2026 한국인터넷거버넌스포럼

AI 학습용 합성데이터의 재식별 위험과 거버넌스 공백 : 금융 합성데이터의 기술적 평가와 법 및 정책적 대안

RE-IDENTIFICATION RISKS AND GOVERNANCE GAPS IN SYNTHETIC DATA FOR AI TRAINING

김고은 · 김효주 · 류봄 · 박서정 · 우서현 · 이혁재

목차

금융 합성데이터의 정의 및 재식별 위험성

정의 및 현황

- 실제 개인정보를 직접 포함하지 않고 생성된 데이터
- 원본 데이터의 통계적 특성, 분포를 학습하여 생성
- 개인정보 활용의 대안으로 주목받고 있음
- 금융 분야에서 실제 데이터 확보가 어려운 영역에서의 활용이 논의 되고 있음

FDS(이상거래탐지) AML(자금세탁방지) CSS(신용평가)

재식별 위험성

- 학습 과정에서 원본과 유사한 패턴이나 학습 포함 여부를 추론할 단서를 남길 수 있음
- 생성 방식과 원본 유사성에 따라 MIA, 추론공격 가능
- 외부 데이터 결합 시 개인 식별 가능성 증가
- 금융정보 결합 시 개인의 경제적 특성 추론 가능

⚠ **합성데이터는 개인정보를 포함하지 않더라도 재식별 및 추론 위험이 발생할 수 있음**

배경

AI 확산과 합성데이터

- 2022 ChatGPT 이후 LLM 시장 급성장 (Google·MS 등 참여)
- 통계 특성 재현 + 원본 미사용 → 보호·활용 동시 달성 수단
- 사전학습·미세조정에서 금융 ·의료 등 민감 분야로 확산

'완전한 익명'의 한계

- 생성 방식과 원본 유사성에 따라 MIA, 추론공격 가능
- 금융: 소비, 자산, 거래 시계열 → 경제상태 노출
- 시계열과 외부결합의 특성으로 추론 위험 심화 (BankSim·Simulacrum)

거버넌스 공백

- 생성 방식·유사성·익명성·재식별 위험 평가 기준 미비
- 체계적인 정책·규제가 아직 마련되지 않음 → 본 연구의 출발점

목적

금융을 중심으로 합성데이터의 재식별·추론 위험이 기술적·법적으로 얼마나 현실화됐는지 검토하고, 생성 방식·재식별 위험 평가·익명성 보호 측면에서 국내 거버넌스 공백을 진단한다

국내외 논의 흐름 : 법 & 정책

개인정보보호법
개인정보·가명정보·익명정보 구분과 재식별 가능성 판단

가명정보제도
가명처리·결합·안전조치 발전

NIST
활용 환경과 공개 맥락에 따른 리스크 기반 접근

GDPR
합리적으로 사용 가능한 모든 수단을 고려한 식별 가능성 판단

EDPB
AI 합성데이터 맥락에서 익명성 판단과 추론 공격 가능성 검토

FCA
금융서비스에서 모델 리스크, 데이터 거버넌스, 소비자 보호 논의

국내외 논의 흐름 : 선행 연구

2017

1 Shokri et al.

Shadow model 기반 MIA 제안

→ ML 모델 대상 멤버십 추론 공격 제시

2018

2 van Breugel et al.

