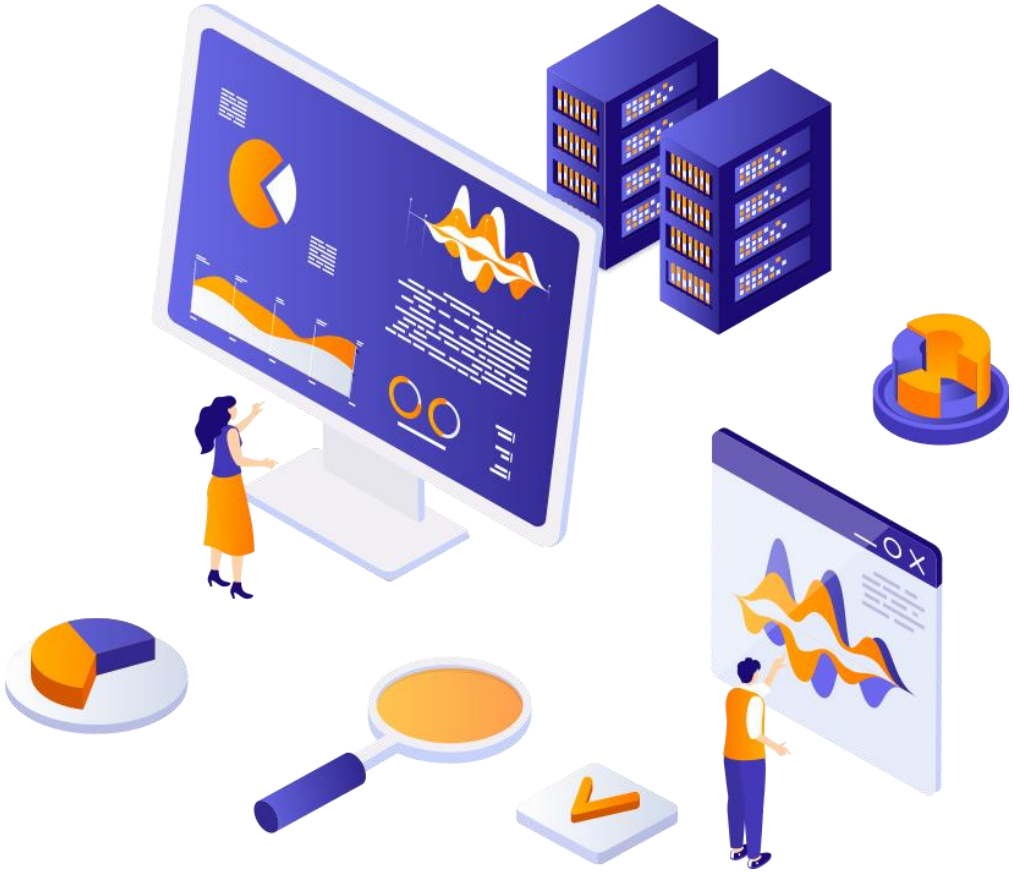
금융미래를 열어가는
금융보안파트너

AI를 활용한 금융사기거래 탐지



금융보안원
FINANCIAL SECURITY INSTITUTE

AI혁신실 김성웅 실장(swkim@fsec.or.kr)

Table of Contents

금융보안원
FINANCIAL SECURITY INSTITUTE

1. AI 기술 활용 현황 및 문제점

FDS를 위한 AI 기술 활용 현황

금융보안원 FINANCIAL SECURITY INSTITUTE

FDS를 위한 AI기술 활용 사례

- **글로벌 주요 금융기관**은 FDS에 AI기술을 적극 도입하여 활용 中
  - (HSBC은행) 매월 13억 건 이상 거래 모니터링
  - (JPMorgan) 실시간으로 고객 행동·신규 계좌 모니터링
- **국내 금융사들**도 머신러닝/딥러닝 등 AI기술을 도입하여 활용 中
  - (은행업권 예시) 전기통신금융사기 탐지 고도화를 위한 도입
  - (카드업권 예시) 신용카드 부정사용 탐지 고도화를 위한 도입

