 기술	시스템/통계적 이질성 기반 FAM 모델 추천 기술	서버탑재형 SW
	사용자 FAM 모델 개인화를 위한 훈련 가속 모듈	서버탑재형 SW
FDD 플랫폼 참여 기관 사이버 보안 체계 개발 및 적용	사용자 데이터 개인정보 보호체계 개발 및 점검	기술 문서
FAM 모델 훈련을 위한 전처리 도구 배포 기술	FDD 플랫폼 참여자 훈련 데이터 유효성 검증 모듈	서버탑재형 SW
연합학습 주요 현안 고려 고도화 알고리즘 구현 및 적용	연합학습 핵심 도전 과제 도출 및 대응방안 도출	현안 및 자문회의 정리 보고서
	데이터 불균형 완화 연합학습 모델 기술	공개 SW 1건
	모델 이질성 문제 완화 연합학습 모델 기술	공개 SW 1건, 논문 1건

연구개발 전략 - 추진체계

주관기관 에비드넷
책임자 이완희

에비드넷	한국전자기술연구원	연세대학교	코어시큐리티
연합학습 기반 신약 개발 플랫폼 설계 및 개발	연합학습 인센티브 및 개인화 기술 개발	데이터 불균형 문제 연구 및 모델 최적화	클라우드 환경 보안 메커니즘 개발 및 구현
플랫폼 개발 총괄	연합학습 핵심 기술 개발	연합학습 모델 개발 및 검증	보안 시스템 구축
플랫폼 운영	데이터 전처리 도구 개발 지원	데이터 분석 및 모델 고도화	데이터 보안 체계 개발
플랫폼 사업화	플랫폼 고도화 지원	모델 이질성 문제 해결	보안 취약점 점검 및 고도화

에비드넷 – 주관연구개발기관

evidnet KETI Korea Electronics Technology Institute 연세대학교 YONSEI UNIVERSITY 코어시큐리티(주)

evidnet
에비드넷

- ☑ 최대 분산 연구 의료 빅데이터 플랫폼(피더넷®, FeederNet®) 운영
- ☑ 피더넷®에서의 연합학습 기능 구현 및 운영

연구 책임자, 이완희

▶ (프로필)

- (주)에비드넷 CTO (상무)
- 2019년 ~ 현재 : (주)에비드넷
- 1999년 ~ 2019년: LG-CNS, 코난테크놀로지, 티맥스소프트, 네이블커뮤니케이션즈 등 근무
- 서울대학교 컴퓨터공학 학사(경력 25년)

▶ (연구분야) 분산 빅데이터 처리, 자연언어 처리(NLP)

주요 연구개발 실적

- ▶ 분산 네트워크 기반 의료 특화 데이터 '연합학습'을 활용한 예후 예측 지원시스템 개발 및 실증
- ▶ OMOP-CDM 기반 감염병 환자 정보관리 통합 시스템 구축 | 보건복지부 (한국보건산업진흥원)
- ▶ '마이헬스링크' 플랫폼을 통한 건강관리 올인원 서비스 | 과학기술정보통신부(한국데이터산업진흥원)
- ▶ 선행 공통데이터모델 기반 분산형 바이오헬스 통합 데이터망 구축 기술 개발 (연구수행 2만 건 돌파)

국내 최대 규모 의료 데이터 플랫폼 운영

- ▶ 압도적인 데이터 양과 품질관리 경험
- ▶ 데이터 표준화 및 품질 관리 노하우
- ▶ 데이터 보안 및 정보보호 역량

연합학습 플랫폼 구축 및 운영

- ▶ 자체 개발 연합학습 플랫폼
- ▶ 플랫폼 확장성 및 유연성
- ▶ 사용자 친화적인 인터페이스

의료 데이터 분석 및 활용 전문성

- ▶ 의료 전문가 네트워크
- ▶ 데이터 분석 및 시각화 도구
- ▶ 맞춤형 데이터 분석 서비스

한국전자기술연구원 – 공동연구개발기관

evidnet KETI Korea Electronics Technology Institute 연세대학교 YONSEI UNIVERSITY 코어시큐리티(주)

