경북대학교 법학연구원『법학논고』
제94집 (2026.07) 179~207면.
Kyungpook Natl. Univ. Law Journal
Vol.94 (July 2026) pp.179~207.
http://doi.org/10.17248/knulaw..94.202607.179

공공데이터 개방과 생성형 AI의 개인정보 재식별 위험에 관한 연구*

- AI 데이터 영향평가 제도의 도입방안을 중심으로 -

박 정 인** · 정 현***

〈국문초록〉

생성형 인공지능(Generative AI)의 급속한 발전은 공공데이터의 활용 가치를 크게 증대시키는 한편, 기존 개인정보 보호체계가 충분히 대응하지 못한 새로운 법적 위험을 초래하고 있다. 특히 공공데이터가 생성형 AI의 학습데이터로 활용되는 과정에서 가명정보 또는 비식별정보가 다른 데이터와 결합되거나 인공지능의 추론 기능을 통하여 특정 개인이나 조직을 다시 식별할 수 있는 재식별(Re-identification) 위험이 현실적인 문제로 대두되고 있다. 현행 「개인정보 보호법」상의 개인정보 영향평가는 개인정보파일의 안전성 확보를 중심으로 설계되어 있으며, 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」상의 고영향 인공지능 영향평가 역시 AI 시스템이 인간의 권리와 자유에 미치는 영향을 중심으로 규율하고 있어, 생성형 AI 환경에서 발생하는 데이터 결합, 재식별 및 추론 위험을 충분히 평가하기에는 한계가 있다. 본 연구는 공공데이터 개방과 생성형 AI 활용 과정에서 발생하는 개인정보 재식별 위험의 특성과 법적 문제를 분석하고, 현행 영향평가 제도의 한계를 검토한 후 새로운 AI 데이터 영향평가 제도의 도입 필요성을 제시하는 것을 목적으로 한다. 이를 위하여 공공데이터 개방정책과 개인정보 보호 관련 법제를 검토하고, 미국의 Sweeney 사건, Netflix Prize사건, AOL 검색기록 사건, 뉴욕 택시 데이터 사건, Strava Heat Map 사건 및 보건의료데이터 활용 사례 등을 분석하였다. 또한 EU GDPR 제35조의 데이터보호영향평가(Data Protection Impact Assessment, DPIA) 제도를 비교법적으로 검토하여 우리나라 법제에 대한 시사점을 도출하였다. 연구 결과, 생성형 AI 시대의 개인정보 침해는 단순한 정보 유출이 아니라 데이터 결합, AI 기반 추론, 집단정보 및 조직정보의 재식별 등 새로운 형태로 나타나고 있으며, 이러한 위험은 기존 개인정보 영향평가나 고영향 AI 영향평가만으로는 충분히 통제하기 어려운것으로 분석되었다. 특히 공공데이터 개방 확대와 생성형 AI의 데이터 활용이 지속적으로 증가하는 상황에서는 개인정보뿐만 아니라 조직정보와 국가안보정보까지 보호할 수 있는 사전적 위험관리 체계가 요구된다. 이에 본 연구는 GDPR의 데이터보호영향평가 제도를 참고하여 공공데

* 2025년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2025S1A5B5A16006202)
** 덕성여대 AI DYNAINFO 연구소 학술연구교수, 법학박사
*** 단국대 법학과 초빙교수, 법학박사

이터 개방, 가명정보 결합 및 생성형 AI 학습데이터 구축 등 고위험 데이터 처리 행위를 대상으로 하는 「AI 데이터 영향평가(AI Data Impact Assessment)」 또는 「재식별 영향평가(Re-identification Impact Assessment)」 제도의 도입을 제안한다. 이러한 제도는 재식별 가능성, 데이터 결합 위험, AI 추론 가능성 및 민감정보 노출 위험 등을 사전에 평가함으로써 공공데이터 활용과 개인정보 보호의 균형을 실현하고, 생성형 AI 시대에 적합한 새로운 데이터 거버넌스 구축에 기여할 수 있을 것으로 기대된다.

주제어 : 공공데이터, 생성형 인공지능, 개인정보 보호, 재식별, AI 데이터 영향평가, 데이터보호영향평가(DPIA), GDPR, 공공데이터 개방, 가명정보, 데이터 거버넌스

• 투고일 : 2026.06.30. / 심사일 : 2026.07.25. / 게재확정일 : 2026.07.25.

Ⅰ. 서론

21세기 디지털 경제에서 데이터는 새로운 생산요소로 평가받고 있으며, 국가 경쟁력과 산업혁신의 핵심 자원으로 인식되고 있다. 특히 인공지능(AI), 빅데이터, 클라우드 컴퓨팅 등 디지털 기술의 발전은 데이터 활용의 중요성을 더욱 증대시키고 있다. 이러한 흐름 속에서 각국 정부는 공공기관이 보유한 데이터를 적극적으로 개방하여 산업 발전과 행정 혁신을 도모하고 있으며,[1] 우리나라 역시 2013년 「공공데이터의 제공 및 이용 활성화에 관한 법률」을 제정하여 공공데이터 개방정책을 적극 추진하고 있다.

공공데이터는 교통, 보건, 환경, 교육, 복지, 문화 등 다양한 분야의 정보를 포함하고 있으며, 민간기업과 연구기관은 이를 활용하여 새로운 서비스를 개발하고 있다. 최근에는 생성형 AI(Generative AI)의 발전으로 인해 공공데이터의 중요성이 더욱 커지고 있다.

생성형 AI는 대규모 언어모델(LLM)을 기반으로 작동하기 때문에 방대한 양의 학습데이터를 필요로 하며, 공공데이터는 신뢰성과 정확성을 갖춘 양질의 학습 데이터로 평가받고 있다. 이에 따라 정부는 공공데이터 개방 확대를 AI 산업 경쟁력 확보를 위한 중요한 정책수단으로 활용하고 있다.[2] 그러나 공공

1) 헌법재판소 2019. 7. 25. 선고 2017헌마1329 결정, 알 권리는 정보에 접근하고 수집·처리함에 있어 국가권력의 방해를 받지 않을 자유권적 성질과, 의사형성이나 여론형성에 필요한 정보를 적극적으로 수집할 청구권적 성질을 가진다. 정보공개청구권은 알 권리의 당연한 내용으로서 헌법 제21조에 의해 직접 보장되며 법률이 기본권을 제한할 경우 과잉금지원칙(목적의 정당성, 수단의 적합성, 침해의 최소성, 법익의 균형성)을 준수해야 한다.

데이터 개방 확대는 개인정보 보호와의 새로운 충돌을 야기하고 있다.[3] 과거에는 개인정보의 보호가 주로 정보 유출이나 해킹과 같은 보안문제 중심으로 논의되었다. 그러나 생성형 AI 시대에는 데이터 결합(Data Combination), 프로파일링(Profiling), 추론(Inference) 기술이 급속히 발전함에 따라 직접 식별정보가 제거된 가명 정보 또는 비식별정보라 하더라도 특정 개인이나 조직을 식별할 수 있는 가능성이 증가하고 있다.

실제로 미국의 Sweeney 사건에서는[4] 우편번호, 생년월일, 성별만으로 특정인의 의료정보를 식별할 수 있음이 입증되었고, Netflix 사건에서는[5] 영화 평점 데이터와 외부 공개정보를 결합하여 이용자를 재식별할 수 있음이 확인되었다. 또한 AOL 검색기록 사건[6], 뉴욕 택시 데이터 사건[7], Strava Heat Map 사건[8], 코로나19 확진자 동선 공개 사례[9], 영국 NHS－DeepMind 사건[10] 등은 모두

---

2) 경건, 한국의 공공데이터 개방의 법제와 쟁점, 행정법이론실무학회, 행정법 연구 47권, 2016.12, 3면

3) 권은정, 공공데이터 영역 리스크 관리에 관한 법적 소고, 행정법이론실무학회, 행정법 연구 제60권, 2020.2 170면 이하, 2020년 2월 「개인정보 보호법」 개정으로 정보 주체의 동의 없이 '가명정보'를 제한된 목적으로 처리할 수 있도록 함에 따라 민간기업은 물론이고 공공기관의 공공 빅데이터 제공에도 더욱 힘이 실리게 되었다. 그런데 개정법은 민간의 빅데이터 활용, 기업 간 정보 결합 등에 초점을 두고 안전조치의무, 과징금, 형사처벌 등을 도입한 반면에, 공공기관의 공공데이터 개방 확대에 대응하여 특별히 강화된 리스크 관리 책임을 부담시키고 있지는 않다. 공공데이터 개방에 따라 공공부문에서도 보안 위협을 비롯한 '데이터 리스크'가 증폭될 것으로 예상되는 한편, 재식별 위험성이 있음에도 불구하고 정보주체의 동의권은 영구히 배제되는 '개인정보자기결정권 침해 리스크'도 인정된다.

