

2026 Vol. 1

Privacy Report

2026 개인정보 이슈
심층 분석 보고서

프라이버시 강화 기술(PETs)

목차

Contents

1

연합학습·안전한 연산 기반 데이터 공동활용과 책임

박유량 연세대학교 의과대학 의생명시스템정보학교실 교수

7

CPO 관점에서 본 양자내성암호 전환

조지훈 삼성SDS 마스터

14

AI 학습용 합성데이터의 개인정보 보호 위험 분석과 안전한 활용 방향

안정연 KAIST 산업및시스템공학과 교수

연합학습·안전한 연산 기반 데이터 공동활용과 책임



박유랑

연세대학교 의과대학
의생명시스템정보학교실 교수

1. 개요

인공지능 모델의 성능과 일반화 가능성은 학습 데이터의 규모와 다양성에 크게 좌우되기 때문에, 한 기관의 데이터만으로는 한계가 있습니다. 금융기관의 이상거래 탐지, 통신·유통 데이터를 결합한 정책 분석, 여러 병원의 진료기록을 활용한 질병 예측 모델처럼 기관 간 데이터 공동활용 수요는 분야를 가리지 않고 늘어나고 있습니다. 그러나 원본 데이터를 한 곳에 모으는 방식은 개인정보의 제3자 제공 또는 결합에 따른 법적 절차, 재식별 위험, 유출 사고 시 책임 부담이라는 장벽에 부딪힙니다. 가명정보 결합 제도가 마련되어 있지만, 데이터를 결합전문기관으로 보내 결합하고 반출 심사를 거치는 과정에는 상당한 시간과 비용이 듭니다.

이러한 배경에서 데이터를 옮기는 대신 분석을 옮기는 연합학습(Federated Learning)과, 동형 암호·다자간 안전연산 등 안전한 연산(secure computation) 기술이 대안으로 주목받고 있습니다. 그러나 원본 데이터가 기관 밖으로 나가지 않는다고 해서 개인정보 쟁점이 사라지는 것은 아닙니다. 모델 업데이트와 학습된 모델에서도 개인에 관한 정보가 추론될 수 있고, 여러 기관과 기술사업자가 함께 참여하는 구조에서는 누가 개인정보처리자이며 어떤 책임을 지는지가 불분명해집니다. 본 글에서는 이 기술들의 활용 가능성과 개인정보 쟁점을 검토하고, 민감정보 활용 요구가 가장 높은 임상유전체·디지털 헬스케어 분야를 적용 예시로 살펴본 뒤, 참여기관·기술사업자·데이터 보유기관 간 역할분담과 보호조치 방안을 제안하고자 합니다.

2. 기술의 구조와 활용 가능성

(1) 연합학습의 구조

연합학습은 각 기관이나 기기가 자체 데이터로 모델을 학습하고, 그 결과인 모델 파라미터(또는 그라디언트)만 집계 서버로 보내 하나의 모델로 통합하는 방식입니다.^[1] 스마트폰 키보드의 다음 단어 예측처럼 이용자의 입력 데이터를 서버로 모으지 않고 기기 안에서 학습하는 서비스가 대표적인 초기 사례입니다.^[2] 같은 항목을 가진 서로 다른 집단을 모으는 수평적 연합학습과, 같은 사람에 대해 서로 다른 항목을 가진 기관(예: 카드사와 통신사, 병원과 보험사)이 협력하는 수직적 연합학습으로 나눌 수 있습니다.

(2) 안전한 연산 기술의 종류와 한계

연합학습만으로는 보호가 충분하지 않기 때문에 보통 다른 보호기술과 결합하여 사용합니다. 동형암호는 데이터를 암호화한 상태로 연산하여 집계 서버도 개별 기관의 업데이트 내용을 볼 수 없게 합니다. 다자간 안전연산과 안전한 집계(secure aggregation)는 여러 참여자가 입력값을 공개하지 않고 집계만 계산할 수 있게 합니다. 신뢰실행환경(TEE)은 하드웨어로 격리된 영역에서 연산을 수행하고, 차등프라이버시는 결과에 통계적 잡음을 더해 특정 개인의 포함 여부를 추론하기 어렵게 합니다.^[3]

다만 각 기술에는 연산 비용, 통신량 증가, 모델 정확도 저하 등의 비용이 따르며, 막아주는 위험도 서로 다릅니다. 예를 들어 동형암호는 전송과 집계 단계의 노출을 막지만 최종 모델을 대상으로 한 추론 공격은 막지 못합니다. 반대로 차등프라이버시는 최종 모델을 통한 추론 위험을 줄이지만 모델 정확도가 일부 낮아지며, 전송 중인 업데이트의 기밀성을 보장하는 기술은 아닙니다. 따라서 어떤 기술을 썼는지보다 어떤 위험을 어떤 기술 조합으로 줄였는지를 기준으로 보호 수준을 평가해야 합니다. 보호기술을 덧붙이는 것 외에 학습 방식 자체를 바꾸는 접근도 있습니다. 예를 들어 여러 차례 업데이트를 주고받는 대신 각 기관의 모델 가중치를 한 번만 전달하여 통합하면, 통신 부담과 함께 업데이트가 외부로 노출되는 횟수 자체를 줄일 수 있습니다.^[4] 미국 NIST가 2025년 차등프라이버시 보장 수준을 평가하는 지침을 확정한 것도 같은 방향의 움직임입니다.^[5]

(3) 적용 예시: 임상유전체와 디지털 헬스케어

이 기술들이 가장 활발하게 검증되고 있는 분야는 보건의료입니다. 건강정보와 유전정보는 민감정보이고 의료법과 생명윤리법이 함께 적용되어 원본 반출이 특히 어렵기 때문입니다. 코로나19 유행 당시 여러 대륙의 20개 의료기관이 흉부 영상과 임상정보를 공유하지 않고 산소 치료 필요성 예측 모델을 공동 학습하였으며, 단일 기관 모델보다 일반화 성능이 향상되었습니다.^[6] 71개 기관이 참여한 희귀 뇌종양 영상 연구도 연합학습으로 데이터 규모의 한계를 극복할 수 있음을 보였습니다.^[7]

임상유전체 분야에서는 유전체 데이터가 그 자체로 개인을 식별할 수 있는 고유한 정보라는 점에서 안전한 연산의 필요성이 더 큼니다. 다자간 안전연산을 이용하면 개별 유전형 정보를 노출하지 않고도 대규모 코호트 수준의 전장유전체 연관분석(GWAS)을 수행할 수 있음이 보고되었습니다.^[8] 국제표준화 측면에서도 ISO/TC 215(보건의료정보)에서 임상유전체 데이터의 2차 활용을 위한 연합학습 프레임워크 표준(ISO/NP 26502)이 신규 작업항목(NP) 단계로 개발되고 있으며,^[9] 동형암호 적용에 대한 표준화도 함께 논의되고 있습니다. 이러한 보건의료 분야의 사례는 금융·공공 분야에서 민감한 데이터를 공동활용하는 경우에도 참고가 됩니다.

3. 개인정보 쟁점

(1) 모델 업데이트와 학습된 모델은 개인정보인가

첫 번째 쟁점은 기관 밖으로 전송되는 모델 업데이트와 학습된 모델 자체가 개인정보에 해당하는 지입니다. 선행 연구에 따르면 공유된 그래디언트로부터 원본 학습 데이터를 복원하거나,^[10] 학습된 모델에 특정인의 데이터가 포함되었는지를 추론하거나,^[11] 대규모 언어모델에서 학습 데이터 원문을 추출하는 것이 가능합니다.^[12] 유럽 개인정보감독기구(EDPS)도 연합학습이 데이터 최소화에 기여한다고 평가하면서, 모델 업데이트를 통한 정보 유출 가능성 등 한계를 함께 지적하였습니다.^[13] 특히 유전체 데이터처럼 식별력이 높은 정보는 업데이트나 요약 통계만으로도 특정인의 참여 여부가 드러날 위험이 상대적으로 큼니다.