공개 합성데이터만으로 MIA공격이 가능함을 보임

→ 합성데이터를 안전한 익명정보라 단정할 수 없음

2020

3 El Emam et al.

완전 합성데이터의 신원 노출 위험성 평가

→ 완전 합성데이터 역시 신원 노출 위험이 존재할 수 있으며, 별도의 위험평가가 필요

2024.2

4 「합성데이터 개인정보 노출 위험 평가 방안 연구」

합성데이터의 재식별 위험을 정량적으로 측정하는 방법론 제안

→ 합성데이터의 안전한 활용을 위해 위험 기반 평가체계가 필요

2024.4

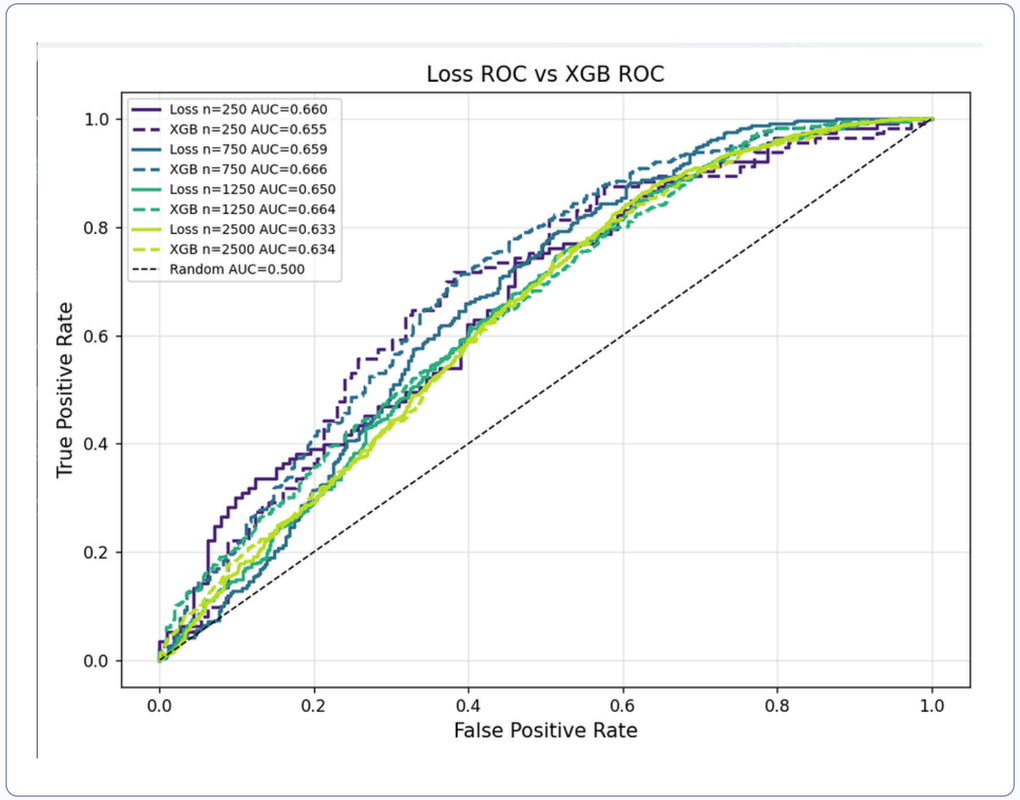
5 「오픈 데이터 환경에서 개인정보 노출 위험 측정을 위한 통계적 방법론 연구」

합성데이터 또한 개인정보 노출 위험이 존재

→ 활용 전에도 체계적인 위험 평가가 필요

실험결과

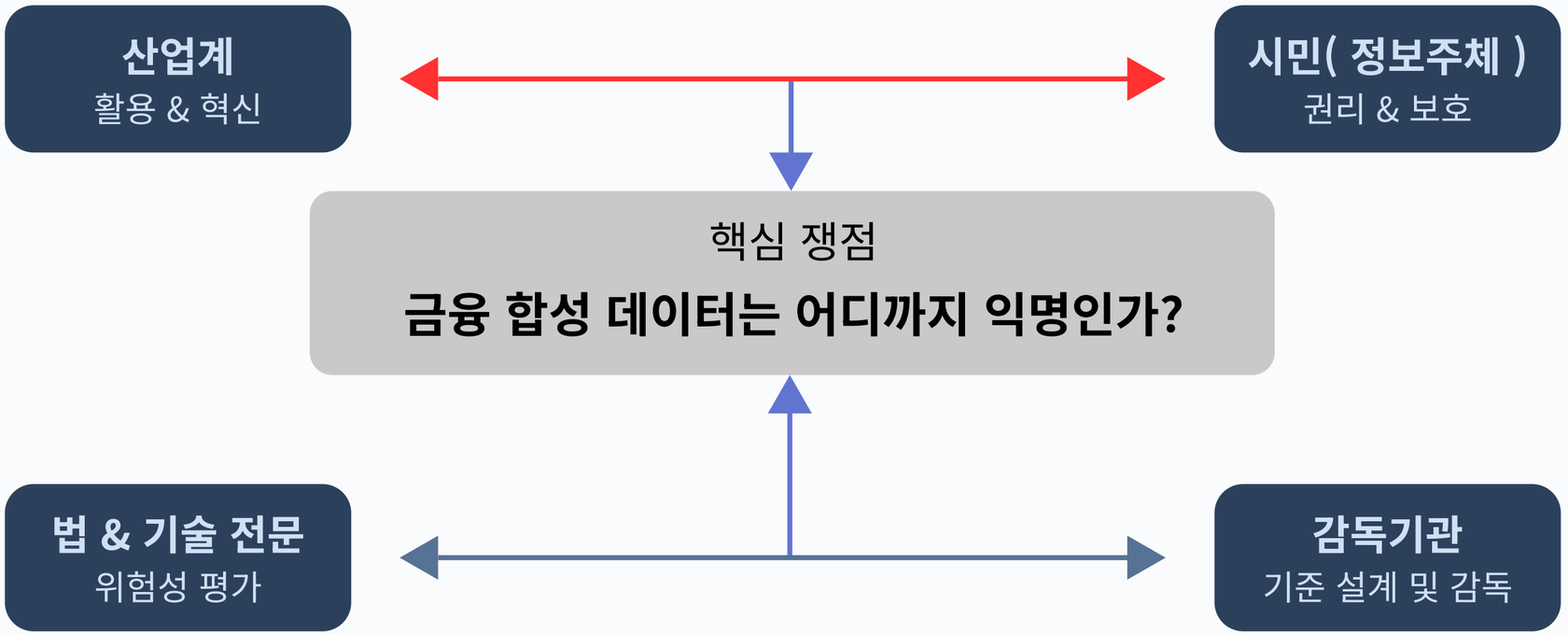
ROC곡선 (Loss vs XGBoost)