4) Latanya Sweeney, Simple Demographics Often Identify People Uniquely, Data Privacy Working Paper No. 3, Carnegie Mellon University, 2000, p. 2.

5) Narayanan, Arvind · Shmatikov, Vitaly, 「Robust De-anonymization of Large Sparse Datasets」, Proceedings of the 2008 IEEE Symposium on Security and Privacy, IEEE Computer Society, 2008. "We successfully identified the Netflix records of known users." 우리는 알려진 이용자들의 넷플릭스 기록을 성공적으로 식별하였다. "An adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record." 공격자는 특정 이용자에 대한 약간의 정보만 알고 있어도 해당 이용자의 기록을 쉽게 식별할 수 있었다.

6) Barbaro, Michael & Zeller Jr., Tom, "A Face Is Exposed for AOL Searcher No. 4417749", The New York Times, Aug. 9, 2006.; https://en.wikipedia.org/wiki/AOL_search_log_release? 최종검색일 : 2026.6.23

7) Tockar, Anthony, "Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset", Neustar Research Blog, Sept. 15, 2014.;Pandurangan, Vijay, "On Taxis and Rainbows: Lessons from NYC's Improperly Anonymized Taxi Logs", Medium, June 21, 2014.

8) Nathan Ruser, Twitter Post, Jan. 27, 2018; "Strava Fitness App Can Reveal Military Bases, Analysts Say," BBC News, Jan. 29, 2018.

9) https://www.pdjournal.com/news/articleView.html?idxno=71317, 박수선기자, 확진자 동선

직접 식별정보가 제거된 데이터라 하더라도 다른 데이터와 결합될 경우 개인정보 또는 조직정보가 재식별될 수 있음을 보여주는 대표적 사례들이다.

생성형 AI는 기존의 데이터 분석기술과 달리 방대한 양의 데이터를 자동으로 결합·분석하고 새로운 정보를 추론할 수 있다는 점에서 재식별 위험을 더욱 증폭시키고 있다. 문제는 현행 개인정보보호법 제33조 개인정보 영향평가는 개인정보파일의 안전성 확보를 중심으로 설계되어 있기 때문에 재식별 위험을 처리하는데 적합하지 않다. 한편「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」제32조 내지 제35조상 고영향 AI 영향평가 제도 역시 AI 시스템이 인간의 권리와 자유에 미치는 영향에 초점을 두고 있기 때문에 이러한 재식별 위험의 사전 대응을 충분히 평가하지는 못하고 있다.

반면 EU GDPR 제35조는 데이터보호영향평가(Data Protection Impact Assessment, DPIA)를 통하여 개인정보 처리행위 자체가 정보주체의 권리와 자유에 미치는 위험을 사전에 평가하도록 하고 있다. 특히 재식별 위험, 데이터 결합 위험, 프로파일링 위험 등 새로운 개인정보 침해 가능성을 사전에 검토하도록 한다는 점에서 생성형 AI 시대의 개인정보 보호에 중요한 시사점을 제공한다.

이에 본 연구는 공공데이터 개방과 생성형 AI 활용 과정에서 발생할 수 있는 개인정보 재식별 위험을 분석하고, 현행 개인정보보호법상 영향평가 제도와 AI 기본법상 영향평가 제도의 한계를 검토한 후, GDPR 제35조를 참고하여「AI 데이터 영향평가(AI Data Impact Assessment)」또는「재식별 영향평가(Re-identification Impact Assessment)」제도의 도입 필요성을 제시하는 것을 목적으로 한다. 이를 통하여 공공데이터 개방과 AI 산업 발전이 개인정보 보호와 조화를 이루는 새로운 법적·제도적 개선방안을 제안하고자 한다.

보도 '성소수자 혐오' 우려 큰데...교회언론회 "공익 보도였다", 피디저널, 2020.5.11., 최종검색일 : 2026.6.23.,2020년 1월부터 3월 사이에는 중앙정부와 지방자치단체가 경쟁적으로 확진자 정보를 공개하였는데 당시 공개된 정보에는 성별, 연령, 국적, 거주지역, 직장, 방문장소, 이동시간, 이동수단 등이 포함되는 경우가 많았다. 일부 지자체는 "30대 여성", "○○회사 근무", "○○아파트 거주" 등의 정보까지 공개하였고, 온라인 커뮤니티에서는 이를 토대로 특정 개인의 신원을 추정하는 사례가 발생하였다.

10) https://www.wired-gov.net/wg/news.nsf/articles/Royal%2BFree%2BGoogle%2BDeepMind%2Btrial%2Bfailed%2Bto%2Bcomply%2Bwith%2Bdata%2Bprotection%2Blaw%2B03072017131000 최종검색일 : 2026.6.23

Ⅱ. 공공데이터 개방정책과 생성형 AI의 재식별 위험의 현황과 법적 한계

1. 공공데이터 개방정책의 의의

현대 정보사회에서 데이터는 국가경쟁력과 산업혁신의 핵심 자원으로 인식되고 있다. 특히 공공기관이 보유·관리하는 데이터는 국민의 세금으로 생산된 공공자산이라는 점에서 공공성과 활용성을 동시에 가진다. 이에 따라 세계 각국은 공공데이터 개방정책(Open Government Data Policy)을 추진하고 있으며, 우리나라 역시 「공공데이터의 제공 및 이용 활성화에 관한 법률」(이하 "공공데이터법")을 제정하여 공공데이터 개방을 제도적으로 지원하고 있다. 공공데이터법 제1조는 "공공데이터의 제공 및 이용 활성화를 통하여 국민의 공공데이터에 대한 이용권을 보장하고 국민경제 발전에 이바지함"을 목적으로 규정하고 있다. 이는 공공데이터가 단순한 행정정보가 아니라 국가경제 발전과 산업혁신을 위한 전략적 자원이라는 점을 선언한 것으로 평가된다.

전통적인 정보공개제도가 국민의 알 권리 보장을 중심으로 발전하였다면, 공공데이터 개방정책은 데이터를 재이용하여 새로운 경제적·사회적 가치를 창출하는 데 목적이 있다. 따라서 공공데이터는 단순히 열람 대상이 아니라 창업, 연구개발, 서비스 개발 및 인공지능 학습 등에 활용되는 생산요소로 기능한다.

특히 디지털 플랫폼 경제와 인공지능 산업의 발전으로 인해 데이터의 경제적 가치가 급격히 증가하고 있으며, 공공데이터는 정확성, 신뢰성, 공공성을 갖춘 데이터라는 점에서 민간 데이터와 차별화된다. 이러한 이유로 공공데이터 개방은 국가 차원의 디지털 전환과 AI 산업 육성을 위한 핵심 정책으로 자리매김하고 있다. 우리나라 공공데이터 개방정책의 본격적인 출발점은 박근혜 정부의 2013년 정부3.0 정책이라고 평가할 수 있다. 정부3.0은 정부가 보유한 정보를 적극적으로 공개하고 부처 간 칸막이를 제거하여 국민 중심의 서비스를 제공하는 것을 목표로 하였다. 정부3.0 이전의 행정은 정부가 정보를 독점하고 국민이 정보공개청구를 통하여 제한적으로 접근하는 방식이었다. 그러나 정부3.0은 정부가 보유한 정보를 원칙적으로 개방하고 국민과 기업이 자유롭게 활용할 수 있도록 한다는 점에서 기존의 행정 패러다임과 차이를 보인다.

이에 따라 정부는 공공데이터포털을 구축하여 교통, 환경, 교육, 문화, 복지,

보건, 기상, 행정 등 다양한 분야의 데이터를 공개하기 시작하였다. 이러한 데이터는 민간기업의 서비스 개발과 연구기관의 정책연구에 폭넓게 활용되고 있으며, 데이터 기반 산업 생태계 형성에 중요한 역할을 수행하고 있다. 그러나 정부3.0 정책은 데이터 개방의 확대에 중점을 둔 나머지 개인정보 보호 문제를 상대적으로 충분히 검토하지 못하였다는 비판도 존재한다. 특히 초기에는 비식별화 조치만 이루어지면 개인정보 침해 위험이 크지 않다고 보았으나, 데이터 분석기술과 인공지능 기술이 발전하면서 이러한 전제는 더 이상 유지되기 어려운 상황이 되었다.