우리 개인정보 보호법은 다른 정보와 쉽게 결합하여 특정 개인을 알아볼 수 있는지를 기준으로 개인정보 여부를 판단하므로(제2조제1호나목), 모델 업데이트 역시 적용된 보호기술과 공격 가능성을 함께 고려하여 판단해야 합니다. 문제는 이를 위한 기준이 아직 없다는 점입니다. 모델 파라미터는 개인정보의 일부를 삭제하거나 대체하는 일반적인 가명처리 결과와 형태가 달라서, 기존 가명처리 기준을 그대로 적용하기 어렵습니다. 2026년 3월 전면 개정된 가명정보 처리 가이드라인은 활용 주체(내부 활용 또는 제3자 제공)와 처리 환경(제공기관의 통제 가능 여부)을 기준으로 기본 위험도를 저·중·고로 판단하도록 하고, 동형암호와 차등프라이버시 등을 가명·익명처리 기술로 소개하고 있습니다.^[14] 이 가이드라인이 연합학습이나 모델 업데이트를 직접 다루지는 않지만, 제공기관이 처리 환경을 통제할 수 있는지를 위험도 판단의 핵심 기준으로 삼는다는 점에서, 원본이 기관의 통제 범위를 벗어나지 않는 연합학습 구조를 평가할 때 참고할 수 있는 틀이 됩니다.

(2) 처리 행위의 법적 성격: 이용, 위탁, 제공

각 기관이 서비스 제공 목적으로 수집한 데이터로 모델을 학습하는 것은 원래 수집 목적과 다른 이용에 해당할 수 있습니다. 따라서 원본이 이동하지 않더라도 정보주체의 동의, 추가적 이용요건 충족, 가명처리 후 과학적 연구 목적의 처리(제28조의2) 등 별도의 근거가 필요합니다. 보건의료 분야에서는 연구의 성격에 따라 생명윤리법상 기관생명윤리위원회(IRB) 심의나 데이터심의위원회 심의도 함께 거쳐야 합니다. 다음으로 모델 업데이트를 집계 서버로 보내는 행위가 위탁인지 제3자 제공인지가 문제 됩니다. 기술사업자가 참여기관의 지시에 따라 집계만 수행한다면 위탁(제26조)으로 볼 수 있지만, 통합 모델을 자사 제품 개발에 활용한다면 제3자 제공 또는 독립적인 처리로 볼 여지가 생깁니다.

(3) 수직적 연합학습과 가명정보 결합

수직적 연합학습에서는 서로 다른 기관이 같은 사람을 연결하기 위해 비공개 교집합 연산(PSI)

등을 사용합니다. 이는 실질적으로 서로 다른 개인정보처리자 간의 결합에 해당할 수 있는데, 현행법은 가명정보 결합을 결합전문기관을 통해 수행하도록 하고 있습니다(제28조의3). 암호학적으로 안전한 연결 방식을 결합전문기관 절차와 같은 수준의 보호조치로 인정할 수 있는지에 대한 해석이 필요합니다.

(4) 국외 공동연구와 국외 이전

국제 공동연구나 글로벌 서비스에서 집계 서버가 해외에 있다면 모델 업데이트의 전송이 개인정보 국외 이전(제28조의8)에 해당하는지도 검토해야 합니다. 원본이 아닌 업데이트만 이동한다는 점은 위험을 낮추는 요소이지만, 이를 법적으로 어떻게 평가할지는 아직 정리되어 있지 않습니다. 국가 간 유전체 공동연구처럼 원본 반출이 사실상 불가능한 분야에서는 이 판단이 공동연구의 성사 여부를 좌우하기도 합니다.

4. 참여기관, 기술사업자, 데이터 보유기관 간 역할분담

연합학습을 실제로 운영할 때 가장 먼저 부딪히는 문제는 책임 구조입니다. 원본 데이터는 각 기관에 있고, 학습 목적은 공동사업 주관기관이 정하며, 인프라는 기술사업자가 운영하기 때문입니다. EU GDPR은 둘 이상의 주체가 처리 목적과 수단을 함께 결정하는 경우 공동처리자(joint controllers)로 보고 각자의 책임을 투명한 협약으로 정하도록 합니다.^[15] 유럽사법재판소도 개인정보에 직접 접근하지 않더라도 처리 목적과 수단의 결정에 참여한 주체는 공동처리자가 될 수 있다고 판단하였습니다.^[16] 우리 법에는 공동처리자 개념이 명시되어 있지 않으므로, 다음과 같이 역할을 구분하고 협약으로 책임을 명확히 할 필요가 있습니다.

데이터 보유기관(예: 병원, 바이오뱅크, 금융기관)은 로컬 학습 단계의 개인정보처리자입니다. 로컬 데이터의 처리 근거 확보, 로컬 노드에서의 가명처리와 데이터 품질 관리, 접근통제와 접속 기록 관리 등 안전성 확보조치, 정보주체에 대한 고지와 거부 방법 안내를 책임집니다. 또한 기관 밖으로 나가는 정보가 합의된 모델 업데이트에 한정되는지를 스스로 확인할 수 있어야 합니다.

참여기관(공동연구 주관기관 또는 컨소시엄)은 학습 목적, 대상 데이터의 범위, 결과 모델의 이용 범위를 결정하므로 실질적으로 공동처리자의 역할을 합니다. 공동처리 협약에는 목적 제한, 보호기술의 적용 수준(예: 안전한 집계 적용 여부, 차등프라이버시 파라미터), 결과 모델의 소유·공개·상용화 범위, 사고 발생 시 통지와 책임 분담, 사업 종료 후 모델과 중간 산출물의 파기 절차를 명시해야 합니다.

기술사업자(플랫폼 및 집계 서버 운영자)는 원칙적으로 수탁자로서 위탁 계약의 범위 안에서만 처리해야 합니다. 집계 서버가 개별 기관의 업데이트를 볼 수 없도록 설계 단계부터 보호조치를 반영하고(privacy by design), 소프트웨어 무결성 검증, 참여 노드 인증, 악의적 업데이트 탐지 등 기술적 책임을 집니다. 통합 모델이나 학습 로그를 자사 목적으로 재사용하려면 별도의 법적 근거와 계약이 있어야 하며, 그렇지 않다면 계약으로 금지해야 합니다.

이와 함께 정보주체 입장에서는 누구에게 문의하고 권리를 행사해야 하는지 알기 어렵기 때문에, 대표 창구를 지정하고 공동활용의 목적, 참여기관, 적용된 보호조치를 공개하는 것이 필요합니다.

5. 결론 및 제언

연합학습과 안전한 연산은 기관 간 데이터 공동활용의 유력한 수단이지만, 원본이 이동하지 않는다는 사실만으로 개인정보 보호가 완성되지는 않습니다. 기술적 보호와 법적 책임 구조가 함께 설계되어야 하며, 이를 위해 다음 네 가지를 제안합니다.

첫째, 모델 업데이트와 학습된 모델의 개인정보 해당성 판단 기준을 마련해야 합니다. 적용 기술의 조합, 공격 시나리오, 처리 환경에 따른 위험도 기반 판단 기준을 가명정보 처리 가이드라인과 연계하여 제시하면 현장의 불확실성을 크게 줄일 수 있습니다.

둘째, 연합학습 공동처리 협약 표준안을 보급해야 합니다. 데이터 보유기관, 참여기관, 기술사업자의 역할과 책임, 결과 모델의 이용 범위, 사고 대응 절차를 표준화하면 기관별 법률 검토 부담을 줄일 수 있습니다.

셋째, 보호기술의 검증 체계가 필요합니다. 동형암호, 안전한 집계, 차등프라이버시 등의 적용 수준을 제3자가 검증하는 테스트베드와 평가 기준을 마련하고, 기준을 충족한 경우 결합·제공 절차를 간소화하는 방안을 검토할 수 있습니다. 유럽 건강데이터공간 규정(EHDS)이 민감한 건강 데이터의 2차 활용을 보안처리환경(secure processing environment) 안에서 허용하는 구조도 참고할 수 있습니다.^[17]

넷째, 국제 표준과의 정합성을 확보해야 합니다. ISO/NP 26502 등 ISO와 GA4GH에서 진행 중인 연합학습·동형암호 표준화 논의와 국내 제도를 연계하면 국제 공동연구와 국외 이전 판단에 서로 예측 가능성을 높일 수 있습니다.

