개인정보 보호 강화기술(PET)의 개념 및 사례 동향

송은지 한국인터넷진흥원 책임연구원

1. 머리말

생성형 AI의 등장으로 AI 광풍이 불며 LLM(거대언어모델, Large Language Model)이 주목받고 있다. LLM은 자연어 안에서 문법, 구문, 단어 등에 대한 규칙성을 찾아내는 AI 모델 중 하나이며, 거대한 데이터를 그 기반으로 한다. AI 기술의 정확도를 높이기 위해선 데이터 학습이 중요한데, 이 학습 데이터에는 미처 거르지 못한 개인정보가 포함될 수 있다. SNS나 포털에 공개된 개인정보 등이 여기 해당한다. 기업이 이러한 고객의 데이터를 활용해 데이터베이스를 구축하는 경우, 개인정보가 필연적으로 포함된다.

개인정보 컴플라이언스는 보편적인 국제 규제이자 소비자의 요구사항이다. 반면 기업 입장에서 개인정보는 보호 대상인 동시에 최대한 효과적으로 활용해야 하는 자원이기도 하다. 이런 맥락에서 현재 개인정보를 보호하면서도 그 효용을 보장하는 PET(개인정보 보호 강화기술, Privacy-Enhanced Technology 혹은 Privacy-Enhancing Technology로 해외에서는 혼용해 표기) 활용이 주목받고 있다. 그러나 초기 단계이기에, 글로벌 기업을 중심으로 점차 적용이 이뤄지고 있는 상황이다. 기술의 효용에 비해 그 중요성에 대한 인식 역시 아직 낮은 편이다.

이번 원고에선 해외 최근 자료들을 기반으로 PET의 개념과 동향을 알아보고, 실제 기업과 주요국 공공 분야 등에서 적용하고 있는 사례를 살펴본다. 이는 향후 PET 확산을 위해 나아가야 할 방향을 수립하고, 과제를 산출하는 데 참고할 수 있을 것이다.

2. PET의 개념

2023년 6월 ICO(영국 개인정보감독기구, Information Commissioner's Office)는 금융, 의료, 리서치, 정부 등 각 부문 정보보호책임자(DPO)를 대상으로 'PET 가이드라인(Privacy-Enhanced Technology guideline)'을 발표했다. 본 가이드라인은 PET에 대해 '개인정보 보호를 위한 방안을 모두 포괄하며, 데이터 보호 원칙을 구현하는 수단'이라고 정의했다. 또한 OECD(경제협력개발기구, Organisation for Economic Co-operation and Development)에서 2023년 2월에 발표한 '개인정보 보호 강화 기술의 등장(EMERGING PRIVACY ENHANCING TECHNOLOGIES)' 보고서에 따르면, PET는 '개인정보의 기밀성을 보호하면서 정보를 수집, 처리, 분석, 공유할 수 있는 디지털 기술과 접근 방식'을 총칭한다.

즉 PET는 '특정 개인정보 혹은 데이터 보호 기능을 달성하고, 더불어 개인 또는 자연인 그룹의 개인정보를 위협으로부터 보호하기 위한' 기술 프로세스, 방법, 지식을 모두 포괄하는 개념이라

볼 수 있다. 예를 들어, 목적 범위 내 정보만 최소한으로 처리하기, 개인정보 익명화 또는 가명화 솔루션, 접근권한 통제처럼 데이터 보호법에서 요구하는 기술적인 조치들이 있다. 앞서 언급한 OECD 보고서는 학술 기관 등의 연구를 기초로 PET를 <표 1>과 같이 분류하고 있다. 크게 ① 데이터 난독 처리 도구, ② 암호화된 개인정보 처리, ③ 연합 및 분석, ④ 데이터 책임 도구 등 네 가지 범주로 나눈 것이다.