2. 생성형 AI와 개인정보 재식별의 개념

(1) 생성형 AI 와 재식별 위험의 증가

최근 각국은 인공지능을 국가 전략산업으로 육성하고 있으며, 데이터 확보를 AI 경쟁력의 핵심 요소로 인식하고 있다. 생성형 AI의 발전은 방대한 양의 학습데이터를 필요로 하며, 데이터의 양과 질은 AI 성능을 결정하는 중요한 요인으로 작용한다. 우리나라는 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」을 제정하여 AI 산업 육성을 적극 추진하고 있으며, 국가 AI 전략과 디지털플랫폼정부 정책을 통하여 공공데이터 활용 확대를 추진하고 있다. 특히 공공데이터는 품질과 신뢰성이 높아 AI 학습데이터로서 높은 가치를 가진다.

그러나 AI 산업 발전을 위한 데이터 활용 확대는 개인정보 보호와 충돌할 가능성이 존재한다. 과거에는 데이터의 직접적인 공개 여부가 문제였다면, 오늘날에는 공개된 데이터가 생성형 AI의 학습자료로 활용되는 과정에서 새로운 개인정보 침해 가능성이 발생하고 있다.

생성형 AI(Generative AI)는 대규모 데이터를 학습하여 새로운 텍스트, 이미지, 음성, 영상 등을 생성하는 인공지능 기술을 의미한다. 대표적인 예로 ChatGPT, Gemini, Claude, Copilot 등이 있다. 전통적인 AI가 특정 문제를 분류하거나 예측하는 데 초점을 두었다면, 생성형 AI는 인간과 유사한 수준의 언어 생성과 추론을 수행할 수 있다는 점에서 차이를 가진다. 특히 생성형 AI는 대규모 언어모델(LLM)을 기반으로 작동하며, 수십억 개 이상의 문서와 데이터를 학습한다. 문제는 생성형 AI가 단순히 정보를 저장하는 것이 아니라 서로 다른 데이터 간의 관계를 학습하고 새로운 정보를 추론할 수 있다는 점이다. 이러한 특성은 AI의 활용 가능성을 확대하는 장점이 있지만, 동시에 개인정보 재식별 위험을 증가시키는 요인으로 작용한다.

(2) 개인정보 · 가명정보 · 익명정보와 재식별

개인정보보호법은 개인정보를 살아있는 개인에 관한 정보로서 해당 개인을 알아볼 수 있는 정보로 정의하고 있다. 개인정보는 크게 직접식별정보와 간접식별정보로 구분될 수 있다. 이름, 주민등록번호, 전화번호 등은 직접식별정보에 해당하며, 연령, 성별, 직업, 지역정보 등은 간접식별정보에 해당한다. 가명정보는 개인정보 일부를 삭제하거나 대체하여 특정 개인을 직접 식별할 수 없도록 처리한 정보이다. 그러나 다른 정보와 결합하면 특정 개인을 식별할 가능성이 존재한다. 반면 익명정보는 어떠한 방법으로도 특정 개인을 식별할 수 없는 상태의 정보를 의미한다. 따라서 익명정보는 개인정보보호법의 적용대상이 되지 않는다. 문제는 생성형 AI의 발전으로 과거에는 익명정보로 평가되었던 데이터도 재식별될 가능성이 증가하고 있다는 점이다. 이는 개인정보와 가명정보, 익명정보의 경계를 더욱 모호하게 만들고 있다.

재식별(Re-identification)이란 비식별화 또는 가명처리된 정보가 다른 데이터와 결합되거나 분석됨으로써 다시 특정 개인을 식별할 수 있게 되는 현상을 의미한다. 과거에는 개인정보 재식별이 주로 통계학적 분석이나 데이터 결합을 통하여 이루어졌으나, 최근에는 AI 기반 분석기술이 발전하면서 재식별의 가능성과 속도가 크게 증가하였다. 대표적으로 Sweeney 사건에서는 우편번호, 생년월일, 성별만으로 특정 개인의 의료정보를 식별할 수 있음이 입증되었으며, Netflix 사건에서는 영화평점 데이터와 공개된 영화리뷰 정보를 결합하여 특정 이용자를 재식별할 수 있음이 확인되었다. 재식별은 단순한 개인정보 침해를 넘어 개인정보 자기결정권 침해와 정보인권 침해로 이어질 수 있다는 점에서 중요한 법적 문제를 제기한다.

(3) AI 추론(Inference) 위험의 의의

생성형 AI 시대의 개인정보 위험은 단순한 재식별에 그치지 않는다. 생성형 AI는 명시적으로 제공되지 않은 정보까지 추론할 수 있기 때문이다. 예를 들어 특정 개인의 검색기록, 구매이력, 위치정보, SNS 게시물 등을 분석하면 건강상태, 정치성향, 종교, 경제수준, 장애 여부 등을 추론할 수 있다. 이를 AI 추론(Inference) 위험이라고 한다. AI 추론 위험은 전통적인 개인정보 보호체계가 충분히 고려하지 못한 새로운 유형의 위험이다. 왜냐하면 실제 개인정보가 공개되지 않았더라도 AI가 이를 유추하여 사실상 동일한 결과를 도출할 수 있기 때문이다. 생성형 AI는 다양한 데이터셋을 결합하여 새로운 정보를 생성할 수 있으므로, 재식별 위험과 추론 위험은 상호 결합되어 나타나는 경우가

많다. 따라서 생성형 AI 시대의 개인정보 보호는 단순한 정보유출 방지를 넘어 재식별과 추론 가능성까지 고려하는 새로운 접근이 요구된다.

3. 공공데이터 개방에 따른 재식별 위험 사례

(1) 미국 매사추세츠 의료데이터 재식별 사건[11]

Sweeney 교수는 공공데이터 개방이 비식별화되더라도 얼마나 개인정보에 위협적인지 보여주는 연구로서 매사추세츠주 의료데이터 공개 사례를 분석하면서 우편번호(ZIP Code), 생년월일(Date of Birth), 성별(Sex)이라는 단 세 가지 준식별자(quasi-identifiers)만으로도 미국 인구의 약 87%를 식별할 수 있음을 입증하였다. 이는 비식별화된 공공데이터라 하더라도 다른 데이터셋과 결합될 경우 개인정보 재식별이 가능함을 보여주는 대표적 연구로 평가된다. 즉, 1997년 미국 매사추세츠주는 공공정책 연구를 위해 주 공무원의 의료기록 데이터를 비식별 처리하여 공개하였다. 당시 정부는 이름 삭제, 주민번호 삭제, 주소 삭제 등을 수행하였기 때문에 개인정보가 아니라고 판단하였고 안전하다고 판단하였다. 그러나 카네기멜런대학교의 Latanya Sweeney 교수는 공개된 의료데이터와 유권자 명부를 결합하여 당시 매사추세츠 주지사 William Weld의 의료기록을 찾아내는 데 간단하게 성공하였다. 그 결과 우편번호, 생년월일, 성별만으로도 상당수 국민을 식별할 수 있다는 사실이 밝혀졌다. 이는 오늘날 AI가 다양한 데이터셋을 자동 결합하는 경우 재식별 위험이 훨씬 커질 수 있음을 보여주는 대표 사례로 제시되는 사례이다.

(2) 넷플릭스(Netflix Prize) 데이터 재식별 사건[12]

Narayanan과 Shmatikov는 넷플릭스가 공개한 익명화 영화평점 데이터를 분석하여 IMDb의 공개 리뷰 데이터와 결합하는 방식으로 특정 이용자를 재식별할 수 있음을 입증하였다. 이 연구는 이름이나 회원번호와 같은 직접식별정

11) Latanya Sweeney, Simple Demographics Often Identify People Uniquely, Data Privacy Working Paper No. 3, Carnegie Mellon University, 2000, p. 2.