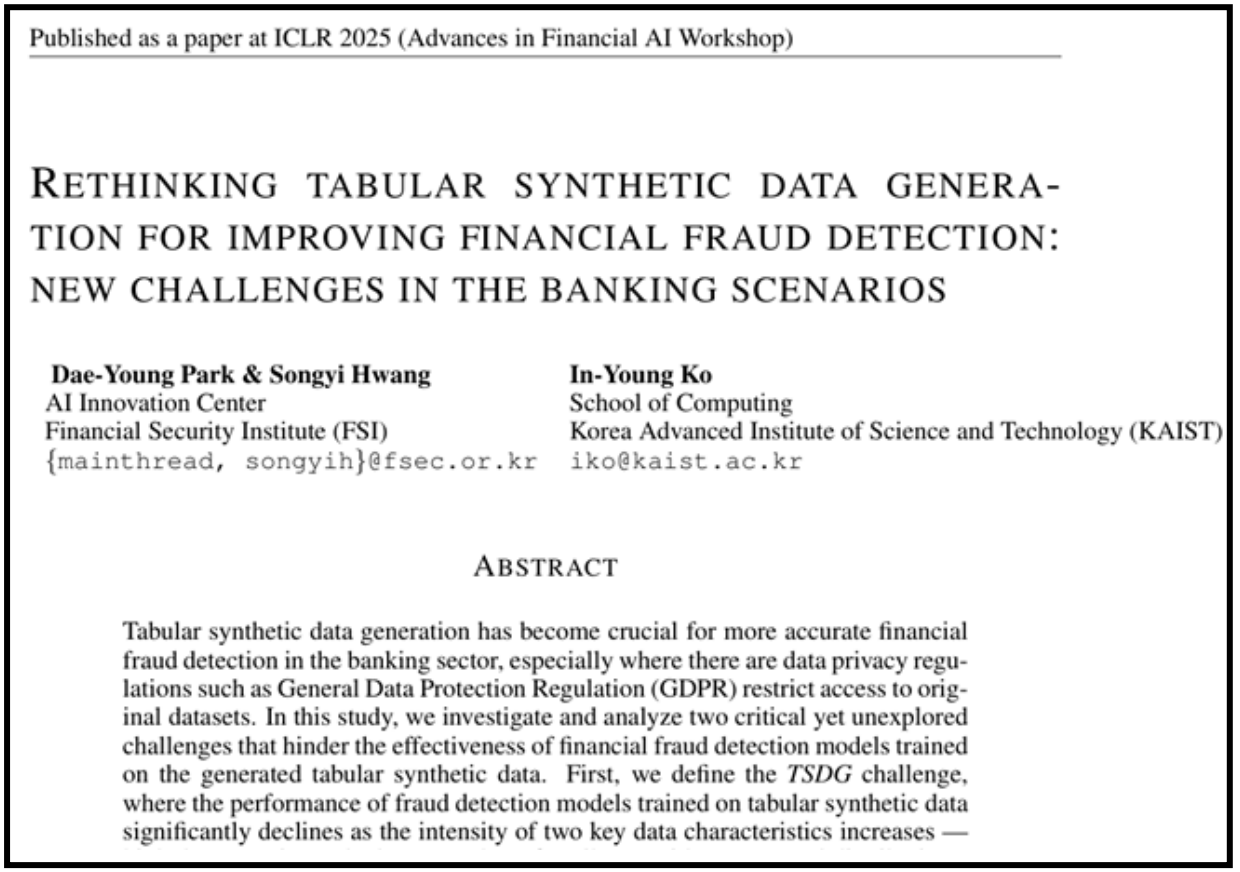
AI기술 활용의 문제점

- (SWIFT, NICT) 각 기관이 보유한 학습데이터 편중 및 범위의 한계 지적
  - AI모델은 각 기관 **자체의 과거 데이터만으로 학습**되므로 사기범이 여러 은행을 이용한 공격에는 취약함
  - 금융 사기는 갈수록 **정교하고 범국가적 양상**을 보이므로, 개별 은행 단독으로 대응하는 데 한계가 있음
- (BNP파리바 은행) 은행업권 전반의 협력이 필수적이며, **학습데이터**를 공유하여 AI를 활용해야 사기 탐지 역량을 **집단적으로 강화**할 수 있다고 강조

금융보안원 자체 연구

AI 기반 FDS 성능 비교 · 분석 연구 결과

- FDS에 활용가능한 이상금융거래 생성/탐지 모델에 대한 성능 및 장단점 비교 · 분석
  - 최대 규모 AI 국제학회 중 하나인 ICLR `25에서 발표 예정



Published as a paper at ICLR 2025 (Advances in Financial AI Workshop)

RETHINKING TABULAR SYNTHETIC DATA GENERATION FOR IMPROVING FINANCIAL FRAUD DETECTION: NEW CHALLENGES IN THE BANKING SCENARIOS

Dae-Young Park & Songyi Hwang
AI Innovation Center
Financial Security Institute (FSI)
{mainthread, songyih}@fsec.or.kr

In-Young Ko
School of Computing
Korea Advanced Institute of Science and Technology (KAIST)
iko@kaist.ac.kr