<표 1> PET의 주요 유형

구분	PET 도구	설명
데이터 난독 처리 도구	차분 프라이버시	개인과 연결된 데이터에 무작위성을 부여하거나 노이즈를 적용해 재식별 가능성을 낮춤
	합성 데이터 생성(SDG)	기존 지식을 사용해 완전히 새로운 데이터를 생성
	영지식 증명	정보를 노출하지 않고 진실 여부 검증
암호화된 개인정보 처리	동종 암호화(HE)	일반 텍스트를 공개하지 않고 암호화된 데이터의 연산 수행
	신원 기반 암호화(IBE)	전통적인 공개키 인프라 대신 개인키 생성을 통해 발신자에서 수신자 방향의 메시지에 암호화 적용
	안전한 다자 연산(SMPC)	분산 컴퓨팅을 수행하면서 정확성과 최소한의 입력 및 출력 학습을 우선시해 연산 과정 보호
	신뢰받는 실행 환경	데이터의 기밀성을 훼손하지 않고 암호화된 키와 민감 데이터를 평문으로 안전하게 접근할 수 있게 함
연합 및 분산 분석	연합학습(FL)	개별 엔드포인트가 기계학습 모델 훈련에 참여하면서 학습 데이터를 기기에 유지하고 요약 데이터만 중앙 데이터 저장소에 전송할 수 있도록 허용하는 기술
	분산 분석	프라이버시를 보호하는 기계학습
데이터 책임 도구	책임 시스템	데이터에 접근할 수 있는 시기에 대한 규칙 설정 및 집행
	개인정보 관리 시스템	정보주체에게 자신의 개인정보에 대한 통제권 제공

※출처: OECD, EMERGING PRIVACY ENHANCING TECHNOLOGIES-OECD DIGITAL ECONOMY PAPERS, 2023. pp.15

첫 번째 데이터 난독 처리 도구는 노이즈를 추가하거나 상세 식별 정보를 제거하는 것이다. 이를 통해 데이터를 변형시켜 읽을 수 없게 만든다. 대표적인 예로 차분 프라이버시, 합성 데이터, 영지식 증명(ZKP) 등이 있다. 단점으론 개인정보가 재식별될 수 있다는 문제와 함께, 현재 기술 수준과 역량이 부족한 것이 꼽힌다. 이때문에 유망한 분야이지만, 실제 사용 사례는 아직 적다.

두 번째는 암호화된 개인정보 처리 도구다. 개인정보 처리는 그간 보안상 가장 큰 취약점이었는데, 이는 암호화하더라도 데이터를 처리할 때는 복호화가 필수였기 때문이다. 최근 기술이 발전하면서 데이터를 활용하는 동안에도 암호화된 상태를 유지할 수 있도록 패러다임이 변했다. 암호화된 개인정보 처리 도구의 예로는 동종 암호화(HE), 다자간 연산(SMPC), 신뢰할 수 있는 실행 환경 등이 있다. 그러나 암호화 처리도 단점이 있다. 암호화된 데이터의 연산 비용이 일반적인 경우보다 훨씬 높아 비효율적이라는 것이다. 또한 데이터가 유출되지 않는 것을 보장하지는 못한다는 약점이 있다.

세 번째는 연합 및 분산 분석이다. 이 기술은 작업을 실행하는 자가 접근할 수 없는 데이터를 분석할 수 있도록 해준다. 작업 실행자에겐 통계와 결과만 전달되는 방식이다. 대표적인 연합학습은 원시 데이터가 전처리돼 처리자에게는 결과만 전달되고 유사 데이터와 결합된다. 따라서 데이터를 처리하는 과정에서 리스크가 상당 부분 줄어든다. 반면 분산 분석은 여러 노드에 걸쳐

분석을 분산시키는 방식이다. 연합 학습과 마찬가지로 이 접근 방식은 처리자가 데이터에 직접 접근하는 것을 허용하지 않는다는 공통점이 있다. 그러나 연합 및 분산 분석도 여전히 한계는 있다. 여전히 정보가 유출될 가능성이 있다는 것과 안정적인 연결이 필수라는 점이다. 연구자들은 정보 유출의 문제를 해결하기 위해 암호화된 개인정보 처리 기법을 병행해 사용할 것을 권하고 있다.

3. 민간 분야의 적용 사례

PET가 부상하면서 실제 현업에서 기술을 적용하는 사례들이 글로벌 기업을 중심으로 나타나고 있다. 기업 입장에선 기술 적용을 통해 '해당 기업이 개인정보 보호 설계를 적용한다'는 사실을 고객에게 입증할 수 있다. 이는 유럽 GDPR(일반개인정보보호법, General Data Protection Regulation)과 같은 글로벌 규정 준수에도 도움이 된다.

<표 2>는 애플(Apple)과 구글(Google), 페이스북(Facebook), IBM 등 글로벌 ICT 기업에서 PET를 적용한 실제 사례들이다.