12) Narayanan, Arvind · Shmatikov, Vitaly, 「Robust De-anonymization of Large Sparse Datasets」, Proceedings of the 2008 IEEE Symposium on Security and Privacy, IEEE Computer Society, 2008. “We successfully identified the Netflix records of known users.” 우리는 알려진 이용자들의 넷플릭스 기록을 성공적으로 식별하였다. “An adversary who knows only a little bit about an individual subscriber can easily identify this subscriber’s record.” 공격자는 특정 이용자에 대한 약간의 정보만 알고 있어도 해당 이용자의 기록을 쉽게 식별할 수 있었다.

보가 제거되었더라도 다른 공개 데이터와의 결합을 통해 개인 식별이 가능함을 보여준 대표적 사례로 평가된다. 특히 연구진은 재식별된 정보를 바탕으로 이용자의 정치적 성향이나 개인적 취향 등 민감정보까지 추론할 수 있음을 지적하였다. 이는 오늘날 생성형 AI가 대규모 데이터를 자동으로 결합·분석하는 환경에서 공공데이터의 가명처리만으로는 개인정보 보호가 충분하지 않을 수 있음을 시사한다. 2006년 넷플릭스는 추천 알고리즘 개발을 위해 약 50만 명의 영화 시청 데이터를 익명화하여 공개하였는데 이 때 넷플릭스는 이름 삭제, 회원번호 삭제 등을 수행하였으므로 안전하다고 보았다. 그러나 연구자들은 IMDb(Internet Movie Database) 리뷰 정보와 결합하여 특정 사용자의 영화 취향을 역추적하는 데 성공하였다. 이를 통해 정치성향, 종교, 성적 취향 등 민감정보까지 추론 가능하다는 사실이 밝혀졌고 생성형 AI는 이러한 추론 능력을 훨씬 강화할 수 있다는 점에서 중요한 사례로 알려지게 되었다. 즉, 50만 명 이용자는 영화평점을 기록하였을 뿐이었는데 그에 대한 데이터셋(Netflix Prize Dataset)을 분석하여 특정 사용자를 찾아낼 수 있었는데 당시 넷플릭스는 개인정보영향평가에 있어 자신감을 보이며 이용자의 이름, 주소, 회원번호 등을 삭제한 후 데이터를 공개하였으므로 개인정보가 아니라고 판단하였다. 그리하여 익명화된 데이터라고 하더라도 다른 공개 데이터와 결합하면 개인을 재식별할 수 있음이 입증되었다.

(3) AOL 검색기록 공개 사건[13)]

2006년 8월 4일, 압두르 초드후리가 이끄는 AOL 리서치는 3개월 동안 65만 명 이상의 사용자를 대상으로 2천만 건의 검색 쿼리를 담은 압축 텍스트 파일을 한 웹사이트에 공개했다. AOL은 2006년 약 65만 명 이용자의 2천만 건 이상의 검색 기록을 제공하면서 이 내용은 익명화되어 공개되기 때문에 문제가 되지 않는다고 하였으나 특정 이용자가 식별되는 문제가 발생하였다. AOL은 8월 7일까지 해당 파일을 사이트에서 급히 삭제했지만, 그 전에 복사되어 이미 인터넷에 배포된 뒤였다. AOL이 검색한 부분에 대해 특정 사용자 수치로 식별한 계정을 귀속하도록 분류했기 때문에, 이용자 번호 4417749의 검색기록을 분석하여 개인을 식별하고 계정 및 검색 기록과 매칭할 수 있었는데 이는 익명화된 검색 기록을 전화번호부 목록과 교차 조회하여 실제 인물인 Thelma Arnold를

13) Barbaro, Michael & Zeller Jr., Tom, “A Face Is Exposed for AOL Searcher No. 4417749”, The New York Times, Aug. 9, 2006.; https://en.wikipedia.org/wiki/AOL_search_log_release? 최종검색일 : 2026.6.23

찾아냈던 일로 AOL은 이것이 실수임을 인정하고 데이터를 삭제했다.

그리하여 이 일은 익명화된 데이터라도 다른 정보와 결합될 경우 재식별이 가능함을 보여주는 대표적 사례로 평가되었으며 이 때 이용자 아놀드는 이름 대신 숫자 ID를 사용하였으나 이용자가 검색한 내용 중에는 집주소에 접근하는 방법, 병원, 가족관계, 지역정보 등이 포함되어 있었고 누구나 특정 이용자를 실제 인물로 식별하는 데 성공할 수 있었던 것이다. 결국 2006년 9월, 캘리포니아 북부 지방법원에 AOL은 집단 소송으로 피소되었다[14]. AOL은 전자통신프라이버시법(EFA)을 위반했다고 하면서 노출된 사람당 최소 5,000달러의 배상금을 요구하였다. 전자통신 프라이버시법(Electronic Communications Privacy Act of 1986)은 전자통신의 내용을 고의로 공개해서는 안 되며[15] 원고들은 AOL이 약 65만 명의 이용자, 약 2천만 건의 검색기록을 인터넷에 공개함으로써 검색기록 자체가 전자통신 내용(contents of communication), 가입자 정보(record concerning subscriber)에 해당하므로 ECPA §2702를 위반했다고 주장하였다. 특히 원고들은 검색어(query)는 단순한 로그정보가 아니라 이용자의 사상, 건강상태, 정치적 성향, 종교, 가족관계 등을 포함하는 통신 내용(content)이라고 주장하였다. AOL 검색기록 공개 사건에서는 익명화된 검색 데이터라 하더라도 검색어 자체에 개인정보가 포함될 수 있다는 점이 문제되었는데 원고들은 AOL이 약 65만 명 이용자의 2천만 건 이상의 검색 기록을 공개함으로써 전자통신프라이버시법(ECPA) 제2702조 위반이 주장되었던 것이다. 결국 해당 소송은 이후 Landwehr v. AOL Inc. 사건으로 병합되어 진행되었으며, 2013년 AOL이 500만 달러 규모의 집단소송 합의에 동의함으로써 종결되었다.[16] 이는 비식별화된 데이터라 하더라도 검색 패턴과 질의 내용만으로 개인 식별이 가능하다는 점을 보여준 대표적 사례로 평가된다.

---

14) Landwehr v. AOL Inc., No. 1:11-cv-01014-CMH-TRJ (E.D. Va. May 24, 2013).

15) 18 U.S.C.(ECPA) §2702(a)(1) Except as provided in subsection (b), a person or entity providing an electronic communication service to the public shall not knowingly divulge to any person or entity the contents of a communication while in electronic storage by that service. 제(b)항에서 정한 경우를 제외하고, 일반 대중에게 전자통신 서비스를 제공하는 자 또는 기관은 해당 서비스에 의해 전자적으로 저장된 통신의 내용을 고의로 어떠한 개인이나 기관에게도 공개하여서는 아니 된다. §2702(a)(3) A provider of remote computing service to the public shall not knowingly divulge a record or other information pertaining to a subscriber or customer of such service.일반 대중에게 원격 컴퓨팅 서비스를 제공하는 사업자는 가입자 또는 고객에 관한 기록이나 기타 정보를 고의로 공개하여서는 아니 된다.

16) Landwehr v. AOL Inc., No. 1:11-cv-01014-CMH-TRJ, United States District Court for the Eastern District of Virginia, May 24, 2013.

(4) 뉴욕 택시 데이터 재식별 사건[17]

뉴욕 택시 데이터 재식별 사건은 뉴욕시 택시ㆍ리무진위원회가 공개한 택시 운행 데이터에서 비롯되었는데 해당 데이터에는 뉴욕의 옐로캡 운행의 승차ㆍ하차 시간, 승차ㆍ하차 위치, 요금, 팁 금액, 택시 면허번호와 차량번호를 해시 처리한 값 등이 포함되어 있었다. Anthony Tockar는 이 데이터가 겉으로는 익명화되어 있었지만, 시간ㆍ장소 정보와 외부 공개정보를 결합하면 특정 승객의 이동경로를 추정할 수 있음을 보여주었다. 이 사건의 핵심은 위치정보와 시간정보의 결합 위험성으로 데이터에는 이름이나 전화번호 같은 직접 식별자는 없었지만, “언제, 어디서 택시를 탔고 어디서 내렸는가”라는 정보 자체가 매우 강한 준 식별자 역할을 했다. 이에 대해서는 연구자 Vijay Pandurangan도 2014년 뉴욕시 택시 데이터 공개 사건을 분석하면서, “On Taxis and Rainbows: Lessons from NYC’s Improperly Anonymized Taxi Logs”[18]라는 유명한 글에서 뉴욕시가 공개한 1억 7,300만 건 이상의 택시 운행기록의 공개는 분석 결과, 익명화 과정에 중대한 결함이 존재함을 지적하였다. 즉, 해시(hash) 처리된 택시 식별번호가 역산될 수 있음을 보여주었으며, 이를 통해 특정 운전기사의 운행정보를 복원할 수 있다고 분석한 것이다. Anthony Tockar가 비식별처리를 했더라도 승객의 재식별 가능성을 보여주었다면 Vijay Pandurangan은 익명화 기술 자체의 취약성을 밝혔다. 그리하여 뉴욕시는 택시번호를 익명화하였다고 주장했으나 실제로는 단순한 MD5 해시(Hash)를 사용하였고, 이를 역산하여 원래 택시번호를 복원할 수 있음을 입증하여 ① 운전기사 식별 가능 ② 승객 이동경로 추적 가능 ③ 외부 데이터 결합 가능이라는 점을 보여주었다. 즉, 뉴욕시가 공공데이터로 공개한 데이터는 승차위치, 하차위치, 시간정보 등이었으나 연구자들은 이를 활용하여 유명인의 이동경로를 추적하고 특정 운전기사와 승객을 식별할 수 있을 뿐만 아니라 AI 기반 위치분석기술까지 발전하면서 위치정보 데이터의 재식별 위험이 크게 증가하고 있음을 경고하고 있는 것이다.