ABSTRACT

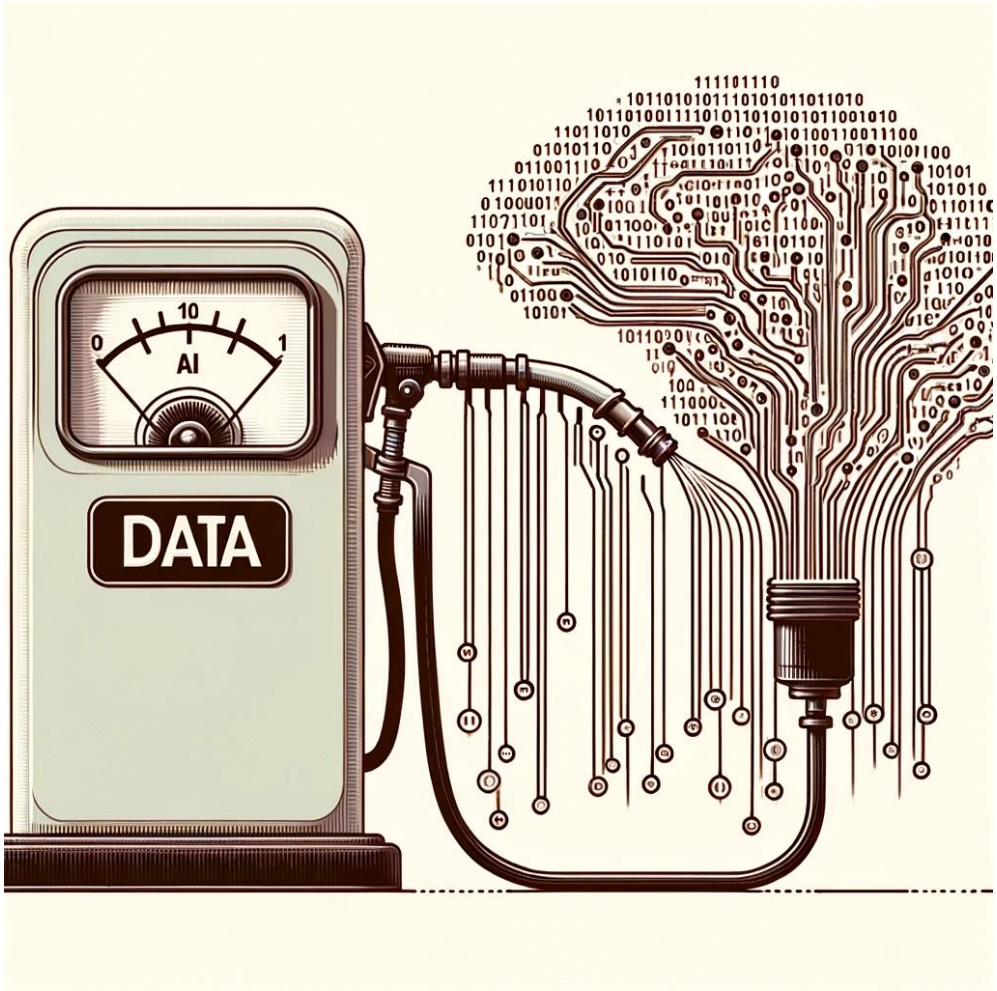
Tabular synthetic data generation has become crucial for more accurate financial fraud detection in the banking sector, especially where there are data privacy regulations such as General Data Protection Regulation (GDPR) restrict access to original datasets. In this study, we investigate and analyze two critical yet unexplored challenges that hinder the effectiveness of financial fraud detection models trained on the generated tabular synthetic data. First, we define the *TSDG* challenge, where the performance of fraud detection models trained on tabular synthetic data significantly declines as the intensity of two key data characteristics increases —

⇒ 원본 금융거래정보의 **희소성과 불규칙성**은 **개별 금융회사의 노력만으로는 극복에 한계**가 있으므로 **공동대응**이 필요

그러나.. 공동대응을 위한 학습데이터 공유 등은 **프라이버시 이슈**를 야기하므로 **연합학습 등 AI 신기술을 통한 공동모델 개발 등 협력 필요**

2. 연합학습 활용 FDS AI모델 공동 개발

인공지능의 연료 = 데이터



- AI로 해결하고 싶은데 데이터가 부족하네..
- 다른 기관이랑 공유하면 안될까?
- A 데이터는 많은데 B 데이터는 없네…
- Y 기관은 반대일텐데 서로 공유 안되나?