기관 간 데이터 공동활용의 수요는 앞으로 더 커질 것입니다. 연합학습과 안전한 연산이 신뢰받는 공동활용 수단이 되려면, 기술로 줄일 수 있는 위험과 남는 위험을 구분하고 이에 맞게 책임을 나누는 제도가 함께 마련되어야 합니다.

참고 문헌

- [1] McMahan HB, et al., Communication-Efficient Learning of Deep Networks from Decentralized Data, Proceedings of AISTATS (2017).
- [2] Hard A, et al., Federated Learning for Mobile Keyboard Prediction, arXiv:1811.03604 (2018).
- [3] ICO, Privacy-enhancing technologies (PETs) guidance (2023.06.); OECD, Emerging privacy-enhancing technologies: Current regulatory and policy approaches, OECD Digital Economy Papers No. 351 (2023.03.).
- [4] Park JA, Sung MD, Kim HH, Park YR, Weight-Based Framework for Predictive Modeling of Multiple Databases With Noniterative Communication Without Data Sharing: Privacy-Protecting Analytic Method for Multi-Institutional Studies, JMIR Medical Informatics 9(4):e21043 (2021).
- [5] NIST, SP 800-226 Guidelines for Evaluating Differential Privacy Guarantees (2025.03.).
- [6] Dayan I, et al., Federated learning for predicting clinical outcomes in patients with COVID-19, Nature Medicine 27:1735-1743 (2021).
- [7] Pati S, et al., Federated learning enables big data for rare cancer boundary detection, Nature Communications 13:7346 (2022).
- [8] Cho H, Wu DJ, Berger B, Secure genome-wide association analysis using multiparty computation, Nature Biotechnology 36:547-551 (2018).
- [9] ISO/NP 26502, Genomics informatics — Federated Learning Frameworks for Secondary Use of Clinical Genomic Data (ISO/TC 215, 개발 중).
- [10] Zhu L, Liu Z, Han S, Deep Leakage from Gradients, Advances in Neural Information Processing Systems 32 (2019).
- [11] Shokri R, et al., Membership Inference Attacks Against Machine Learning Models, IEEE Symposium on Security and Privacy (2017).
- [12] Carlini N, et al., Extracting Training Data from Large Language Models, USENIX Security Symposium (2021).
- [13] EDPS·AEPD, TechDispatch #1/2025 – Federated Learning (2025.06.10.).
- [14] 개인정보보호위원회, 가명정보 처리 가이드라인[본권] 가명정보 제도 안내편, 37면 및[별권] 가명정보 제도 실무편, 51-52면(2026.03.).
- [15] Regulation (EU) 2016/679 (GDPR) 제26조(공동처리자).
- [16] CJEU, C-210/16 Wirtschaftsakademie Schleswig-Holstein (2018.06.05.); CJEU, C-40/17 Fashion ID (2019.07.29.).
- [17] Regulation (EU) 2025/327 on the European Health Data Space (2025.02.11.).

CPO 관점에서 본 양자내성암호 전환



조지훈
삼성SDS 마스터

1. 들어가며

양자컴퓨팅의 위험은 대규모 오류보정 양자컴퓨터가 등장한 뒤 시작되는 것이 아니다. 공격자는 현재 탈취한 암호문을 보관했다가 미래에 복호화할 수 있다. 이른바 ‘수집 후 해독(Harvest Now, Decrypt Later)’이다. 유전체, 의료, 생체, 아동, 교육, 금융 정보처럼 오래 보호해야 하고 유출 후 변경하기 어려운 데이터일수록 현재의 암호 선택이 미래의 침해로 이어질 수 있다.^[1]

또한 양자내성암호(Post-Quantum Cryptography, 이하 ‘PQC’) 전환은 소프트웨어나 하드웨어의 일회성 교체가 아니다. 표준, 암호분석, 구현 취약점, 공급망이 변하면 알고리즘, 프로토콜, 키, 인증서를 다시 바꿔야 한다. 따라서 개인정보와 처리시스템, 소프트웨어, 클라우드 공급자, 그리고 암호자산의 관계를 지속적으로 파악해야 한다.

핵심은 반복 가능한 위험관리와 암호 민첩성(crypto agility)이다. 민감도와 보호기간에 따라 우선순위를 정하고, 서비스 중단을 최소화하며 암호 구성요소를 바꿀 수 있어야 한다.^[2] 불필요한 수집과 보유를 줄이고 목적 달성 후 파기하는 것도 재암호화 대상과 양자위험을 함께 줄일 수 있다.^[3]

여기서 CPO 관점은 개인정보 보호책임자(Chief Privacy Officer, 이하 ‘CPO’)가 암호기술을 단독으로 결정한다는 뜻이 아니다. CPO는 정보주체 피해와 개인정보의 민감도와 보호기간을 기준으로 우선순위를 제시하고, CISO-IT-개발-조달-법무-사업부서-공급자의 책임을 연결해야 한다. 이 글은 장기보관 개인정보의 양자위험과 국내외 정책을 살펴본 뒤, ISMS-P에 개인정보 분류, 암호자산 가시성, 암호 민첩성과 자동화된 디지털 증적을 결합하는 방안을 제시한다.

2. 장기보관 개인정보에 대한 양자위험

양자내성암호 전환 시점은 보호 필요기간(X), 전환기간(Y), 양자위험 도달기간(Z)을 함께 보는 Mosca 방식으로 판단할 수 있다. $X+Y > Z$ 보다 크면 준비를 시작해야 한다.^[4] NIST IR 8547 초안은 112비트 보안강도의 RSA-ECC 서명과 키 설정을 2030년 이후 ‘사용 중단 권고(deprecated)’로, 양자추약 공개키 방식을 2035년 이후 금지하는 일정을 제시한다. 확정 규범은 아니지만 역산 계획의 참조선이 된다.^[5]

표 1 보호 필요기간에 따른 예시적 우선순위

데이터 유형	보호 필요기간과 위험 특성	우선순위 판단
유전체·생체정보	· 변경·재발급이 어렵고 사실상 평생 영향을 미침	· 최우선
의료·장기보험 정보	· 법정 보존기간 또는 계약기간이 길고 민감도가 높음	· 높음
신원·인증·전자서명 기록	· 장기검증이 필요하고 위조 시 소금 피해 발생	· 높음
금융·거래 기록	· 법정 보존기간과 처리규모를 함께 고려	· 중간~높음
단기 로그·접속기록	· 단기간 내 가치가 감소하며 파기 여부가 중요	· 파기 우선

우선순위는 보유기간뿐 아니라 암호문의 외부 노출·복제 범위, 전송·접근 경로, 민감도와 재발급 가능성, 키 보호방식 및 공급자 전환 난이도를 함께 평가해 정해야 한다. 의료·금융·보험 정보처럼 개별법의 보존의무로 기간을 줄이기 어렵다면 그 기간 전체를 보호기간으로 보되, 법적·업무상 필요가 끝난 데이터는 재암호화보다 파기를 우선해야 한다. 불필요하게 긴 보존의무는 소관부처와 개인 정보 감독기관이 함께 재검토할 필요가 있다.

3. 해외 및 국내의 전환정책

(1) 미국·EU의 위험기반 전환

미국 정책은 알고리즘 선정에서 전환관리로 확장됐다. NIST는 2024년 ML-KEM·ML-DSA·SLH-DSA를 FIPS 203·204·205로 확정했고, OMB M-23-02는 연방기관의 우선순위 인벤토리에 알고리즘과 소프트웨어·운영체제·클라우드 공급자, 데이터 유형·보호기간을 함께 기록하도록 했다.^[6]

2026년 행정명령 14412와 OMB M-26-15는 이를 기한과 산출물이 있는 과제로 구체화했다. 고가치·고영향 시스템의 키 설정은 2030년 말, 전자서명은 2031년 말까지 전환하고, 책임자, 전환계획, 암호자재명세서(CBOM), 자동화된 탐지, 인벤토리 및 공급자 책임을 갖추도록 했다.^[7] 핵심은 암호자산 가시성, 책임체계와 자동화된 증적을 선행조건으로 삼은 데 있다.