17) Tockar, Anthony, “Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset”, Neustar Research Blog, Sept. 15, 2014.;Pandurangan, Vijay, “On Taxis and Rainbows: Lessons from NYC’s Improperly Anonymized Taxi Logs”, Medium, June 21, 2014.

18) https://medium.com/vijay-pandurangan/of-taxis-and-rainbows-f6bc289679a1, Vijay Pandurangan, “On Taxis and Rainbows: Lessons from NYC’s Improperly Anonymized Taxi Logs,” Medium, June 21, 2014, available at Medium Article (최종방문일: 2026. 6. 23.)

(5) Strava 군사기지 노출 사건

Nathan Ruser는 호주의 전략정책연구소(ASPI) 연구원이었으며 2018년 1월 27일 자신의 X(당시 Twitter)에 Strava Heat Map 분석 결과를 공개하였다.[19] 미국의 운동기록 앱인 Strava Heat Map은 이용자들이 달리기, 자전거, 등산, 군사훈련 등을 기록할 수 있는 GPS 기반 운동 플랫폼이다. 2017년 11월 Strava는 전 세계 약 2,700만 명의 이용자로부터 수집한 약 30억 건 이상의 GPS 운동기록을 기반으로 Global Heat Map을 공개하였다. Strava는 이름 삭제, 계정 정보 삭제 등의 조치를 취하였기 때문에 개인정보 문제가 없다고 판단하였다. 2018년 1월 Nathan Ruser는 Heat Map을 분석하던 중 이상한 현상을 발견하였는데, 사하라 사막, 시리아, 아프가니스탄, 이라크 등 인구가 거의 없는 지역에서 특정 경로가 매우 밝게 표시되고 있었던 것이다. 그는 이를 분석한 결과 해당 경로들이 미군 기지, NATO 군사시설, CIA 관련 시설, 특수부대 작전지역과 일치한다는 사실을 발견하였다. 문제는 아프가니스탄의 외딴 군사기지에서는 일반 시민이 거의 존재하지 않는데 Heat Map에는 병영, 숙소, 식당, 경계 순찰로가 선명하게 표시되어 있었던 것이다. 이는 군인들이 운동할 때 Strava 앱을 사용했기 때문이다. 결과적으로 공개된 Heat Map만으로 기지의 위치, 규모, 주요 시설 배치, 경비 순찰 경로까지 추정할 수 있게 되었던 것이었다. 그리하여 CIA 및 정보기관 시설이 모두 노출되었는데 일부 지역에서는 CIA 전진기지, 정보 수집시설, 비공개 작전기지로 추정되는 시설까지 노출되었다. 특히 시리아와 아프리카 지역에서는 공개적으로 알려지지 않았던 시설의 존재가 Heat Map 분석을 통해 추정되었다. 중요한 점은 Strava가 직접 개인정보를 공개한 것이 아니라는 점이다. 문제는 ① GPS 위치정보 ② 반복적인 이동 패턴 ③ 공개 위성사진 ④ 군사시설 위치정보를 결합함으로써 개별 이용자 또는 군사 조직을 식별할 수 있게 되었다는 것이다. 즉, 익명화된 데이터와 공개 데이터, 분석 기술은 재식별이라는 공식이 현실에서 입증된 사례였다. Strava 사건 당시에는 연구자가 직접 Heat Map을 분석하였으나 생성형 AI 시대에는 상황이 훨씬 심각해진다. AI는 GPS 데이터, SNS 게시물, 공개 지도, 위성 사진, 뉴스기사를 자동으로 결합할 수 있기 때문이다. 따라서 과거에는 전문가가 수개월 동안 분석해야 했던 작업을 AI는 수분 내에 수행할 수 있다.

---

19) Nathan Ruser, Twitter Post, Jan. 27, 2018; “Strava Fitness App Can Reveal Military Bases, Analysts Say,” BBC News, Jan. 29, 2018.

(6) 보건데이터 재식별 위험 사례

우리나라에서는 2020년 1월 20일 중국 우한에서 입국한 35세 중국인 여성이 최초 확진자로 확인되면서 코로나19 대응이 시작되었는데 이후 2020년 2월 대구·경북 지역을 중심으로 대규모 집단감염이 발생하면서 확진자가 급증하였고, 정부는 감염병 위기경보를 최고 단계인 “심각”으로 격상하였다. 이후 방역당국은 광범위한 역학조사와 접촉자 추적을 실시하였으며, 확진자의 이동경로 공개를 주요 방역수단으로 활용하였다.[20] 「감염병의 예방 및 관리에 관한 법률」 제34조의2 ‘감염병위기 시 정보공개’는 질병관리청장, 시·도지사 및 시장·군수·구청장은 국민의 건강에 위해가 되는 감염병 확산으로 인하여 「재난 및 안전관리 기본법」 제38조제2항에 따른 주의 이상의 위기경보가 발령되면 감염병 환자의 이동경로, 이동수단, 진료의료기관 및 접촉자 현황, 감염병의 지역별·연령대별 발생 및 검사 현황 등 국민들이 감염병 예방을 위하여 알아야 하는 정보를 정보통신망 게재 또는 보도자료 배포 등의 방법으로 신속히 공개하여야 한다. 다만, 성별, 나이, 그 밖에 감염병 예방과 관계없다고 판단되는 정보로서 대통령령으로 정하는 정보는 제외하여야 한다.고 규정하고 있다. 이 때 감염병 환자의 이동경로, 이동수단, 진료의료기관 및 접촉자 현황 등이 포함된 확진자 동선이 공개되었는데 이 과정에서 성별, 연령, 방문장소, 이동시간 등이 공개되었고, 온라인 커뮤니티에서는 특정 개인을 식별하는 사례가 발생하였던 것이다. 이에 대해 특정 집단에 비난이 쏟아지자 많은 사람이 질병관리청의 과도한 동선 공개가 개인정보 침해 위험을 초래한다고 지적하였다.

그 밖에 우리나라에서 공공데이터와 가명정보 활용의 대표적인 사례로는 국민건강보험공단과 건강보험심사평가원이 제공하는 보건의료 빅데이터를 들 수 있다. 국민건강보험공단은 전 국민 건강보험 가입자의 진료내역, 건강검진 정보, 처방정보, 보험자격 정보 등을 보유하고 있으며, 건강보험심사평가원은 의료기관의 진료비 청구자료, 질병정보, 약제처방 정보 등을 축적하고 있다. 이들

20) 강태경,유진, 코로나19 위기관리정책의 인권 형사정책적 개선방안 연구, 한국형사정책연구원, 2021, 26면, 시라쿠사 원칙에 따라 공중보건 또는 국가 비상사태를 이유로 인권을 제한하는 법률 조항의 정당성과 관련된 일반적 해석 원칙을 제시하고 있다. 공중보건 또는 국가 비상사태를 이유로 인권을 제한하는 조항은, (1) 규약을 최대한 존중하여야 하고 (규약의 최우선성), (2) 규약 상 권리를 최대한 보호해야 하고(권리보장의 최대성), (3) 법률에 근거해야 하고(법률성), (4) 제한을 목적 달성에 필요한 범위 내로 국한해야 하고(제한의 최소성), (5) 제한을 자의적으로 적용하지 말아야 하고(제한의 비자의성), (6) 이의제기 및 구제책을 두어야 하고(이의제기 가능성), (7) 제한의 정당화에 대한 책임을 국가에 부과해야 한다(국가의 책무성). 이때 ‘필요한’ 범위란, 제한이 규약에 근거를 두고 공익적이고 합법적인 목적에 비례하는 범위를 의미한다.