규 제

금융, 의료 등 규제산업에서는 어려운 영역

데이터 공유 시 개인정보 유출 등 이슈 상존

인공지능의 연료 = 데이터

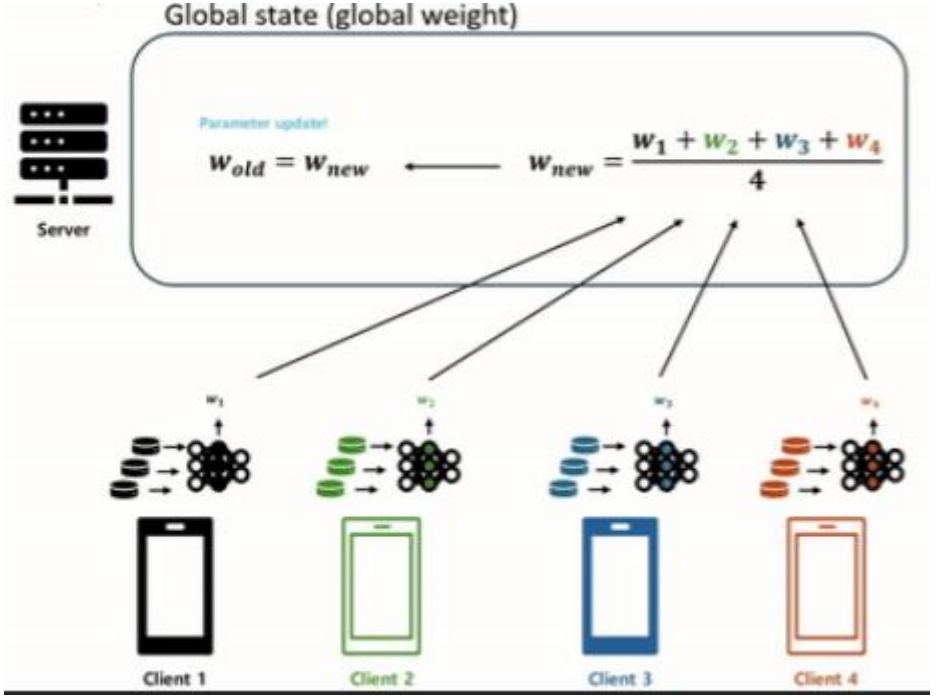
데이터 대신
모델을 공유한다면?

연합학습 개요

연합학습 정의

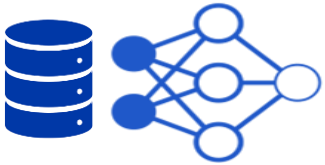
- AI모델 개발 목적이 같은 기관 간에 데이터를 공유하지 않고 공동 활용가능한 고성능 AI 모델을 개발할 수 있는 기술
- 각 참여자가 개별 학습한 뒤, 학습된 모델만 공유하여 통합 모델 구축
- 데이터 프라이버시를 보호하면서 다양한 데이터 활용의 효과

※ (참고) 개인정보보호위원회, 「개인정보보호 활용 기술 표준화 로드맵」 (`23.1.)



연합학습 필요성

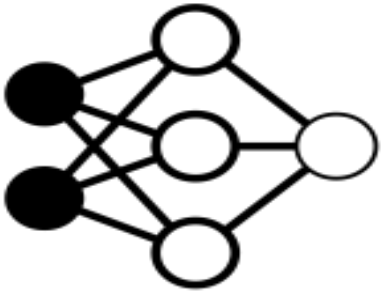
연합학습 미활용 시 – 데이터 공유 필요



A 기관

B 기관

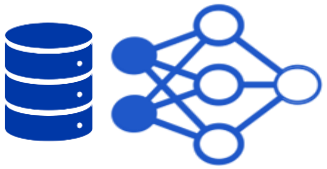
C 기관



ABC 공동모델

연합학습 필요성

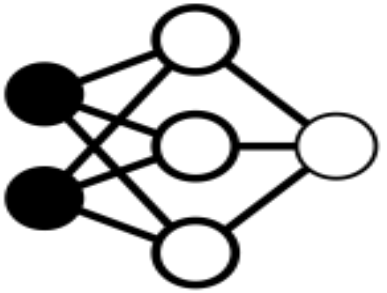
연합학습 활용 시 – 모델만 공유



A 기관

B 기관

C 기관



ABC 공동모델

금융보안원 FINANCIAL SECURITY INSTITUTE

국외 금융권 활용 사례

기업/기관	협력 대상	적용 업무	적용 효과
국제은행간 통신협회(SWIFT)	글로벌 은행	해외 결제 사기 방지	결제사기 탐지 정확도 개선
홍콩응용과학 기술연구소(ASTRI)	금융기관	• 금융범죄탐지 • 신용평가모형	• 탐지 성공률 증가 • 신용평가점수 산정 효율성 증대
영국 금융감독청, 미국 금융범죄 단속네트워크	금융기관	금융범죄예방	범죄예방을 위한 연합학습 기반 솔루션 공동 개발 中
일본 정보통신연구기구 (NICT)	일본은행 (5개)	사기 관련 계좌 탐지	기존 모델 대비 탐지건수 증가