EU도 회원국이 2026년 말까지 전환을 시작하고 고위험 활용사례를 2030년 말까지 보호하도록 권고한다. GDPR은 PQC를 직접 규정하지 않지만, 보유기간 제한, 보호설계, 위험과 기술수준에 상응하는 안전조치를 요구하므로 장기보관 데이터의 양자위험과 전환기간을 개인정보 위험 평가에 반영할 필요가 있다.^[8]

(2) 국내 전환정책과 법제

국내에서는 2023년 「범국가 양자내성암호 전환 마스터플랜」에 이어 2025년 9월 「범국가 양자내성 암호체계 전환 종합 추진계획」이 발표됐다. 이 계획은 2035년 전환 완료로 목표로 위험·영향도 평가, 암호자산 탐색·가시화, CBOM, 하이브리드 적용, 암호 민첩성, PKI 고도화와 기관별 준비계획을 제시한다.^[9] 국가 일정은 공통기준이지만, CPO는 개인정보 보호기간과 전환 소요 기간에 따라 더 이른 내부기한을 정할 수 있다.

로드맵의 암호자산 식별기준·우선순위 방법론과 기관별 준비계획은 연차별로 마련된다. 그 전에도 개인정보 처리방침·영향평가, 자산목록, 형상·구성 정보와 계약자료를 결합해 기초 인벤토리를 구축해야 한다. 특히 변경이 어려운 생체·유전체 정보와 장기검증이 필요한 전자서명 기록은 국가 목표보다 앞선 내부기한을 검토할 수 있다.

2026년 11월 시행 예정인 「양자과학기술 및 양자산업 육성에 관한 법률」은 공공기관 등이 기술지침에 따라 PQC·QKD 등 양자보안체계의 확보 및 적용계획을 수립·추진하도록 하고, 국가안보나 국민 생활에 큰 영향을 주는 양자기술 활용사업에는 별도의 영향평가 근거를 둔다. 이는 「개인정보 보호법」상 개인정보 영향평가와 법적 근거와 대상이 다르다. 다만 두 평가가 모두 필요한 사업이라면 데이터 흐름 및 위험·보호조치 검토가 겹칠 수 있으므로, 평가 시기를 조정하고 암호자산 현황, 데이터 흐름도, 개선계획을 공통 증적으로 활용하는 실무지침이 필요하다.^[10]

공공부문의 적용경로도 열리고 있다. 2025년 12월 개정 「암호모듈 구현안내서」는 검증대상 키 설정과 전자서명에 PQC를 결합한 하이브리드 구성을 허용한다. 그러나 PQC 부분은 KCMVP 시험범위 밖이므로, 조달 요구사항에 알고리즘 정확성, 상호운용성, 부채널 대응, 교체 및 업데이트 조건을 넣고 결과를 관리해야 한다.^[11]

2026년 9월 발표한 KCMVP 검증대상 PQC 후보군은 키 캡슐화 3종, 전자서명 4종이다.^[11] 기관은 후보군과 최종 검증대상을 구분하고, 일정, 인증서 구조 및 장기검증 요구가 다른 키 캡슐화와 전자서명을 별도 전환 트랙으로 관리해야 한다.

4. ISMS-P 기반의 PQC 전환 관리

(1) 기존 관리체계 안에서의 위험관리

국내 조직은 개인정보 보호법상 안전조치 의무와 ISMS-P 절차에 PQC를 연결하는 것이 현실적이다. 위험평가에는 개인정보 보호기간과 양자위험을, 자산목록에는 알고리즘·키·인증서·프로토콜을, 변경관리에는 기한·시험·복구계획을 반영할 수 있다.^[12]

CPO는 현재상태와 목표상태를 정하고, 무엇보다 개인정보, 시스템, 암호자산, 공급자의 의존관계를 식별해야 한다. 데이터 인벤토리에는 개인정보 유형, 위치, 보유·파기, 제3자 제공 등을, 시스템 인벤토리에는 애플리케이션, 라이브러리, API, 클라우드, SBOM 등을, 암호자산 인벤토리에는 알고리즘·프로토콜·키·인증서와 저장·통신·서명 적용지점을 담는다. 이를 ‘개인정보-처리시스템-SW공급자-책임자’로 연결해야 한다.

이 연결이 없으면 발견한 암호자산이 어떤 개인정보 위험과 관련되는지, 장기보관 개인정보가 어떤 통신경로, 라이브러리와 키로 보호되는지 판단하기 어렵다. 따라서 연결률을 관리지표로 두고, 시스템 도입, 클라우드 이전, 위탁계약 변경 및 개인정보 영향평가 때 인벤토리가 자동 또는 의무적으로 갱신되도록 해야 한다.

공공기관을 포함한 개인정보처리자는 ISMS-P 인증 의무와 관계없이 개인정보 보호법상 안전 조치 의무를 부담한다. PQC 전환에는 ISMS-P의 위험관리와 처리단계별 보호조치를 공통 실행기준으로 활용하되, 민감도, 처리규모, 재식별 가능성, 보호기간 및 제3자 구조를 결합한 ‘개인정보-양자위험 프로파일’을 마련해야 한다.

SI 활용은 비정형 데이터의 가시성을 더 떨어뜨린다. 학습·평가 데이터와 프롬프트 로그로 개인정보 사본이 늘면 원본 중심 인벤토리와 실제 소재가 어긋나고 암호 적용지점도 분산된다. 데이터 발견·분류·등급표지 자동화를 암호자산 탐지와 연결해 비정형 데이터까지 관리범위를 넓혀야 한다.^[13]

(2) 일회성 점검에서 지속적 보증으로

2026년 ISMS·ISMS-P 실효성 강화방안은 취약점 진단·모의침투, 현장실증 중심 심사와 인증 후 상시점검을 제시했다.^[14] 실제 통제의 작동을 확인한다는 점에서 의미가 있지만, 침투테스트는 특정 시점과 공격경로의 표본점검이다. 변화하는 자산·구성·공급망에서 암호통제가 유지되는지는 자동화된 주기 점검과 이력 증적으로 보완해야 한다.

FedRAMP 20x의 ‘living evidence’는 이러한 지속적 보증의 참고사례다. 통제 상태를 가능한 범위에서 자동 검증하고, 원천데이터·수집 및 변환로직·대상범위·판정기준·추세를 재현 가능하게 관리한다.^[15] 국내 인증에 그대로 적용하기보다 ISMS-P 통제의 운영효과성을 높이는 원칙으로 참고할 수 있다.

지속적 보증이 모든 통제의 실시간 감시를 뜻하지는 않는다. 위험과 변경 빈도에 따라 주기를 정하고, 승인 알고리즘·키 길이, 인증서 만료, TLS·VPN 설정, 라이브러리 버전과 PQC 적용 여부처럼 자동화 가능한 항목은 기계판독 증적으로 전환한다. 예외 사유, 정보주체 피해와 잔여위험 수용은 표본검증과 CPO·의사결정기구의 판단을 병행해야 한다.

암호자산 인벤토리 도구는 계획단계에서 양자취약 알고리즘과 전환대상을 찾고, 전환 후에는 소스코드·빌드·배포·운영환경에서 취약 암호의 재유입, 미승인 설정변경, PQC 미적용과 암호 후퇴를 주기적으로 탐지할 수 있다.^[16] 결과를 ISMS-P 내부점검과 증적 저장소에 연결하면 DevSecOps 전반의 점검시각·범위·판정기준·예외·조치이력을 축적할 수 있다. 다만 자동탐지 결과는 보증의 입력자료이므로 탐지범위, 오탐·미탐, 수집로직과 증적 무결성도 검증해야 한다.

5. CPO 중심의 PQC 전환 거버넌스

ISMS-P에 PQC 항목을 반영한 뒤에는 의사결정과 실행책임은 하나의 증적체계로 연결해야 한다. CPO는 정보주체 피해를 기준으로 목표 보호수준·우선순위·예외 원칙을 정하고, CISO·IT·개발·조달·법무는 이를 기술변경과 계약조건으로 이행한다. 표 2는 이를 ISMS-P 운영주기에 배치한 것이다.