기관은 보건의료 연구와 정책개발, 신약개발 및 의료 인공지능 연구 등을 지원하기 위하여 가명처리된 형태로 데이터를 제공하고 있다.[21] 보건의료 데이터는 질병 연구와 공중보건 정책 수립에 매우 중요한 공공적 가치를 가진다. 실제로 건강보험 빅데이터는 코로나19 대응, 희귀질환 연구, 의료비 분석, 의료서비스 개선 및 인공지능 의료기술 개발 등에 광범위하게 활용되고 있다. 특히 최근에는 생성형 AI와 의료 AI의 발전에 따라 대규모 의료데이터에 대한 수요가 급격히 증가하고 있으며, 정부 역시 데이터 기반 정밀의료와 디지털 헬스케어 산업 육성을 적극 추진하고 있다. 그러나 의료정보는 개인정보 중에서도 가장 높은 수준의 민감정보에 해당한다. 질병명, 처방내역, 입원기록, 수술이력 등은 개인의 건강상태뿐 아니라 경제적 상황, 직업, 생활습관, 장애 여부, 정신건강 상태 등 다양한 개인정보를 포함하고 있기 때문이다. 더욱이 의료정보는 한 번 유출되면 변경이 사실상 불가능하다는 특성을 가지고 있어 일반 개인정보보다 보호 필요성이 더욱 크다. 문제는 가명처리가 이루어졌다고 하더라도 재식별 위험이 완전히 제거되는 것은 아니라는 점이다. 예를 들어 특정 지역에 거주하는 희귀질환 환자의 경우 질병명, 연령, 성별, 거주지역과 같은 정보만으로도 특정 개인을 추정할 가능성이 존재한다. 특히 인구가 적은 지역에서 특정 연령대의 희귀질환 환자 정보가 포함될 경우 재식별 가능성은 더욱 높아진다. 또한 공개된 SNS 게시물, 환우회 활동정보, 언론보도, 지역사회 정보 등이 결합될 경우 가명정보만으로도 개인 식별이 가능해질 수 있다.[22] 이러한 위험은 생성형 AI의 등장으로 더욱 심화되고 있다. 과거에는 재식별을 위해 전문 연구자가 여러 데이터베이스를 수작업으로 비교·분석해야 하였으나, 생성형 AI는 방대한 양의 데이터를 자동으로 수집·분석하고 상호 연관성을 추론할 수 있다. 특히 대규모 언어모델(LLM)은 공개정보와 의료데이터 사이의 관계를 분석하여 특정 개인의 건강상태나 질병 이력을 추정할 가능성이 있다. 이는 기존의 비식별화 또는 가명처리 기법만으로는 충분한 개인정보 보호가 어렵다는 점을 시사한다.

---

21) 김지희, 보건의료 빅데이터의 활용과 개인정보보호, 경인문화사,2022, 44면~47면

22) Nicholas Carlini et al., “Quantifying Memorization Across Neural Language Models”, ICLR, 2023.Nicholas Carlini et al., “Extracting Training Data from Large Language Models”, USENIX Security Symposium, 2021.

(7) ChatGPT 개인정보 노출 사례[23]

생성형 AI 자체가 개인정보 위험의 원인이 될 수 있다는 것은 2023년 3월에 발생한 OpenAI의 사건보고서 때문이었다. ChatGPT 서비스에서 오픈소스 라이브러리(Redis)의 버그로 인해 일부 이용자의 개인정보가 다른 이용자에게 노출될 가능성이 있었음을 공개하면서 OpenAI 조사 결과, 일부 이용자에게 다른 사용자의 대화 제목(Chat Title), 이름, 이메일 주소, 결제주소 일부, 신용카드 정보 일부(마지막 4자리 및 만료일) 등이 노출될 수 있었던 것으로 확인되었는데 OpenAI는 문제를 발견한 직후 서비스를 일시 중단하고 보안패치를 실시하였다고 하였지만 일부 이용자의 대화 제목, 결제정보 일부가 노출되는 사건은 생성형 AI 시스템이 대규모 데이터를 처리하는 과정에서 개인정보 유출 가능성이 현실화된 사례라 할 수 있을 것이다.

OpenAI에 따르면 일부 사용자는 다른 이용자의 이름, 이메일 주소, 결제주소 및 신용카드 정보 일부를 열람할 수 있었던 것으로 확인되었는데 이 사건은 생성형 AI 서비스가 대규모 데이터를 처리하는 과정에서 기술적 결함만으로도 개인정보 침해가 발생할 수 있음을 보여주는 대표적 사례로 평가되었다.

4. 생성형 AI 시대 재식별 위험의 특징

생성형 AI의 발전은 개인정보 보호의 패러다임을 근본적으로 변화시키고 있다. 전통적인 개인정보 침해는 주로 개인정보의 무단 수집, 유출 또는 해킹과 같은 형태로 발생하였다. 그러나 생성형 AI 시대에는 개인정보가 직접 유출되지 않더라도 다양한 데이터의 결합과 인공지능의 추론 기능을 통하여 특정 개인이나 조직에 관한 정보를 식별하거나 추정할 수 있는 새로운 유형의 위험이 등장하고 있다. 생성형 AI는 대규모 언어모델(LLM)을 기반으로 방대한 데이터를 학습하고 서로 다른 데이터 간의 연관관계를 분석할 수 있기 때문에 기존의 비식별화 또는 가명처리 기술만으로는 충분한 개인정보 보호를 보장하기 어렵다. 이러한 점에서 생성형 AI 시대의 재식별 위험은 기존의 개인정보 침해와 구별되는 독자적인 특성을 가진다. 첫째, 데이터 결합 위험이다. 생성형 AI 시대 재식별 위험의 가장 중요한 특징은 데이터 결합(Data Combination)에

23) https://openai.com/index/march-20-chatgpt-outage/, OpenAI, “March 20 ChatGPT Outage: Here’s What Happened”, Mar. 24, 2023.BBC News, “ChatGPT Bug Exposed Users’ Chat Histories”, Mar. 24, 2023.Reuters, “OpenAI Says ChatGPT Bug Exposed Some User Information”, Mar. 24, 2023.

의한 재식별 가능성의 증가이다. 과거에는 개인정보를 보호하기 위하여 이름, 주민등록번호, 주소 등 직접식별정보를 삭제하거나 가명처리하는 방식이 일반적으로 활용되었다. 이러한 방법은 개별 데이터셋만을 기준으로 할 경우 일정 수준의 보호 효과를 가질 수 있었다. 그러나 생성형 AI는 하나의 데이터셋만을 분석하는 것이 아니라 여러 데이터셋을 동시에 학습하고 상호 연관성을 분석할 수 있다는 점에서 기존의 비식별화 방식에 근본적인 한계를 드러내고 있다. 대표적으로 Sweeney 사건에서는 우편번호, 생년월일, 성별이라는 단순한 정보만으로도 특정 개인의 의료정보를 식별할 수 있음이 입증되었다. 또한 Netflix 사건에서는 익명화된 영화평점 데이터와 공개된 IMDb 리뷰 데이터를 결합하여 특정 이용자를 재식별할 수 있었다. AOL 검색기록 사건 역시 이용자의 이름이나 계정정보 없이 공개된 검색어만으로 특정 개인을 찾아낼 수 있음을 보여주었다. 생성형 AI는 이러한 데이터 결합을 인간보다 훨씬 빠르고 광범위하게 수행할 수 있다. 예를 들어 공공데이터, SNS 게시물, 위치정보, 뉴스기사, 의료정보, 구매기록 등이 동시에 학습될 경우 개별 데이터 자체는 비식별 상태라 하더라도 전체적으로는 특정 개인을 식별할 수 있는 정보로 변환될 수 있다. 공공데이터는 신뢰성과 정확성이 높기 때문에 민간 데이터와 결합될 경우 재식별 가능성이 더욱 증가한다. 따라서 생성형 AI 시대의 개인정보 보호는 개별 데이터의 안전성뿐 아니라 데이터 결합 가능성 자체를 평가하는 방향으로 발전할 필요가 있다. 둘째, 추론(Inference) 위험으로 전통적인 개인정보 침해는 개인정보 자체가 유출되거나 공개되는 경우를 전제로 하였다. 그러나 생성형 AI는 특정 정보가 직접 제공되지 않더라도 다양한 데이터를 분석하여 새로운 정보를 추론할 수 있다. 예를 들어 특정 개인의 구매기록, 검색기록, SNS 활동정보만으로도 건강상태, 정치성향, 종교, 경제수준, 가족관계 등을 추정할 수 있다. 이러한 정보는 당사자가 공개한 적이 없음에도 불구하고 AI가 데이터 간의 상관관계를 분석하여 도출하는 결과이다. 생성형 AI는 인간이 쉽게 발견하기 어려운 패턴을 찾아내는 데 강점을 가지고 있다. 예를 들어 특정 지역에서 특정 약품을 반복적으로 구매한 기록과 병원 방문기록이 결합될 경우 희귀질환 여부를 추정할 수 있으며, 특정 도서 대출기록과 온라인 활동정보를 결합하면 정치적 성향이나 종교적 성향을 추론할 수도 있다. 문제는 이러한 추론이 반드시 사실이라고 단정할 수 없음에도 불구하고 현실에서 차별이나 불이익의 근거로 활용될 가능성이 있다는 점이다. 또한 정보주체는 자신에 관한 정보가 어떠한 방식으로 추론되었는지 알기 어렵기 때문에 개인정보 자기결정권의 침해가 발생할 수 있다. 결국 생성형 AI 시대의 개인정보 보호는