FDS 고도화(개선)을 위한 연합학습 추진 경과

금융보안원 FINANCIAL SECURITY INSTITUTE

1. 금융보안원 AI혁신실 자체 개념검증 (완료)
2. 금융보안 AI 워킹그룹 및 은행업권 연합학습 소그룹 구성 (완료)
   - (참여기관) 하나은행, 농협은행, 케이뱅크, 카카오뱅크, 토스뱅크

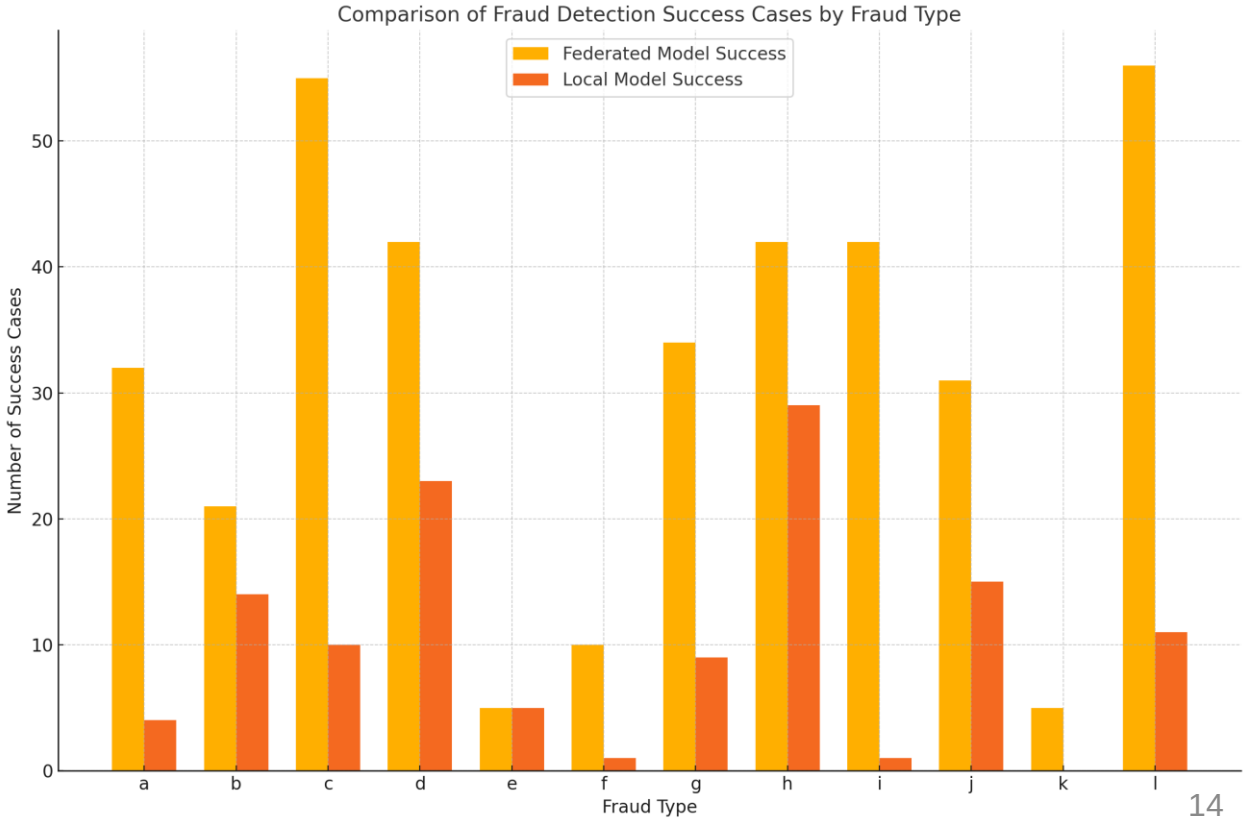
     ※ 일부 은행에서 FDS 고도화를 위한 연합학습 기반 공동 모델 개발에 대한 금융보안원의 주도적 역할 요청
   - (선정분야) 전자금융사기 사기계좌/거래 탐지 (피해자·피의자)
3. 서울대 공동연구 수행 (완료)
   - FDS 분야 연합학습 기술에 대한 실증 연구
4. 법률검토 (완료)
   - 개인정보로 학습된 AI모델의 개인정보 여부 등
5. 은행업권 연합학습 소그룹 운영 (완료)
   - FDS 연합학습 PoC 추진 (진행 중)
     - 데이터 레이아웃, 규격 및 세부 카테고리 논의
     - 개별 은행 학습데이터 준비
6. 카드업권 연합학습 소그룹 구성 (진행 중)