<표 2> ICT기업들의 PET 적용 사례

기업명	PET 도구	설명
APPLE	차분 프라이버시	메시지 앱, 검색어 추천에 적용(2016년~)
Google	연합학습	키보드 앱에서 개인정보 전송 기능에 적용(2020년)
Facebook	다자간 계산	광고 및 마케팅에 활용할 개인정보 수집
IBM	-	광고 식별자 대안 개발을 위해 PET 실무 그룹 운영(2022년~)

출처: 각종 해외 언론보도 요약

먼저, 애플은 2016년 "이용자의 프라이버시를 침해하지 않고 행동 패턴을 파악하는 기술을 도입한다"며 차등 프라이버시 기술 도입을 발표했다. 해당 기술은 ios 10 업데이트 버전에 포함됐으며, 애플은 이에 대해 "자사 소프트웨어 엔지니어들로 하여금 '불특정 다수 이용자가 아이폰, 아이패드 등 모바일 기기를 어떻게 사용하는지 파악'하는 데 도움이 될 것"이라고 평가했다. 애플은 이 외에도 iCloud에 저장하는 데이터를 암호화하고 iMessage에서도 권한이 없는자의 접근을 어렵게 하는 새로운 보안 기능을 출시하기도 했다.

구글은 키보드에 AI 기능을 탑재하면서 연합학습 기술을 적용했다. 이는 평소 이용자의 메시지 내용과 습관을 분석해 적절한 답장을 보낼 수 있도록 자동으로 추천해 주는, '스마트 답장' 기능을 지원하기 위한 것이다. 제대로 된 추천이 이뤄지려면, 이용자의 내밀한 메시지 내용이나 키보드를 입력하는 습관 등 각종 개인정보를 수집하고 학습해야 한다. 연합학습을 통해 사용자 기기의 알고리즘이 개인정보 등을 학습해 모델을 구축하면, 이를 서버로 전송한다. 서버에선 여러 스마트폰에서 전송한 모델을 분석하고 업데이트해 스마트폰으로 배포하는 방식을 활용했다. 이는 개인정보가 아니라 모델만 공유하기 때문에 개인정보 유출 가능성과 리스크가 적다는 장점이 있다.

페이스북의 경우, 다자간 연산 방식으로 PET를 적용했다. 페이스북은 맞춤형 마케팅 분야에서 다자간 연산 방식을 사용해 개인정보 보호 조치를 적용하고 있다. 이를 통해 광고주에게 결과를

제공하면서 개인정보는 제한하고 있는 것이다. 메타(Meta)로 사명이 변경된 이후에도 PET에 대한 투자 의지는 이어지고 있다. 개인정보 광고 분야 부사장인 데니스 부크하임(Dennis Buchheim)은 이에 대해 "메타는 PET가 디지털 광고의 다음 시대를 지원하는 혁신이라고 믿는다"고 밝혔다.

IBM 역시 맞춤형 광고와 관련해 연구를 진행하고 있다. IBM은 디지털 광고 기술 표준 제정 기관으로서, 2022년 2월 PET 워킹그룹을 신설했다. 이는 웹상의 쿠키와 기타 식별자에 대한 대안 개발을 목적으로, 새로운 표준과 기술의 필요성이 높아졌기 때문이다. 해당 워킹그룹의 목표는 고급 암호화 작업, 데이터 연구자, 개인정보보호와 보안 시스템 전문가 등이 모여 디지털 광고 산업을 위한 도구를 개발하는 것이다. IBM에선 PET가 데이터 보안을 극대화하고 개인정보 활용을 최소화하는 대표적인 수단이라고 인식하고 있다. 더불어 광고 산업에서도 PET가 이용자의 개인정보 보호와 보안 유지를 위한 유용한 수단이라는 판단이다.

4. 공공 분야의 적용 사례

주요국 공공 분야에서도 PET에 투자하고 이를 활용하는 움직임이 활발하게 나타나기 시작했다. NSF(미국 국립과학재단, National Science Foundation)와 DARPA(방위고등연구계획국, Defense Advanced Research Projects Agency)는 동형암호와 차분 프라이버시 기술 개발을 위한 프로젝트를 수행하고 있다. 또한, CEDI(영국 데이터윤리 혁신센터, Centre for Data Ethics and Innovation)는 우수사례 발굴 및 홍보 등 산업계의 PET 적용을 적극 추진하고 있다. NHS(영국 국민보건서비스, National Health Service)에선 PET 지원 시스템 구축을 위해 무려 3,500만 파운드(약 610억 원)의 대규모 예산을 투자했다[9]. 이는 민감정보인 의료정보보호를 위한 것이다. 이에 더해, 미국과 영국은 금융범죄 근절과 해결을 위한 해결 방안으로서, 2022년부터 'PET 챌린지'를 개최해 수상하고 상금을 수여하는 등 다양한 방식으로 관련 기술 개발과 인재 양성을 장려하고 있다.