표 2 CPO 중심의 개인정보·PQC 전환 운영주기

운영단계	CPO의 주요과제	공동 이행과제	주요 산출물·증적
방향설정	· 개인정보 위험기준, 목표 보호수준, 우선순위와 예외 원칙 설정	· 책임자·예산·공급자 요구사항 확정	정책, 책임체계, 목표상태
식별·평가	· 개인정보 유형·민감도·보유기간·흐름 파악	· 시스템·SW·암호자산·공급자 의존관계와 취약 암호 식별	연결형 인벤토리, 위험등록부
보호·전환	· 최소화·파기 및 전환 우선순위 결정	· 통신·저장·서명 영역별 PQC·하이브리드 적용과 복구시험	전환설계, 시험·복구계획
탐지·보증	· 통제의 점검기준·주기·예외 설정	· 개발·배포·운영환경에서 취약 암호와 통제 미작동 자동탐지	기계판독 증적, 커버리지·이력
대응·개선	· 정보주체 영향과 잔여위험 평가·보고	· 전환 실패·공급자 지연 대응, 재키잉·인증서 재발급	대응기록, 시정조치, 갱신된 평가

CPO의 핵심은 알고리즘 선택이 아니라 개인정보 위험의 판단기준을 제시하는 데 있다. 보호기간과 암호문 노출 가능성이 큰 데이터를 우선하고 최소화·파기 가능성을 검토한 뒤, 통신·저장·서명 영역별 목표기한을 정한다. 예외는 대상 개인정보, 보안통제, 책임자, 만료일과 재검토 조건을 갖춘 위험수용 결정으로 기록해야 한다.

CISO는 알고리즘·프로토콜·키 관리와 보안검증을, CIO·개발조직은 변경·상호운용성·복구시험을 담당한다. 조달·법무는 공급자의 PQC 지원일정, 취약점 통지, 업데이트·교체 책임과 증적 제공조건을 계약에 반영한다. 재래핑·재암호화와 하이브리드 적용처럼 장애 영향이 큰 변경에는 실패모드와 롤백 시험이 필요하다.

이 분담은 암호 민첩성으로 이어져야 한다. 키·인증서·HSM·프로토콜과 업무서비스의 의존관계를 관리하고 제품 종속을 피하도록 상호운용성과 교체 가능성을 검증한다. 시스템 도입·배포 때 인벤토리와 정책검사를 갱신하면 취약 암호의 재유입과 후퇴를 조기에 찾을 수 있다.

CPO는 개인정보-암호자산 연결률, PQC 적용률, 취약 암호 재발견률, 자동화 증적 커버리지, 예외 해소기간과 복구시험 완료율을 관리한다. 이 지표는 완료 선언이 아니라 잔여위험과 다음 개선순서를 결정하는 근거이며, 정기 보고에는 미달 원인·정보주체 영향·책임조직·해소기한을 함께 제시해야 한다.

6. 나가며

개인정보 양자위험은 미래의 양자컴퓨터가 가동되는 날이 아니라, 오래 보호해야 할 암호문이 오늘 복제돼 통제 밖으로 나가는 순간부터 시작된다. 데이터가 AES로 암호화돼도 키를 RSA·ECC로 보호하거나 전송 키 설정이 양자취약하면 약한 연결고리가 남는다. 따라서 보유기간뿐 아니라 암호문의 복제·노출 범위와 키 보호방식을 전환 우선순위에 반영해야 한다.

이를 위해 ISMS-P의 위험·자산·변경·공급자 관리와 점검·개선 절차에 PQC를 반영해야 한다. 개인정보·시스템·소프트웨어·공급자·암호 적용지점을 연결한 인벤토리를 구축하고, 취약 암호의 재유입과 PQC 적용상태를 주기적으로 확인한다. 모든 통제를 실시간 감시할 필요는 없지만, 자동화 가능한 항목은 주기와 판정기준을 정해 반복 검증해야 한다.

CPO는 예상 피해를 기준으로 목표 보호수준·전환순위·예외범위를 정하고, CISO·IT·개발·조달·법무는 기술전환·복구·공급자 의무를 이행한다. 적용률·인벤토리 연결률·취약 암호 재발견률·예외 해소기간으로 이행상태와 잔여위험을 판단하고 후속조치를 정해야 한다. 결과는 일회성 점검자료가 아니라 정기적이고 가능한 범위에서 자동화된 디지털 증적으로 축적돼야 한다.

결국 PQC 전환의 성패는 알고리즘의 이름이 아니라, 어떤 개인정보가 어디에서 어떤 암호에 의존하며 보호통제가 실제로 작동하는지를 지속적으로 파악·입증할 수 있는가에 달려 있다. PQC 전환은 일회성 기술교체가 아니라 암호환경 변화에 따라 위험을 재평가하고 통제를 검증·개선하는 지속적 개인정보 위험관리다.

참고문헌

- [1] European Commission, “Post-Quantum Cryptography,” updated 3 Sep. 2026.
- [2] NIST, Considerations for Achieving Crypto Agility: Strategies and Practices, NIST CSWP 39-upd1, 19 Dec. 2025 (updated 29 Jun. 2026).
- [3] 「개인정보 보호법」 제3조·제21조·제29조; 개인정보보호위원회 고시 「개인정보의 안전성 확보조치 기준」.
- [4] Michele Mosca, “Cybersecurity in an Era with Quantum Computers: Will We Be Ready?,” IEEE Security & Privacy, vol. 16, no. 5, pp. 38–41, Sep./Oct. 2018.
- [5] NIST, Transition to Post-Quantum Cryptography Standards, NIST IR 8547 Initial Public Draft, 12 Nov. 2024.
- [6] NIST, FIPS 203·204·205, 13 Aug. 2024; U.S. OMB, Migrating to Post-Quantum Cryptography, M-23-02, 18 Nov. 2022.
- [7] The White House, Executive Order 14412, 22 Jun. 2026; U.S. OMB, M-26-15, 24 Jun. 2026.
- [8] NIS Cooperation Group & European Commission, A Coordinated Implementation Roadmap for the Transition to Post-Quantum Cryptography, 23 Jun. 2025; Regulation (EU) 2016/679 (GDPR), Arts. 5, 25, 32.
- [9] 관계부처 합동, 「범국가 양자내성 암호체계 전환 종합 추진계획」, 2025.9.; 「범국가 양자내성암호 전환 마스터플랜」, 2023.
- [10] 「양자과학기술및양자산업육성에관한법률」 일부개정법률(2026.5.19. 공포, 2026.11.20. 시행) 제10조·제18조의2; 「개인정보보호법」 제33조.
- [11] 국가정보원·국가보안기술연구소, 「암호모듈 구현안내서(GVI) Part 1」, 2025.12.10., 부속서 C.5; 국가정보원, 「양자내성암호 검증대상 후보군 및 추진 일정 공지」, 2026.9.1.
- [12] 한국인터넷진흥원, 「정보보호및개인정보보호관리체계(ISMS-P) 인증기준」, ISMS-P 누리집(접속 2026.9.18.); 「개인정보 보호법」 제29조·제32조의2.
- [13] NIST NCCoE, Data Classification Practices, NIST SP 1800-39 Initial Public Draft, 12 Feb. 2026.
- [14] 개인정보보호위원회·과학기술정보통신부, 「정보보호및개인정보보호관리체계인증제실 효성강화방안」, 2026.4.10.
- [15] FedRAMP, “Approaching FedRAMP 20x Assessments,” FedRAMP Consolidated Rules for 2026, 2026.
- [16] 삼성SDS, 「삼성SDS, ‘양자내성암호’ 국내기술표준이끈다」(암호자산 인벤토리 도구 사례), 2025.3.11.