서울대 공동 연구 - 실증연구

실험 결과 요약 (일부)

사기 유형별로 연합모델은 탐지 성공한 건수가 훨씬 많음

- 노란색: 연합모델은 탐지성공, 로컬모델은 탐지실패한 건수
- 주황색: 연합모델은 탐지실패, 로컬모델은 탐지성공한 건수



서울대 공동 연구 - 실증연구

사기 유형별 설명

사기 유형	시나리오 설명	예시
a	물리적 거리 관련 사기 시나리오	특정 xx km 이상 거리이면서 x시간 차이 이내 같은 계좌에서 거래 발생 등
b	비정상적인 접속 관련 사기 시나리오	루팅/탈옥, 핸드폰 로밍 등 특이상태이면서 계좌 유형별 특정 금액 이상 출금 시도 등
c	악성앱/해킹 관련 사기 시나리오	악성앱 구동과 관련 시그널을 나타내는 속성값들이 임계치 이상인 경우
d	장기간 미사용 단말 관련 사기 시나리오	거래발생 기준 x일 이내 미사용로 접속 등 단말 관련 특이사항들을 충족한 경우
e	ATM 거래 관련 사기 시나리오	ATM 출금 한도 증액 여부, 최근 x일 이내 ATM 출금 일자 등 ATM 관련 특이사항들을 충족한 경우
f	영업점 거래 관련 사기 시나리오	계좌 유형별로 영업점에서 발생가능한 특이 방식으로 거래를 시도한 경우
g	무거래 계좌 관련 사기 시나리오	거래중지계좌 등 특이이력이 존재하는 계좌로 거래를 시도한 경우
h	고액 거래 관련 사기 시나리오	특정계좌유형에서 최근 x개월 동안 최대 이체 금액을 상회하는 거래액으로 이체 시도 등
i	새벽시간대 거래 관련 사기 시나리오	새벽시간대 특정 금액 및 거래건수 이상의 거래를 발생시킨 계좌와 관련된 거래 등
j	수취인계좌 관련 사기 시나리오	비정상적인 거래 이력을 지닌 수취계좌와의 거래 등
k	연속적인 출금 관련 사기 시나리오	특정 시간 내에 특정 횟수 이상 거래 발생 등
l	고령자 관련 사기 시나리오	고령자 고객 중 시스템 변경 (인증서 신규 발급 등)을 동반하여 비정상적인 거래가 발생한 경우

법률검토 결과 공유

연합학습 과정에서 AI 모델 송·수신 시 법적 리스크를 최소화하기 위한 AI 모델의 개인정보 여부에 관한 법률 검토

- 『안전한 인공지능(AI) 데이터 활용을 위한 AI 프라이버시 리스크 관리모델』(개인정보보호위원회, `24.12.19.) 및 『AI 모델의 익명성에 대한 의견』(EU 개인정보보호이사회, `24.12.17.)에 의거, **AI 모델 자체는 개인정보가 아니라는 의견**

> **3. 결론**
>
> 이상과 같은 검토에 의하면 AI 모델은 자체는 (i) 해당 정보의 성명, 주민등록번호 및 영상 등을 통하여 특정 개인을 알아볼 수 있는 정보가 아니며, (ii) 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 특정 개인을 알아볼 수 있는 정보에도 해당하지 않으므로 개인(신용)정보라고 볼 수 없습니다. 따라서 신용정보법 및 개인정보 보호법에 따른 기술적·물리적·관리적 보안대책의 적용 대상도 아닐 것으로 사료됩니다.

→ **다만, AI 모델 공유에 있어서도 프라이버시 이슈가 발생할 수 있다는 견해가 존재하기 때문에 관련부처의 명확한 해석이 필요**

은행업권 연합학습 소그룹 운영

학습데이터 레이아웃 세부 카테고리 논의 (`24.10.~`25.2.)

- 등록일자 – 일반회원가입, 실명고객등록일자 등
- 대출 신청 유형 - 신용대출, 담보대출, 할부금융 등
- 계좌유형 – 자유입출금계좌, 법인계좌(사업자통장), c: 저축계좌 (예적금 등), d: 모임통장 등
- 거래유형 – 현금출금, 예금이자, 대출, 간편이체, 해외송금, 캐시백, 증권계좌, 선불지급수단 등
- 거래채널 – 공과금수납기, 통장정리기, CD/ATM, 제휴사CD, 뱅크라인, 신용카드, 센터컷 등
- 거래시스템 접근 매체 – id/pw, CI, 주민등록번호, 패턴, 지문, 이체인증, 핀, OTP, 실명인증 등

학습데이터 수집기간 통일화 및 샘플링 방식 논의 (`25.1.~3.)