한편 공공 분야에서 PET 적용이 가장 두드러지고 활발한 분야는 통계와 데이터 가공이다. PET가 개인정보가 포함된 통계 분석을 가능케 하는 동시에, 민감정보의 기밀성을 보호하는 방안으로 유용하기 때문이다. <표 3>은 주요국 공공 분야에서 해당 기술을 사용해 데이터를 처리한 사례들이다. 서로 다른 부문에 걸쳐 단일기술 또는 여러 방식을 조합한 기술을 활용한 것이다.

<표 3>을 통해 유럽, 미국, 캐나다처럼 현재 개인정보 활용 방안 논의를 선도하는 국가 중심으로 PET가 활용되고 있음을 알 수 있다. 또한 데이터 처리와 관련되므로, 통계기관의 사례가 많은 것도 특징이다. 한편 네덜란드나 UN(국제연합, United Nations)의 예처럼 AI 학습을 위한 데이터 생성 방안으로 기술 적용과 활용이 확대되고 있는 추세도 볼 수 있다.

PET는 데이터 셋의 개인정보를 손상시키지 않고 인사이트를 얻어낼 수 있다는 장점을 갖고 있다. 추가로 민감한 개인정보에도 접근이 가능하다는 강력한 장점도 있다. 그러나 PET가 데이터 처리의 보안 문제를 완전하게 해결한다고 볼 수는 없다. 이 때문에 합법적이고 공정하며 투명한 처리를 하기 위한 노력이 지속적으로 필요할 것이다. 기술 적용을 고려하기 전 먼저 목적을 분명하게 확인하고, 기술 적용 후의 영향을 평가해 봐야 한다. 또한 개인정보 보호 원칙에 입각해 문제 가능성이 없는지 반드시 고려해야 한다.

<표 3> 주요국 공공기관들의 PET 적용 사례

PET	주체	목적	데이터
안전한 다자 연산 (SMPC)	보스턴 여성노동자협의회	안전한 다자간 계산을 활용해 급여 격차 측정	보스턴의 성별 및 인종 간 임금 격차
	유럽 통계 시스템	스마트 설문조사 기술 개발	참여자의 기기에서 수집된 센서 데이터
	미국 교육부	개인정보보호 기록 연계를 이용한 학생 재정지원 데이터 분석	학부생의 평균 학자금 대출 및 보조금 데이터
신뢰받는 실행 환경	Eurostat	모바일 네트워크 사업자의 데이터 처리	통화기록 및 가입자의 방문 위치 등
	인도네시아 관광부	두 모바일 사업자의 데이터를 공유 및 결합해 통계 생성	망사업자의 IMSI 목록
동종 암호화	캐나다 통계청	단계적 동종 암호화를 통해 기계학습을 위한 개인정보 분류	개인정보 텍스트
합성 데이터	캐나다 통계청	합성 데이터 활용 테스트	교육 및 해커톤을 위한 고품질 데이터
차분 프라이버시	미국 인구조사국	인구조사에서 수집한 민감정보의 노출 방지	미국 인구조사 관련 데이터
안전한 다자 연산, 동종 암호화, 차분 프라이버시	한국 통계청	개인정보보호 통계 데이터 허브 플랫폼 개발	다양한 종류의 통계 데이터
안전한 다자 연산, 동종 암호화, 연합학습	이탈리아 통계청 및 은행	양 기관의 개인정보를 연결해 데이터 분석	인구통계 및 금융 데이터 셋
	네덜란드 통계청	분산된 임상 및 사회경제 데이터에서 개인정보보호 심혈관 위험 예측 모델 개발	1차 및 2차 병원 진료 데이터, 사회경제적 데이터 셋
연합 학습, 동종 암호화, 차분 프라이버시	UN 유럽 경제위원회	스마트 기기에서 수집한 라이프스타일 데이터를 통한 기계학습 모델 개발	스마트장치에서 수집된 데이터 셋
안전한 다자연산, 차분 프라이버시	UN PET Lab	여러 국가 통계청에서 수집된 데이터의 분석	UN Comtrade 데이터 셋에서 가져온 데이터

※출처: THE UNITED NATIONS GUIDE ON PRIVACY-ENHANCING TECHNOLOGIES FOR OFFICIAL STATISTICS, THE PET GUIDE, 2023. pp.61-109 요약

5. 맺음말

PET의 개념과 기술 종류, 기업과 공공 분야의 대표적인 사례들을 알아봤다. PET는 AI 기술이 발전하고 주류로 자리매김하면서, 안전한 개인정보 활용과 데이터 전처리를 위한 핵심 요소가 되고 있다. 향후 데이터 대부분이 개인정보이거나 개인정보가 될 가능성이 높은 상황이며, 이 때문에 PET는 더 유망한 분야로 주목받을 것이라 판단된다.