AI 학습용 합성데이터의 개인정보 보호 위험 분석과 안전한 활용 방향



안정연

KAIST 산업및시스템공학과 교수

1. 들어가며

AI 모형 개발에는 충분한 학습자료가 필요하지만, 그 자료에는 건강상태, 금융거래, 이동기록처럼 민감한 정보가 담기는 경우가 많다. 국내에서는 가명정보 제도가 도입된 뒤에도 보건의료와 금융 분야의 자료를 기관 밖으로 내보내는 절차가 까다롭고, 연구자가 실제 자료를 손에 쥐기까지 긴 시간이 걸린다. 합성데이터는 원본을 그대로 내보내지 않으면서 분석과 모형 개발을 이어 가려는 방법으로 논의되어 왔고, 개인정보보호위원회가 2024년 생성과 활용 절차를 정리한 안내서를 내놓으면서 국내 논의도 구체화되었다.^[1]

자료 접근의 제약은 AI 개발에만 걸리는 문제가 아니다. 정책을 결정할 때에는 여러 기관에 흩어진 자료를 함께 보아야 답할 수 있는 질문이 많다. 청년층의 학업 이력과 취업 성과를 연결하거나 만성질환의 의료 이용을 소득 수준과 함께 살피는 분석이 그런 예이지만, 실제로 기관 사이의 원본 공유는 거의 이루어지지 않는다. 법적 근거의 해석, 심의 절차의 부담, 사고가 났을 때의 책임이 겹치면 개별 기관으로서는 내보내지 않는 쪽이 언제나 안전한 선택이 되기 때문이다. 각 기관이 자기 자료의 합성본을 만들어 이를 연계하고 공유하는 방식은 이 교착을 푸는 현실적인 대안으로 제시되어 왔으며, 여러 부처의 종단자료를 합성해 연구자에게 개방한 사례도 보고되었다.^[4] 아래에서 살펴볼 한계가 분명함에도 합성데이터가 자료 공유 문제의 가장 유력한 해법으로 꼽히는 이유가 여기에 있다.^[10, 12]

합성데이터는 통계적 모형이나 생성 알고리즘으로 만들어 낸 자료이며, 실제 자료에서 관찰되는 분포와 변수 사이의 관계를 학습해 생성된다. 정형자료에서는 일부 변수만 새로 만드는 부분합성과 모든 변수를 새로 만드는 완전합성을 나눈다.^[13] 이 구분은 만드는 방식에 관한 것이지 안전하다는 판정이 아니다. 완전합성 자료에서도 원본에 관한 정보가 새어 나올 수 있다는 점은 여러 실증연구가 보여준다.^[16] 합성되었다는 사실 자체는 생성 방식에 대한 설명일 뿐이므로, 어떤 분석을 가능하게 하는지, 원본 접근을 얼마나 줄여 주는지, 개인정보 위험을 어느 정도 낮추는지를 함께 따져야 한다.

AI 개발에서 합성데이터의 가장 직접적인 이점은 원본을 들여다볼 일을 줄여 준다는 것이다. 개발자는 합성자료로 자료 구조를 파악하고 분석 코드를 작성하며 시스템이 제대로 작동하는지 시험할 수 있는데, 이 단계에서는 변수의 형식과 구조만 맞아도 쓸모가 있다. 예를 들어 상급종합

병원과 외부 연구자가 진료자료로 예측모형을 함께 개발한다면, 초기에는 합성자료로 분석 절차를 맞추고 마지막 단계에서 심의와 승인을 거쳐 실제 자료로 성능을 확인하는 방식을 택할 수 있다. 합성데이터를 원본의 대체물로 보기보다, 원본이 꼭 필요한 단계와 그렇지 않은 단계를 구별해내는 수단으로 쓰는 셈이다.

합성데이터는 드문 상황이나 특정 집단의 사례를 보강하는 데에도 쓰인다. 다만 학습 성능을 높이려고 비슷한 자료를 증강하는 것과, 개인의 정보가 드러나지 않도록 자료를 대체하는 것은 목표가 다르다. 앞의 경우는 특정 패턴을 충실히 되살리는 데 초점이 있지만, 뒤의 경우에는 그 재현이 특정인의 고유한 정보를 드러내지 않는지 확인해야 한다.^[16] 같은 생성모형을 쓰더라도 확인할 항목이 달라지므로, 무엇을 위해 만든 자료인지 처음부터 분명히 해 두어야 한다.

2. 합성데이터의 용도와 한계

합성데이터의 쓸모는 무엇을 하려는지에 따라 달라진다. 변수 하나하나의 평균과 분산을 잘 맞추더라도 변수들이 얹혀 있는 방식까지 살아남는다는 보장은 없다. 그래서 전반적인 분포가 얼마나 닮았는지를 보는 관점과, 관심 있는 분석의 결과가 원본에서의 결과와 얼마나 가까운지를 보는 관점을 나누어 평가한다.^[15] 전체 환자의 연령과 검사수치 분포를 잘 재현한 자료가 특정 연령대의 희귀질환을 가려내는 데에는 쓸모가 없을 수 있고, 거래금액 분포가 비슷해도 이상거래 직전의 흐름은 사라져 있을 수 있다. 분포가 닮았다는 결과 하나로 특정 과제에 쓸 수 있다고 말하기 어렵다.

통계적 결론을 이끌어내는 데 이용하려면 한 가지를 더 살펴야 한다. 합성자료에서 계산한 수치의 불확실성(오차)에는 원본 자체의 표본 오차뿐 아니라 생성모형이 원본을 잘못 읽어 낸 부분과 자료를 만들 때 개입한 우연이 함께 들어 있다. 이를 구분하지 않으면 결과가 실제보다 정밀한 것처럼 보고된다. 자료를 많이 만들어 낼수록 수치는 안정적으로 보이지만 원본이 담고 있던 정보가 늘어나는 것은 아니므로, 자료의 양과 정보의 양을 구분해 이용자에게 알려야 한다. 합성자료로 추정과 검정을 하려는 경우에는 같은 분석을 실제 자료에서 했을 때와 얼마나 어긋나는지를 미리 확인해 두는 절차가 필요하다.

희귀한 사례와 소수집단을 어떻게 다룰지도 까다롭다. 드문 특성을 가진 기록은 보호 측면에서 위험하지만 분석에서는 가장 중요한 기록인 경우가 많고, 사기 탐지나 희귀질환 연구처럼 예외를 잡아내는 것이 목적인 과제에서는 두 요구를 함께 만족시키기가 어렵다. 실제 관측치가 몇 건뿐인 집단을 바탕으로 자료를 대량으로 만들면 집단별 표본 수는 고르게 맞출 수 있지만, 몇 안 되는 특성이 반복해서 강조될 뿐이다. 집단별 숫자를 맞추는 일과 그 집단을 제대로 대표하는 일은 다르며, 특정 집단에서 오류가 몰린다면 활용 범위를 제한하거나 실제 자료를 더 확보하는 쪽을 검토해야 한다.

3. 개인정보 보호상 주요 위험과 보호효과의 한계

가장 먼저 꼽히는 위험은 생성모형이 학습자료를 외워 두었다가 그대로 내놓는 경우이다. 모형이

자료 전반의 규칙성을 배우는 데 그치지 않고 특정 기록에 지나치게 의존하면, 만들어진 자료에 원본과 같거나 거의 같은 내용이 섞여 들어간다. 대규모 언어모형에서는 질문을 던지는 것만으로 학습자료에 있던 문장 일부를 되살릴 수 있다는 점이 실제로 확인되었다.^[6] 정형자료에서 늘 같은 일이 일어난다는 뜻은 아니지만, 새로 만들어 낸 자료라는 이유만으로 노출 가능성을 접어 둘 수는 없다.

원본이 그대로 복제될 때만 문제가 되는 것도 아니다. 이름이나 연락처가 없더라도 드문 특성의 조합을 외부 정보와 맞춰 보면 특정 개인에 관한 사실을 알아낼 수 있다. 그래서 개인을 구별해 낼 위험, 다른 자료와 연결될 위험, 알려지지 않은 속성을 추정당할 위험을 함께 따지는 평가 틀이 제안되어 있다.^[8] 특정인의 자료가 학습에 쓰였는지를 알아내려는 공격도 있는데, 특정 질환 환자만으로 이루어진 자료라면 학습에 포함되었다는 사실 자체가 민감한 정보가 된다. 보호해야 할 대상이 이름과 같은 식별정보에 한정되지 않는 이유이다.^[16]

국내 실무는 가명처리와 익명처리를 다루면서 k-익명성 같은 기준에 익숙하다. 같은 특성을 가진 사람이 적어도 몇 명 이상 함께 묶여야 공개할 수 있다는 발상이다.^[17] 합성자료에는 원본 기록이 남아 있지 않으므로 이 기준을 그대로 적용할 수는 없지만, 이러한 접근방식 자체는 생성 단계로 옮겨 볼 만하다. 자료를 한 변수씩 차례로 만들어 가면서 지금까지 정해진 값들의 조합이 원본에서 충분한 수의 사람으로 뒷받침되는지 확인하고 그렇지 않으면 그 조합을 피하도록 하면, 안전성을 사후에 채점하는 대신 만드는 과정의 규칙으로 삼을 수 있다. 다만 이런 규칙은 드문 조합을 체계적으로 걸러내므로 앞 절에서 본 소수집단 문제와 바로 이어진다. 보호가 얼마나 강해졌는지와 쓸모가 얼마나 줄었는지를 같은 실험에서 함께 측정해야 한다.