KETI Korea Electronics Technology Institute
한국전자기술연구원

- 초거대 인공지능 플랫폼, 교감형 AI, 생성형 AI 등 인공지능 분야 핵심 원천 기술 보유
- 공공, 환경, 산업현장 등 각종 분야의 난제 해결을 위한 인공지능 개발 및 운영 경험 다수

연구 책임자, 최원기

- (프로필)
  - 2021년 연세대학교 컴퓨터과학 박사
  - 2021년 ~ 현재 : 한국전자기술연구원 선임
  - 2021년 ~ 현재 : 디지털 트윈 팀 팀장
- (연구분야) MLOps 시스템, 디지털 트윈, 빅데이터 플랫폼
- 과기부, 산림청, 해수부 등 다수 정부과제에 실무책임자로서 참여

논문 등 주요 연구 실적

- IEEE ToC, TKDE 등 데이터베이스 분야 최상위 저널 게재
  ("OurRocks: offloading disk scan directly to GPU in write-optimized database system" (2020), "MV-FTL: An FTL that provides page-level multi-version management"(2017) 등)
- "워크로드별 최적 동기화 에이전트 선정 방법 및 시스템" 등 시스템 연합 관련 국내 특허 출원 20건(등록 13건), 미국특허출원 3건(등록 1건)을 보유
- TTA, 사물인터넷융합포럼 등 국내 표준화 그룹에서 분과위원으로 활동

연합 디지털 트윈 기술

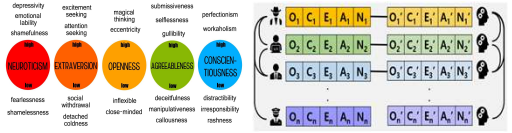
- 다양한 형태의 시스템을 연계·연합하기 위한 메타데이터 관리 및 동기화 기술에 대한 원천 기술 보유
  * 제주 서귀포 인근 관광객 안전을 위한 연합 플랫폼 운영

대규모 MLOps 인프라 운영 경험

- 인공지능 그랜드 챌린지를 운영하며 대회 참여자의 안정적인 AI 모델 개발을 지원하는 복합 MLOps 시스템 개발 경험
- 고성능 AI 인프라 보유 및 대규모 서비스 제공 경험

차세대 인공지능 핵심 원천 기술 보유

- 인간과의 상호작용을 통해 개성을 형성하는 AI 등 인공지능 분야 최첨단 핵심 원천 기술 개발 경험
- 국내외 인공지능 선도 연구팀과의 네트워크 보유



연세대학교 - 공동연구개발기관

연세대학교

- ✓ 다기관 의료 데이터를 활용하여 다양한 Federated Learning 알고리즘 개발 경험 보유
- ✓ 의료 데이터 과학 및 AI 분야에서 다양한 학문 분야와의 융합 연구 진행 및 AI 적용 기술 보유

연구 책임자, 성민동

- ▶ (프로필)
  - 연세대학교 의생명시스템정보학 박사
- ▶ 세브란스병원 내과전문의
- ▶ 세브란스병원 임상조교수
- ▶ (연구분야) 연합학습 관련 SCI/SCIE 포함 국내외 논문 약 9편, 특허출원 2건, 등록 1건

주요 연구 실적

- ▶ 수직 분할 데이터 분석 알고리즘을 이용한 대장암 전주기 예후 예측 (2020~2022)
- ▶ 다중 분할 임상 데이터 기반 분산형 컴퓨팅 기술 개발 및 검증 (2019 ~ 2021)

연합학습 및 플랫폼 사업 참여

- ▶ 산업통산자원부 주관 **"분산 네트워크 기반 의료 특화 데이터 연합학습을 활용한 예후예측 지원시스템 개발 및 실증 사업 (2021-2024)"**에 참여하여, 실제 임상에서 사용 가능한 대장암 환자 사망 예후 예측 알고리즘을 개발하여 **연합학습**에 적용 및 알고리즘을 고도화 하였고, 세브란스병원 데이터를 활용하여 실증함.
- ▶ 과학기술정보통신부 주관 "2023 데이터 기반 디지털 바이오 선도 사업 (2023-2027) " 에 참여하여, 멀티모달 사전학습과 데이터 합성 알고리즘 개발 및 플랫폼 연동 과제 수행 중.