단순히 정보의 공개 여부를 통제하는 것만으로는 충분하지 않으며, AI에 의한 추론 가능성 자체를 평가하는 새로운 보호체계가 필요하다. 셋째, 집단정보 및 조직정보의 재식별 위험이다. 생성형 AI 시대의 재식별 위험은 더 이상 개인에 한정되지 않는다. 최근에는 특정 집단이나 조직에 관한 정보 역시 재식별의 대상이 되고 있다. 전통적인 개인정보 보호법은 개인을 보호의 중심으로 설계되어 있다. 그러나 AI는 개별 개인정보뿐 아니라 특정 집단이나 조직의 특성까지 분석할 수 있다. 예를 들어 특정 지역의 의료정보를 분석하면 지역사회의 건강상태를 추정할 수 있으며, 특정 학교 학생들의 데이터를 분석하면 학교 전체의 교육환경이나 학업성취도를 예측할 수 있다. 또한 특정 기업의 임직원 활동정보, 내부 업무패턴, 위치정보 등이 결합될 경우 기업의 영업전략이나 조직구조가 노출될 가능성도 존재한다. 이러한 경우에는 특정 개인이 식별되지 않더라도 조직 전체가 분석 대상이 되어 새로운 위험이 발생한다. 공공데이터가 생성형 AI의 학습자료로 활용되는 경우에는 특정 지역, 특정 직업군, 특정 연령집단에 대한 분석이 가능해진다. 이러한 분석은 사회적 편견이나 차별을 강화할 위험이 있으며, 특정 집단에 대한 낙인효과(Stigmatization)를 초래할 수도 있다. 따라서 생성형 AI 시대에는 개인정보 보호를 넘어 집단정보(Group Privacy)와 조직정보(Organizational Privacy)에 대한 새로운 보호 논의가 필요하다. 넷째, 국가안보정보 재식별 위험이다. 생성형 AI 시대 재식별 위험의 가장 심각한 형태는 국가안보정보의 재식별 위험이다. 대표적인 사례가 2018년 Strava Heat Map 사건이다. Strava는 전 세계 이용자의 GPS 운동기록을 집계한 Heat Map을 공개하였으나, 연구자들은 이를 분석하여 아프가니스탄, 시리아, 이라크 등에 위치한 미군기지와 군사시설의 위치 및 순찰경로를 식별할 수 있었다. 해당 데이터에는 개인의 이름이나 계정정보가 포함되어 있지 않았음에도 불구하고 위치정보와 이동패턴을 분석함으로써 군사시설의 구조와 운영형태를 추론할 수 있었던 것이다. 이 사건은 데이터 재식별 위험이 더 이상 개인정보 문제에 그치지 않고 국가안보 문제로 확장될 수 있음을 보여주었다. 특히 생성형 AI는 위성사진, 공개지도, 뉴스기사, SNS 게시물, 공공데이터 등을 자동으로 결합할 수 있기 때문에 과거보다 훨씬 높은 수준의 분석이 가능하다. 우리나라 역시 군 장병 건강관리 시스템, 공공기관 위치정보, 국가기반시설 운영정보 등 다양한 데이터가 디지털화되고 있다. 이러한 정보가 비식별 상태로 제공되더라도 생성형 AI가 이를 다른 공개정보와 결합할 경우 특정 부대의 활동 특성, 국가중요시설의 위치, 조직 운영방식 등이 추론될 가능성을 배제하기 어렵다. 결국 생성형 AI 시대의 재식별 위험은 개인정보 보호의 범위

를 넘어 국가안보와 공공안전의 문제로 확대되고 있다. 따라서 향후 데이터 개방정책은 개인정보 보호뿐 아니라 조직정보 보호와 국가안보 보호까지 고려하는 종합적인 위험평가 체계를 구축할 필요가 있다.

Ⅲ. 현행 영향평가 제도의 한계와 AI 데이터 영향평가의 필요성

1. 개인정보보호법상 개인정보 영향평가의 한계

개인정보 보호는 현대 정보사회에서 가장 중요한 법적 과제 중 하나로 평가되고 있다. 개인정보의 수집·이용·제공이 확대됨에 따라 개인정보 침해의 위험 역시 증가하고 있으며, 이에 따라 사전적 예방수단으로서 개인정보 영향평가 제도가 도입되었다.

우리나라 개인정보보호법 제33조는 공공기관이 일정 규모 이상의 개인정보파일을 구축·운영하는 경우 개인정보 영향평가를 실시하도록 규정하고 있다. 개인정보 영향평가(Privacy Impact Assessment, PIA)는 개인정보 처리로 인하여 발생할 수 있는 침해 요인을 사전에 분석하고 개선방안을 마련함으로써 개인정보 침해를 예방하기 위한 제도이다. 개인정보 영향 평가는 개인정보 보호를 위한 사후적 규제방식에서 벗어나 사전적 예방원칙(Preventive Principle)을 구현한 제도로서 중요한 의미를 가지는데 대규모 정보시스템 구축 이전에 개인정보 침해 가능성을 분석함으로써 시스템 설계단계부터 개인정보 보호를 반영하도록 하는 Privacy by Design 개념을 실현하는 수단으로 평가된다. 그러나 현행 개인정보 영향평가는 개인정보파일의 안전성 확보를 중심으로 설계되어 있어 생성형 AI 시대의 새로운 개인정보 위험인 재식별의 위험을 예방하는 기준에 대해서는 충분히 반영하지 못하고 있다. 개인정보보호법상 영향평가는 일정 규모 이상의 개인정보파일을 구축하거나 변경하는 공공기관을 대상으로 하는 것으로 평가대상은 주로 행정정보시스템, 복지정보시스템, 교육정보시스템, 조세정보시스템 등 공공기관이 운영하는 개인정보파일이다. 그리하여 개인정보 수집의 적법성과 필요성, 개인정보 보유·이용의 적정성, 개인정보 접근권한 및 접근통제의 적정성, 암호화 및 기술적·관리적 보호조치의 적정성, 개인정보 파기절차의 적정성을 평가한다. 즉, 현행 영향평가는 개인정보파일이 안전하게 관리되고 있는지 여부를 중심으로 평가하도록 구성되어 있다.