마지막으로 최근 PET 기술 동향과 몇 가지 두드러지는 공통 특징을 꼽아보도록 하겠다. 첫 번째, PET는 개인정보 보호와 활용에 유용한 기술로서 대부분 개인정보 처리가 포함된다는 것이다. PET의 대표적인 데이터 난독화나 암호화, 분석 방안 등은 모두 개인정보의 식별성을 감소시키는 수단이다. 다만 모든 PET가 개인정보를 익명화하는 것은 아니며, PET를 통하지 않고도 익명화가 가능하다는 점을 주의해야 한다.

두 번째, PET는 특정한 기술만을 뜻하는 것이 아니라 기술 프로세스와 방법 또는 지식을 포괄하는 추상적 개념이다. 그 목적은 개인정보 또는 데이터 보호 기능을 달성하거나, 개인 또는 자연인 그룹의 개인정보를 위협으로부터 보호하는 것이다. 따라서 기술이 발전하고 변화함에 따라

구체적인 방안도 진화할 수 있으며, 현재 대표적인 기술들도 좀 더 정교하게 적용하기 위한 개선이 필요할 것이다. 본문에서 제시한 것처럼 PET의 구체적인 사용 사례를 분석하고, 보완점과 이점을 도출해 공유하는 작업이 지속적으로 요구된다.

세 번째, PET를 활용한다고 해서 완전한 안전을 보장하는 것은 아니라는 점이다. 또한, PET는 글로벌 규제들의 준수를 보장할 수 없고, 대체할 수도 없다. 이용자의 개인정보와 권리를 보호할 목적으로 활용하되, 규제와 법을 준수하는 방식 중 하나일 뿐이다. 따라서 PET 분야의 실효성을 높이기 위한 지속적인 연구개발 노력과 투자가 필수라고 생각된다.

그 잠재력과 중요성에도 불구하고, PET는 아직 한정된 분야에서만 적용되고 있고 대중적이지 못한 측면이 있다. 여전히 한정된 수의 개인정보 처리 사례를 제외하고는, "기술의 성숙도 수준이 부족하다"는 지적이 제기되고 있기 때문이다. 따라서 기술이 성숙되기까지, 현 단계에선 "PET가 개인정보의 안전한 활용을 보장하고, 개인의 권리를 증진한다"는 것을 입증해야만 한다. 관련 사례 발굴과 확산이 중요한 이유다.

※ 본 연구는 정보통신기획평가원의 주간기술동향 2127호 ICT신기술 '개인정보 보호 강화기술(PET)의 개념 및 사례 동향'(2024.3월)을 기반으로 작성되었다.

[참고문헌]

[1] ICO, Privacy-enhanced technologies guideline, 2023. pp.17

[2] OECD, EMERGING PRIVACY ENHANCING TECHNOLOGIES-OECD DIGITAL ECONOMY PAPERS, 2023. pp.15

[4] INDEPENDENT, Apple introduces a range of new security features to stop attacks on iPhone users, 2022. 12. 8.

[5] 보안뉴스, 구글 키보드 AI 기능, 개인정보 유출 우려 어떻게 줄였을까, 2020. 10. 13.

[6] Raconteur, PETs: the tech balancing data insights and customer privacy, 2023. 2. 1.

[7] Meta(페이스북), 개인정보 보호 강화기술과 더 나은 미래를 위한 준비, 그레이엄 머드, 2021. 8. 11.

[8] PR Newswire, IAB Tech Lab Launches Privacy Enhancing Technologies(PETs) Working Group to Support the Creation of Long-Term Sustainable Solutions to Privacy, 2022. 2. 7.

[9] digital health, NHS England launches tender for new privacy enhancing technology, 2023. 6. 28.

[10] GOV.UK, Press release, Winners announced in first phase of UK-U.S. privacy-enhancing technologies prize challenges, 2022. 11. 10.

[11] THE UNITED NATIONS GUIDE ON PRIVACY-ENHANCING TECHNOLOGIES FOR OFFICIAL STATISTICS, THE PET GUIDE, 2023. pp.61-109 요약

※ 출처: TTA 저널 제213호