보호와 쓸모가 언제나 반비례하는 것은 아니다. 많은 사람에게 공통으로 나타나는 규칙성을 살리는 일과 특정인의 기록을 사실상 되살리는 일은 다르기 때문이다. 보호 기술의 목표도 자료를 무조건 원본과 다르게 만드는 것이 아니라, 집단 수준의 정보는 남기면서 개인 단위의 노출만 줄이는 데 있다. 원본과 합성자료 사이의 최단 거리나 똑같은 기록의 비율은 유용한 진단이지만, 똑같은 기록이 없어도 정보는 추론될 수 있고 흔한 특성의 기록이 많았다는 이유만으로 노출을 단정하기도 어렵다. 변수가 많아지면 어떤 두 기록을 골라도 거리가 비슷해져서 이런 지표는 판별력을 잃는다.

차등정보보호는 많이 쓰이는 수학적 정보보호 방식으로, 어떤 한 사람의 자료가 들어 있는 빠져 있던 결과가 거의 달라지지 않도록 생성 과정에 잡음을 섞는 방식이며, 그 사람이 자료에 참여했기 때문에 새로 드러나는 정보를 정해진 범위 안에 묶어 준다.^[6] 그러나 이것이 아무것도 알 수 없게 만든다는 뜻은 아니다. 어떤 생활습관과 질환의 관계처럼 많은 사람에게 공통된 사실은 그대로 학습되고, 그 사실에 비추어 누군가의 상태를 짐작하는 일까지 막아 주지는 않는다. 보호하는 주체의 단위로 확인해야 한다. 한 사람이 수백 건의 거래기록을 남기는 자료에서 거래 한 건을 가려 주는 것과 그 사람의 소비 패턴 전체를 가려 주는 것은 전혀 다른 이야기이고, 앞의 방식이 훨씬 약하다.^[11] 그러므로 어떤 기법을 적용했다거나 수치를 얼마로 잡았다는 설명보다, 누구의 무엇을 어디까지 보호했는지를 밝히는 편이 훨씬 유용하다.

기술적으로 안전하다는 것과 법적으로 익명정보라는 것도 다른 말이다. 「개인정보 보호법」 제58조의2는 시간과 비용, 기술을 합리적으로 고려할 때 다른 정보를 함께 쓰더라도 더는 개인을 알아볼

수 없는 정보를 법 적용에서 제외한다.^[2] 판단의 잣대는 식별 가능성이지 완전합성이라는 이름이 아니다. 결과물의 법적 지위와 원본을 다루는 과정의 적법성도 따로 따져야 하며, 국내 안내서가 단계마다 준수사항을 점검하도록 구성된 것도 같은 이유이다.^[1] 살펴야 할 범위 역시 공개되는 자료에 그치지 않는다. 원본을 넘겨받는 단계, 생성모형을 학습하고 외부에 제공하는 단계, 만들어진 자료로 다시 AI를 개발하는 단계가 모두 들어가며, 생성모형에 얼마든지 질의할 수 있게 열어 두면 고정된 자료 한 벌을 내보낼 때와는 위험의 성격이 달라진다.^[5]

4. 안전한 활용을 위한 주요 고려사항

안전한 활용은 자료를 만들기 전에 용도와 공개 범위를 정하는 데에서 시작한다. 자료 구조 확인, 코드 시험, 예측모형 학습, 통계적 추론은 요구하는 재현 수준이 서로 다르고, 내부 개발자가 쓰는 경우와 외부에 제공하거나 일반에 공개하는 경우도 구분해야 한다. 개인정보보호위원회 안내서는 절차를 사전준비, 생성, 안전성 및 유용성 검증, 심의위원회 평가, 활용 및 안전한 관리로 제시하고, 널리 공개하는 자료일수록 안전성을 중심으로 검증하도록 강조한다.^[1] 실제 사업에 옮길 때에는 단계마다 책임자를 두고 만드는 쪽과 검증하는 쪽이 판단 근거를 공유하도록 기록해 두어야 한다.

생성방법은 자료의 성격과 목적에 맞추어 고르면 된다. 복잡한 모형을 썼다는 사실이 품질을 보증하지는 않으므로 단순한 방법과 성능을 견주어 보는 편이 낫고, 심층학습이 모든 자료에 최선은 아니라는 지적도 여러 평가지침에 나온다.^[11] 평가는 자료 자체의 품질과 목적별 성능을 나누어 설계한다.^[15] 앞쪽에는 변수의 범위와 결측 양상, 변수 사이의 관계가 들어가고, 뒤쪽에는 예측 오류와 희귀사건 탐지 능력, 집단별 성능처럼 실제 쓰임에 직결되는 항목이 들어간다. 한 가지 점수가 높다는 이유로 모든 용도에 적합하다고 판단하지 않도록 결과를 항목별로 제시하는 것이 바람직하다.

평가 수치를 어떻게 읽을지도 문제가 된다. 어떤 지표가 0.03이라는 결과는 그 자체로 좋지도 나쁘지도 않고 견줄 기준이 있어야 뜻이 생긴다. 한 가지 방법은 원본을 무작위로 둘로 나눈 뒤 그 두 덩어리 사이에서 같은 지표를 계산해 보는 것이다. 같은 자료에서 나왔는데도 생기는 차이의 폭을 알면, 합성자료에서 나온 값이 그 범위 안에 드는지로 판단할 수 있다. 원본을 생성용, 평가용, 기준용 세 몫으로 나누어 두면 쓸모에 관한 지표와 위험에 관한 지표를 같은 기준 위에서 읽을 수 있고, 서로 다른 생성기법이나 서로 다른 자료에서 나온 숫자를 견주는 일도 가능해진다.

더 어려운 문제는 유용성과 정보보호를 동시에 판단하는 일이다. 이 두 가지 개념은 측정 대상도 스케일도 다르며 일반적으로 한쪽을 올리면 다른 쪽이 내려가는 상충적인 관계여서, 단지 점수를 더하거나 가중평균해서는 의미 있는 결론이 나오지 않는다. 어느 정도의 쓸모면 충분한지, 어느 정도의 공격 성공률이면 받아들일 수 있는지도 자료와 용도에 따라 달라지는데 이를 정해 줄 기준은 아직 없다. 여러 방법을 같은 잣대로 평가하고 비교할 표준이 마련되어 있지 않다는 지적은 관련 분야의 검토 논문에서 거듭 알려진 사실이다.^[9] 국내에서도 유용성 지표와 노출 위험 지표를 함께 적용하여 여러 재현자료 생성기법을 비교한 연구가 이루어졌다.^[3] 당분간은 종합점수 하나를 만들기보다 항목을 나누어 보여 주고 어떤 조건에서 어떤 결과가 나왔는지 함께 밝히는 편이 현실적이다.

AI 학습에 쓸 때에는 합성자료로 학습한 모형을 따로 확보한 실제 자료에서 평가하는 설계를 먼저 고려할 만하다. 이때 평가용 자료는 생성모형을 만들거나 손보는 과정에서 완전히 떼어 두어야 하며, 한 사람에게 여러 기록이 있다면 기록 단위로 무작위 분할하는 것만으로는 같은 사람의 정보가 학습과 평가에 함께 들어가는 것을 막지 못한다. 실제 자료만 쓴 모형, 합성자료만 쓴 모형, 둘을 섞은 모형을 같은 조건에서 비교하면 합성자료가 무엇을 보태고 무엇을 잃게 하는지가 분명해진다.

위험 점검에서는 공격자가 무엇을 알고 있다고 볼지부터 정해야 한다. 합성자료만 내려받을 수 있는지, 생성모형에 질의할 수 있는지에 따라 따져야 할 내용이 달라지며, 학습 참여 여부를 알아 내려는 시도와 속성을 추정하려는 시도, 다른 자료와 연결하려는 시도 가운데 용도에 맞는 상황을 골라 시험한다. 결과를 읽을 때에는 비교 기준을 함께 보아야 한다. 점검 대상 대부분이 학습에 참여하지 않은 사람이라면 모두 비참여자라고 답하는 것만으로도 정확도가 높게 나오므로, 그런 수치는 공격이 통했다는 증거가 되지 못한다. 반대로 여러 방식과 여러 하위집단에 대해 지표를 계속 계산하다 보면 실제로는 위험이 없어도 몇몇 조합에서 큰 값이 우연히 나타나므로, 몇 개의 지표를 몇 번 계산했는지를 함께 밝혀야 한다.