실험이 보여준 것

- 위험은 데이터에 한정되지 않음
- 학습된 모델의 반응값 (loss·confidence)을 통한 추론으로 발현
  - 모든 subset에서 AUC > 0.5
    → membership signal 존재

이해관계자 별 입장



관계

적색 → 충돌

청색 → 협력

쟁점 1

금융 합성데이터의 특징과 재식별 위험성

금융 분야 관점에서:

금융 데이터의 특수성

- 거래내역 & 소비패턴 포함
- 소득, 자산, 신용정보 결합
- 강한 시계열 특성
- 개인 생활양식 반영

» **변수들의 조합만으로도 특정 개인 또는 집단 추론 가능**

주요 공격 유형

- 멤버십 추론 공격 (MIA)
  → 특정 개인의 데이터가 학습에 포함
- 속성 추론 공격
  → 알려진 정보로 민감 속성 추론
- 재식별 공격
  → 외부 정보 결합을 통한 식별

» **직접 식별뿐 아니라 추론 공격까지 고려 필요**

금융 합성데이터 활용 시 고려사항

- 유용성 VS 개인정보 보호
- 희귀 패턴 재현 위험
- SNS·유출 데이터 결합 위험
- 특정 집단 과소대표 위험

» **변수들의 조합만으로도 특정 개인 또는 집단 추론 가능**

쟁점 2

금융 합성데이터 재식별 위험에 관한 글로벌 논의와 대응 흐름

미국 NIST	• 리스크 기반 접근 방식: 합성데이터의 활용 환경과 공개 맥락에 따라 재식별 위험을 함께 고려 • 데이터의 유용성과 개인정보 보호 사이의 균형을 강조
유럽	• GDPR: 개인 식별 가능성을 판단할 때 합리적으로 사용될 수 있는 모든 수단을 함께 고려해야 한다는 점을 강조 • EDPB: AI 및 합성데이터 맥락에서 익명성 판단, 추론 공격 가능성 등 위험 요소 사례별 검토 방향성 제시
영국 FCA	• 금융서비스 영역에서 합성데이터의 활용 가능성 및 모델 리스크, 데이터 거버넌스, 소비자 보호 문제 거론 • 금융소비자 보호를 위해 재식별 위험 관리 측면에서 별도의 검토가 필요하다는 논의

대응 흐름

기술적 조치만으로 판단하기보다,
목적과 범위, 결합 및 공격 가능성, 사후 관리 체계까지 함께 고려하는 방향으로 발전

쟁점 3

국내 금융 분야 합성데이터 재식별 위험 평가를 위한 기준 설계 방향과 시사점 검토

금융 분야 관점에서:

합성데이터의 익명정보 판단 여부

① 원본 데이터와의 유사성

② 외부 정보와의 결합 가능성

③ 생성 방식 및 활용 목적

④ 금융정보의 민감성

» 단일 기준이 아닌 유사성, 결합 가능성,
활용 맥락 등을 종합적으로 고려하여 평가해야함

금융 합성데이터의 재식별 가능성 대응 방안

정량적 요소 + 정성적 요소

전 주기적 평가 체계

» 복합 지표를 종합적으로 평가해야 하며,
생성부터 활용까지 전 주기적 위험 관리 체계가 필요함

**개인정보 보호와 유용성을 고려한
금융 합성데이터의 국내 설계 방향과 시사점**

유용성 vs 개인정보 보호

균형있는 기준 마련 필요

금융 환경 맞춤 설계

» 차등적인 평가 기준을 적용, 재식별 위험 뿐만 아니라
AI 성능 유지 여부 등 유용성까지 고려하는 체계를 마련

법·정책적 대응 방안: 위험 기반 차등 관리

정량·정성 요소를 종합한 위험도 등급화 & 차등 대응

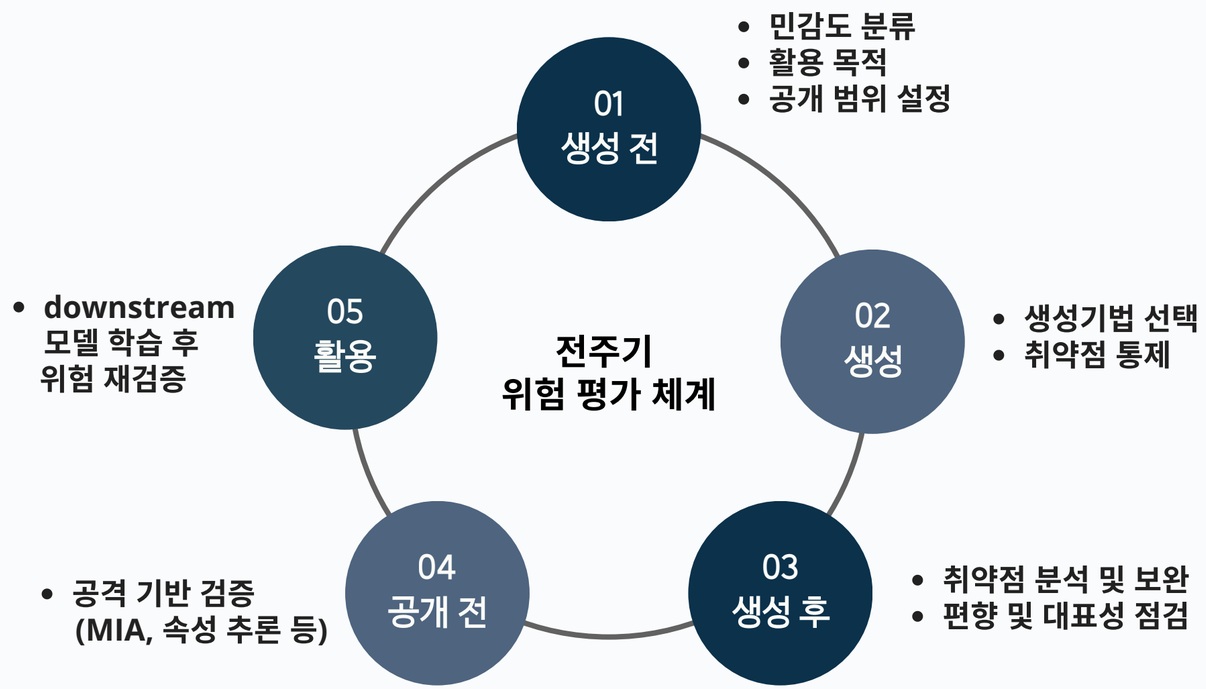
낮음	내부 활용 허용 및 정기 모니터링
중간	일반화 및 노이즈 처리, 접근권한 제한
높음	재생성·폐쇄형 분석환경, 공개 제한

맥락 기반 기준 적용

- 데이터셋 활용 목적 별 분류 및 차등 적용
- 단일 수치가 아닌 복수 지표를 통한 종합적 판단
- 데이터 특성 및 활용 환경 고려

평가 결과 활용 → 후속 조치 결정 (공개 범위 축소, 접근 제한, 재성성 등)

법·정책적 대응 방안: 전주기 위험 평가 체계



⚠ 지속적인 개선 및 반복적인 평가 필요

금융 데이터 맞춤 추가 평가 조치

FDS	AML	CSS
이상 거래 탐지 학습 데이터	자금 세탁 의심 패턴 학습 데이터	신용평가 모델 학습 데이터
희귀 패턴 (고소득, 대출)	변수 결합 (소득, 연체, 신용)	거래 시계열
회귀값의 재현율 점검	준식별자 조합 검토 추가 결합 가능성 검토	시계열 고유성 평가

기대 효과

01.

다층적 인식 형성

- 실제 멤버십 추론 · 재식별 공격 위험성 인지
- 금융 데이터 특수성에 따른 위험성 검토
- 이해관계자 별 합성데이터에 대한 관점 이해

02.

안전한 데이터 활용 환경 기반 마련

- 국내외 제도의 방향성 검토
- 위험 평가 및 검증 기준 논의
- 원본 유사성 및 결합 가능성 평가를 통한 침해 위험 감소

03.

법적 · 실무적 불확실성 완화

- 금융 합성데이터의 법적 지위 및 검증 절차 확립
- 생성자·제공자·활용자 간 책임 구조 정립
- 기업과 연구자의 부담 및 분쟁 가능성 감소

04.

연계 가능성

- 국제 논의 참여의 기반
- 후속 연구 주제 발굴
- 향후 정책 및 제도 개선 논의의 기초 자료

감사합니다

Team Information

김고은 (중앙대학교)

김효주 (서울여자대학교)

류봄 (이화여자대학교)

박서정 (숙명여자대학교)

우서현 (중앙대학교)

이혁재 (명지대학교)