- A은행
  - 2024년 4월 13일 ~ 12월 31일 데이터 활용 가능
- B은행
  - 초기에는 100만 건 내의 record로 학습하는 것을 제안
  - 최대 약 3년까지의 대상 기간 설정이 가능
- C은행
  - 5개년의 데이터 사용가능
  - 23년 월 평균 거래건수는 대략 2900만건, 24년 월 평균 거래건수는 대략 4200만 건

금융보안원
FINANCIAL SECURITY INSTITUTE

3. 향후계획

은행업권 연합학습 PoC 일정 상세(안)

4-5월

- (금보원, 은행) 레이아웃 및 피처 세부 카테고리 통일
- (은행) 데이터 준비 (가공 및 샘플링 등)
- (금보원) PoC용 프로그램 (client-side) 구현

5-6월

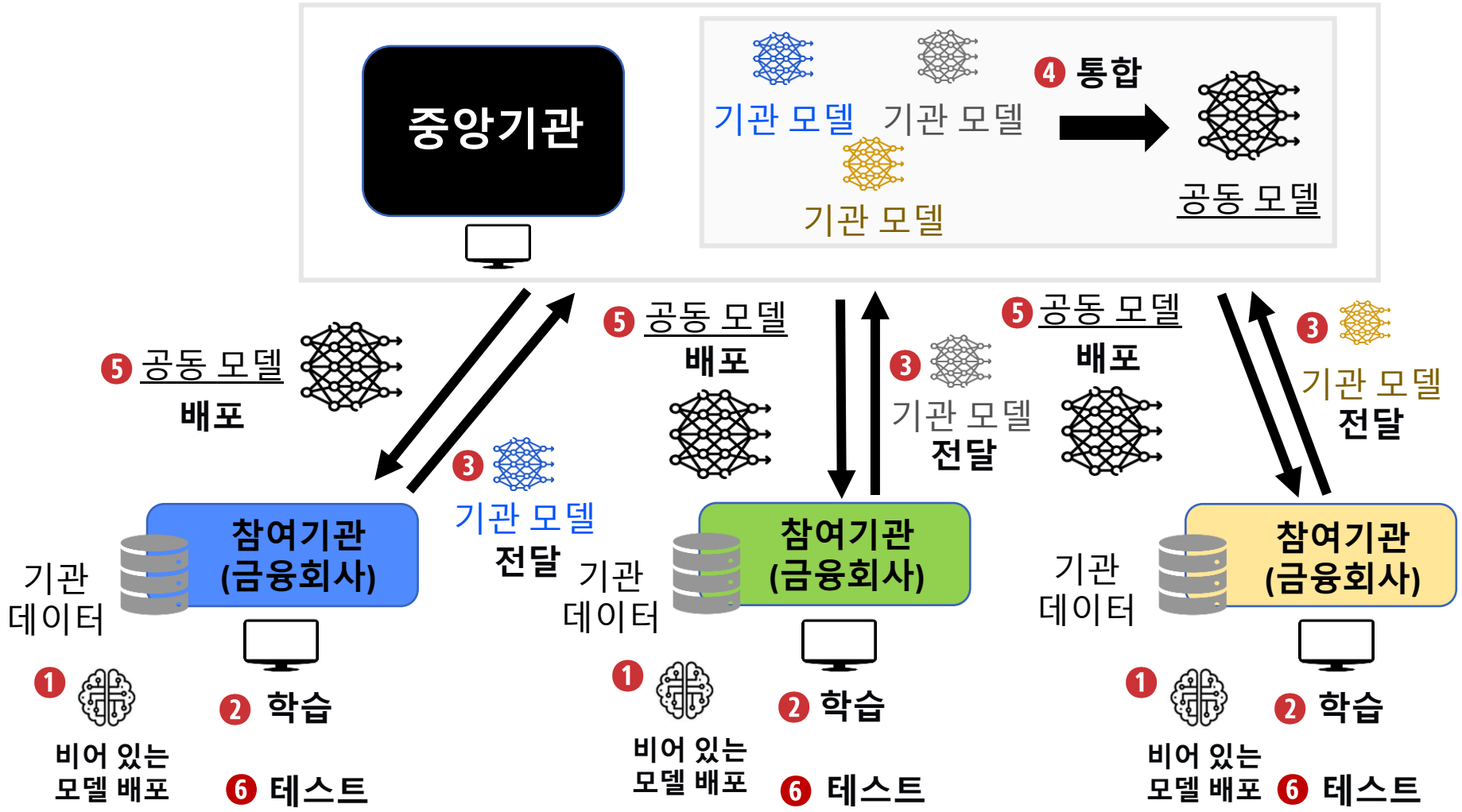
- (금보원) 코드 디버깅 (필요시)
- (은행) 개별 성능평가 후 모델, 연합학습 필요 정보 및 테스트 결과 전송
- (금보원) 연합학습 기반 공동모델 개발 (server-side) 및 전송
- (은행) 공동모델 성능평가 및 결과 전송