원본에 거듭 접근하고 그 결과를 계속 내보내는 상황도 관리 대상이다. 이미 학습을 마친 모형에서 자료를 더 뽑아내는 일 자체는 원본을 다시 들여다보는 것이 아니므로 새로운 위험을 만들지 않지만, 원본으로 모형을 다시 학습하거나 조정해 그 결과를 공개하는 일은 매번 따로 점검해야 한다.^[6] 위험이 남아 있는 자료라면 그대로 공개하기보다 적법한 근거 아래 이용 환경을 제한하는 편이 낫다. 초기 개발은 합성자료로 하고 중요한 검증만 보호된 환경에서 실제 자료로 수행하는 단계적 운영이 현실적이며, 외부 개발자가 합성자료로 코드를 작성하면 원본 보유기관이 이를 실제 자료에 돌려 필요한 결과만 돌려주는 방식도 가능하다.

끝으로 합성자료에 붙는 설명자료의 수준을 높여야 한다. 무엇을 위해 어떤 원본에서 어떻게 만들었는지, 어떤 평가를 거쳤고 어떤 용도까지 확인되었는지를 적어 두고, 개인정보가 없는 안전한 데이터라는 문구 대신 어떤 조건에서 어떤 결과를 얻었는지를 밝히는 편이 유용하다. 기관 사이에 자료를 주고받을 때에는 이런 기록이 곧 신뢰의 근거가 된다.

5. 향후 발전방향

앞으로의 과제는 얼마나 많이 만들어 내느냐가 아니라 용도별 검증체계를 얼마나 촘촘히 갖추느냐에 있다. 보호 수준과 쓸모를 평가하는 공통 원칙은 두되 모든 자료에 같은 지표와 임계값을 들이대서는 안 된다. 의료자료와 금융자료에서 중요한 위험이 다르므로, 공통 절차와 분야별 방법을 구분하고 공개 범위에 따라 요구하는 근거의 수준을 달리 정해야 한다. 평가의 초점도 평균적인 성능에서 취약한 사례와 집단 쪽으로 옮겨 가야 하며, 전체 위험이 낮다는 결과만으로 모든 사람이 보호된다고 말할 수는 없다.

연구가 필요한 지점은 크게 세 가지이다. 첫째는 합성 과정에서 생기는 불확실성을 제대로 반영해 분석 결과의 신뢰성을 계산하는 방법이고, 둘째는 쓸모와 위험을 같은 기준 위에서 읽어 하나의 판단으로 묶는 평가 틀이며, 셋째는 변수가 많은 자료에서 생성과 평가의 성질이 어떻게 달라지는 지 밝히는 일이다. 합성자료가 다시 학습자료로 흘러드는 상황도 살펴야 한다. 만들어진 자료만으로 반복해서 학습하면 모형이 다루는 범위가 점점 좁아진다는 결과가 보고되었고, 실제 자료를 계속 쌓으면서 합성자료를 함께 쓰면 그런 악화가 완화된다는 분석도 있다.^[7, 14]

보호의 측면에서는 수학적 보장과 실제 공격 실험을 함께 써야 한다. 수학적 개념은 무엇을 어디까지 보호하는지 분명히 해 주지만 구현과 운영이 그 조건을 지키는지는 따로 확인해야 하고, 시험한 공격이 실패했다는 사실을 무제한의 안전으로 읽어서도 안 된다. 제도 역시 어떤 기술을 썼는지 확인하는 데 그치지 말고 보호의 대상과 조건, 검증의 한계를 설명하도록 요구하는 편이 낫다. 원본 보유기관과 생성기관, AI 개발기관, 최종 이용기관이 서로 다르다면 누가 적법성을 확인하고 누가 안전성을 검증하며 누가 활용 범위를 관리하는지도 미리 정해 두어야 한다.

합성데이터가 맞지 않는 상황을 인정하는 것도 안전한 활용의 일부이다. 쓸모와 보호를 함께 충족하기 어렵다면 보호된 분석환경에서 원본을 쓰는 편이 낫고, 원본의 한계나 생성모형의 오류가 합성이라는 절차로 사라지지도 않는다. 그럼에도 기관 사이의 원본 공유가 사실상 이루어지지 않는 현실에서, 각 기관의 자료를 합성해 연계하고 공개 수준을 조절하는 방식은 지금 택할 수 있는 가장 실효성 있는 접근이다.^[10, 12] 남은 과제는 이 접근을 접는 것이 아니라 생성과 평가, 활용과 사후 관리를 잇는 검증체계를 갖추는 일이다. 궁극적인 판단기준은 실제 자료를 합성자료로 바꾸었는가 아니라, 어떤 정보를 지키고 어떤 위험을 줄였으며 그 결과를 어떤 조건에서 확인하였는가여야 한다.

참고문헌

- [1] 개인정보보호위원회. (2024). 『합성데이터 생성·활용 안내서』. 개인정보보호위원회.
- [2] 대한민국. 「개인정보 보호법」, 제58조의2. 국가법령정보센터.
- [3] 안성빈, 외. (2023). 유용성과 노출 위험성 지표를 이용한 재현자료 기법 비교 연구. *응용통계연구*, 36(2), 141–166.
- [4] Bonnéry, D., et al. (2019). The promise and limitations of synthetic data as a strategy to expand access to state-level multi-agency longitudinal data. *Journal of Research on Educational Effectiveness*, 12(4), 616–647.
- [5] Carlini, N., et al. (2021). Extracting training data from large language models. 30th USENIX Security Symposium, 2633–2650.
- [6] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407.
- [7] Gerstgrasser, M., et al. (2024). Is model collapse inevitable? Breaking the curse of recursion by accumulating real and synthetic data. ICML 2024 FM-Wild Workshop
- [8] Giomi, M., Boenisch, F., Wehmeyer, C., & Tasnádi, B. (2023). A unified framework for quantifying privacy risk in synthetic data. *Proceedings on Privacy Enhancing Technologies*, 2023(2), 312–328.
- [9] Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. (2022). Synthetic data generation for tabular health records: A systematic review. *Neurocomputing*, 493, 28–45.
- [10] Jordon, J., et al. (2022). Synthetic data—what, why and how? *arXiv:2205.03257*.
- [11] Near, J. P., Darais, D., Lefkowitz, N., & Howarth, G. S. (2025). Guidelines for evaluating differential privacy guarantees. NIST Special Publication 800–226.
- [12] Raghunathan, T. E. (2021). Synthetic data. *Annual Review of Statistics and Its Application*, 8, 129–140.
- [13] Reiter, J. P. (2003). Inference for partially synthetic, public use microdata sets. *Survey Methodology*, 29(2), 181–188.
- [14] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631, 755–759.
- [15] Snoke, J., Raab, G. M., Nowok, B., Dibben, C., & Slavković, A. (2018). General and specific utility measures for synthetic data. *Journal of the Royal Statistical Society: Series A*, 181(3), 663–688.
- [16] Stadler, T., Oprisanu, B., & Troncoso, C. (2022). Synthetic data—Anonymisation groundhog day. 31st USENIX Security Symposium, 1451–1468.
- [17] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 557–570.

Privacy Report

2026 개인정보 이슈 심층 분석 보고서

『2026 개인정보 이슈 심층 분석 보고서』는
개인정보보호위원회 출연금으로 수행한
사업의 결과물입니다.

한국인터넷진흥원의 승인 없이 본 보고서의 무단전재나
복제를 금하며, 인용 출처 『2026 개인정보 이슈 심층 분석
보고서』를 밝혀주시기 바랍니다.

본 보고서의 내용은
한국인터넷진흥원의 공식 견해가 아님을 알려드립니다.

발행

발행일	2026년 10월
발행처	한국인터넷진흥원 개인정보예방팀 전라남도 나주시 진흥길 9

Privacy Report

2026 개인정보 이슈
심층 분석 보고서

2026 Vol. 1