논문 등 연구 실적

- ▶ "WICOX : Weight-Based Integrated Cox Model for Time-to-Event Data in Distributed Databases Without Data-Sharing (2023)", "A Novel privacy-preserving personalized progressive federated learning method for leveraging different healthcare institution-specific features (2024)" 등 연합학습 알고리즘 연구 결과 국제 학술대회 발표 및 학술지 게재
- ▶ 최근 5년 간 JAMA Network Open 등 Impact Factor 10점 이상 논문 10편 이상 발표, 주저자 논문 60편 이상 발표 **(Total IF: 267.79)**

특허 등 지식재산권 현황

- ▶ **7건의 연합학습 관련 지식재산권 출원**
  - ✓ 인공지능 기반의 수직, 수평 및 다중 분할 데이터에 대한 프라이버시 보존 분산 방법 및 장치 (출원번호 : 10-2022-0032316)
  - ✓ 수평 분할 데이터, 수직 분할 데이터 및 이들을 혼합한 다중 분할데이터에 대한 프라이버시 보존 분산 알고리즘 (출원번호 : 10-2021-0171520)
  - ✓ 프라이버시 보존 엣지-서버 시너지 컴퓨팅 장치 및 방법 (출원번호 : 10-2021-0109670)
  - ✓ 그 외 3건

코어시큐리티 - 공동연구개발기관

코어시큐리티(주)

코어시큐리티

- 다양한 산업 도메인의 보안 내재화를 위한 **보안 모델 핵심 기술 보유**
- 레드팀 기반의 취약점 점검, 침투테스트를 통한 **보안모델 및 보안성 컨설팅 경험** 다수 보유

연구 책임자, 윤민규

▶ (프로필)

- 2018년 성결대학교 정보통신공학 학사
- 2022년 ~ 현재 : 고려대학교 컴퓨터공학과 박사과정
- 2021년 ~ 현재 : 디지털 트윈 팀 팀장

▶ (연구분야) IoT/Embedded, 산업제어시스템(ICS), 의료기기 보안, 선박 및 해양 보안, 취약점 분석

주요 연구 실적

▶ (논문) IoT 기기 보안 점검을 위한 쉘 확보 방법론 연구 – 2024 정보보호학회 하계학술대회

▶ (논문) IEC 62443기반 Threat Modeling을 통한 PLC 보안 강화 연구 – 2024 정보보호학회 하계학술대회

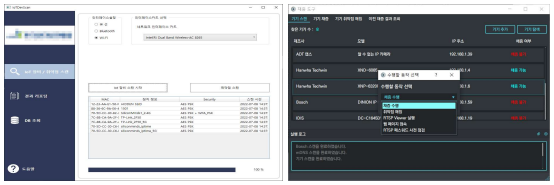
▶ (특허) IoT기기의 보안 취약점 분석 장치 및 방법과 그를 수행하도록 컴퓨터 판독 가능한 기록 매체에 저장된 프로그램(2023)

사이버 보안 전문 기술 보유

▶ 국가정보원과 국가안보실을 중심으로 운영되는 국가사이버안보센터의 국가사이버위관리단 민간협력사로 사이버보안에 대한 전문적인 기술과 경험을 보유

각종 산업 도메인에 특화된 취약점 점검 도구 보유

▶ IoT, IIoT, 의료기기 등 다양한 산업 도메인에 특화된 취약점 자동화 도구를 개발하고, 이에 대한 서비스를 제공



레드팀 기반 실시간 공격/방어 훈련 기술 보유

▶ 실제 공격 그룹의 데이터베이스인 MITRE ATT&CK Framework를 활용한 TTPs 중심의 공격 시나리오를 활용한 실시간 공격/방어 훈련 제공