7-8월

- (금보원) 테스트 결과 분석
- (금보원, 은행) 성능개선 방안 도출 및 개선 수행
- (은행) 개선된 공동모델의 성능 테스트 추가 수행 및 결과 전송
- (금보원, 은행) 추가 성능 개선 시 위 과정 반복

* 금보원 및 은행 협의 내용에 따라 변동될 수 있음

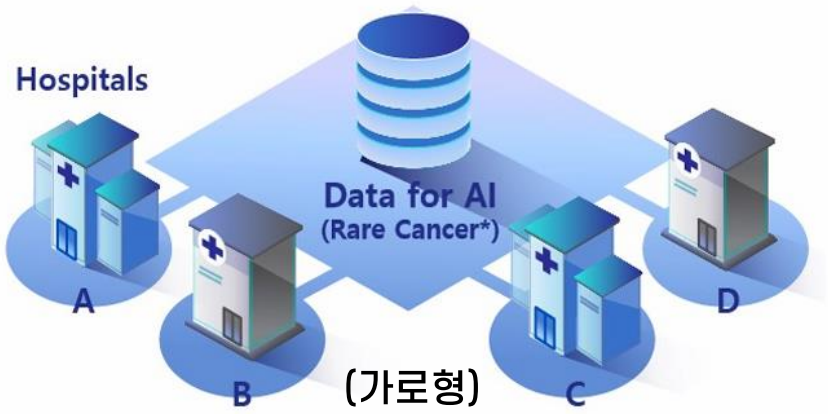
연합학습을 통한 공동모델 개발 및 테스트 과정



FDS 연합학습 업무의 장기적인 방향

가로형 연합학습에서 세로형 연합학습으로 확장 고려

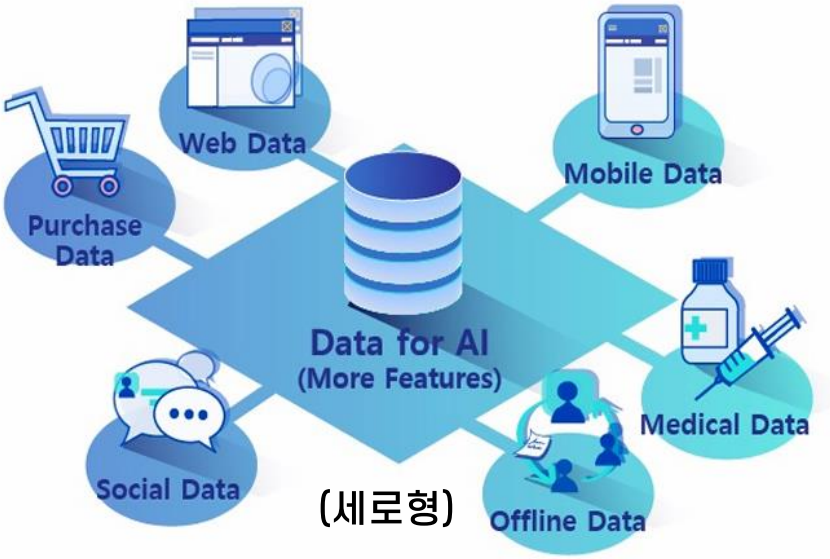
Many private samples



- 동일한 Feature에 대해 데이터 확장하는 효과

[예시] 사전 규격화된 A은행, B은행 간 공동모델 개발
→ 20~60대 전 연령에 걸친 데이터 활용 효과
→ 고성능 은행 전자금융사기 탐지 등

Many private features



- 동일한 데이터에 대해 Feature 확장하는 효과

[예시] C카드, D증권, E은행 간 공동모델 개발
→ 고객별 더 많은 특징, 행동정보 활용 효과
→ 고객 맞춤형 부정카드사용 사기탐지 등

금융미래를 열어가는 금융보안파트너

금융보안원
FINANCIAL SECURITY INSTITUTE

감사합니다.

금융보안원 홈페이지
www.fsec.or.kr