그러므로 현행 개인정보 영향평가는 정보시스템의 안전성 확보를 중심으로 설계되어 있기 때문에 생성형 AI 시대에 발생하는 재식별 위험을 충분히 반영하지 못한다는 한계를 가진다. 예를 들어 건강보험공단의 가명정보가 적법하게 제공되었다고 하더라도 SNS 정보, 위치정보, 뉴스기사, 공개된 공공데이터 등과 결합될 경우 특정 개인을 식별할 가능성이 존재한다. 그러나 이러한 경우에는 개인정보파일 자체가 유출된 것이 아니므로 현행 영향평가 체계에서는 위험으로 인식되지 않을 수 있다. 또한 생성형 AI는 다양한 데이터셋을 자동으로 결합하고 새로운 정보를 추론할 수 있기 때문에 기존의 비식별화 또는 가명처리만으로는 충분한 개인정보 보호가 어려워지고 있다. 그럼에도 불구하고 현행 영향평가 항목에는 재식별 가능성, 데이터 결합 위험, AI 추론 위험 등에 대한 독립적인 평가기준이 존재하지 않는다. 결국 개인정보 영향평가는 “개인정보파일이 안전한가”를 평가하는 제도일 뿐, “데이터가 재식별될 수 있는가”를 평가하는 제도는 아니라는 구조적 한계를 가진다.

2. AI 기본법상 고영향 AI 영향평가의 한계

「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」은 인공지능 기술의 발전을 촉진하는 동시에 국민의 권리와 안전을 보호하기 위하여 고영향 인공지능(High-Impact AI) 제도를 도입하고 있다. 고영향 AI란 개인의 권리와 의무 또는 안전에 중대한 영향을 미칠 수 있는 인공지능 시스템을 의미한다. 대표적으로 채용, 금융심사, 의료, 교육, 공공서비스 등 국민의 기본권에 직접 영향을 미치는 영역에서 활용되는 AI가 이에 해당한다. 고영향 AI는 일반적인 AI보다 높은 수준의 투명성과 책임성이 요구되며, 일정한 영향평가 의무를 부담하게 된다. 인공지능 기본법상 영향평가는 고영향 AI가 인간의 권리와 자유에 미치는 영향을 사전에 평가하기 위한 제도로서 평가대상은 주로 다음과 같은 사항이다. 차별 발생 가능성, 안전성 및 신뢰성, 설명가능성, 인간의 통제 가능성, 투명성, 즉, AI 시스템이 사회적으로 어떠한 영향을 미치는지를 중심으로 평가하는 구조를 가지고 있다. 그러나 인공지능 기본법상 영향평가는 인공지능 시스템에 대한 평가이지 인공지능 학습데이터에 대한 평가는 아니다. 예를 들어 의료 AI가 환자에게 차별을 발생시키는지 여부는 평가할 수 있지만, 해당 AI가 학습한 의료데이터가 다른 데이터와 결합될 경우 특정 환자를 재식별할 수 있는지는 평가하지 않는다. 마찬가지로 공공데이터를 활용한 생성형 AI가 사회적으로 안전하게 운영되는지 여부는 평가할 수 있으나, 해당 공공데이터 자

체가 재식별될 위험이 있는지는 평가대상이 아니다. 따라서 인공지능 기본법상 영향평가는 인공지능 서비스 영향평가로서의 기본권 침해 여부를 판단하는 의미는 있으나 데이터 재식별 위험을 통제하기 위한 제도로는 한계를 가진다.

3. GDPR 제35조 데이터보호영향평가(DPIA)가 주는 시사점

EU GDPR 제35조는 개인정보 처리행위가 정보주체의 권리와 자유에 고위험을 초래할 가능성이 있는 경우 데이터보호영향평가(Data Protection Impact Assessment, DPIA)를 실시하도록 규정하고 있다.[24] GDPR 제35조는 새로운 기술을 활용한 개인정보 처리로 인해 정보주체의 권리와 자유에 높은 위험이 발생할 가능성이 있는 경우 데이터 보호 영향평가를 의무화하고 있는데[25] 특히 자동화된 의사결정, 프로파일링, 대규모 개인정보 처리 등을 고위험 처리행위로 규정함으로써 사전적 위험관리 체계를 구축하고 있다.[26]

감독기관은 데이터보호영향평가 실시가 필요한 처리행위의 목록을 작성하고 이를 공개하여야 할 뿐 아니라 꼭 필요하지 않은 처리행위의 목록도 작성하게 할 수 있는데[27] 이 때 (a) 처리 목적 및 처리행위에 대한 체계적 설명 (b) 처

---

24) GDPR Official Text, Regulation (EU) 2016/679, Article 35.

25) Article 35(1) Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. 새로운 기술을 사용하는 경우를 포함하여 개인정보 처리의 성격, 범위, 맥락 및 목적을 고려할 때 해당 처리행위가 자연인의 권리와 자유에 높은 위험을 초래할 가능성이 있는 경우, 개인정보처리자는 그 처리 이전에 예정된 처리행위가 개인정보 보호에 미치는 영향을 평가하여야 한다.

26) Article 35(3) 다음의 경우에는 특히 DPIA를 실시하여야 한다.(a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person; 프로파일링을 포함한 자동화된 처리에 기초하여 개인의 특성을 체계적·광범위하게 평가하고, 그 결과가 법적 효과를 발생시키거나 이에 준하는 중대한 영향을 미치는 경우 (b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; GDPR 제9조 제1항의 민감정보 또는 제10조의 범죄경력 관련 정보를 대규모로 처리하는 경우 (c)a systematic monitoring of a publicly accessible area on a large scale. 공공에게 개방된 장소를 대규모로 체계적으로 감시하는 경우

27) Article 35(4) The supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment. 감독기관은 DPIA 실시가 필요한 처리행위의 목록을 작성

리의 필요성과 비례성 평가 (c) 정보 주체의 권리와 자유에 대한 위험 평가 (d) 위험 완화를 위한 보호조치, 보안조치 및 기술적·관리적 조치 사항은 포함되어야 한다.[28]

우리나라는 현재 생성형 AI의 학습 과정에서 발생할 수 있는 개인정보 재식별 위험, 민감정보 추론 위험 및 대규모 데이터 결합 위험 역시 이러한 고위험 처리에 해당할 수 있으므로, 우리나라도 개인정보보호법상 「AI 데이터 영향평가」 제도를 도입하여 공공데이터 개방 이전에 재식별 가능성, AI 추론 가능성, 민감정보 노출 위험 등을 사전적으로 평가하는 제도적 장치를 마련할 필요가 있다. 데이터보호영향평가는 개인정보 보호를 위한 사전적 예방제도라는 점에서 우리나라의 개인정보보호법 제33조 개인정보 영향평가와 유사한 측면이 있으나, 평가대상과 평가범위에서 중요한 차이를 가진다. GDPR 제35조는 개인정보 처리행위 자체가 정보주체에게 미치는 위험을 평가하도록 함으로써 개인정보 보호를 보다 실질적으로 구현한다. 그리하여 데이터가 개인정보보호, 비식별처리했다고 하더라도 이후 재식별처리를 할 수 있는지 위험을 보다 확인하고자 한다. 특히 35조 3항에서 의무화하고 있는 데이터보호영향평가로는 첫째, 자동화된 의사결정 및 프로파일링, 둘째, 민감정보의 대규모 처리, 셋째, 공개장소에 대한 체계적 감시, 넷째, 새로운 기술의 활용이므로 AI, 빅데이터, 위

하고 이를 공개하여야 한다. Article 35(5) The supervisory authority may also establish and make public a list of the kind of processing operations for which no data protection impact assessment is required. 감독기관은 DPIA가 필요하지 않은 처리행위의 목록도 작성하여 공개할 수 있다.

28) Article 35(7) Data Protection Impact Assessment, The assessment shall contain at least: (a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
(b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes; (c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and (d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.제35조 제7항 영향평가에는 최소한 다음 사항이 포함되어야 한다. (a)예정된 개인정보 처리행위와 처리 목적에 대한 체계적 설명. 필요한 경우 개인정보처리자가 추구하는 정당한 이익(legitimate interest)을 포함한다. (b) 해당 처리 목적에 비추어 처리행위의 필요성 및 비례성에 대한 평가.(c)제1항에서 언급한 정보주체의 권리와 자유에 대한 위험성 평가.(d)위험에 대응하기 위한 조치에 관한 내용. 여기에는 보호조치(safeguards), 보안조치(security measures), 개인정보 보호를 보장하기 위한 메커니즘 및 GDPR 준수 입증 수단이 포함되며, 정보주체 및 기타 관련자의 권리와 정당한 이익을 고려하여야 한다.

치정보 분석, 의료데이터 활용과 같은 기술은 대표적인 데이터보호영향평가 대상에 해당한다.

이는 재식별 위험 평가체계로서 우리나라 개인정보보호법 제33조의 한계로서 GDPR 제35조가 재식별 위험을 포함한 데이터 처리 위험 자체를 평가한다는 점이다. 이 때 평가항목에는 데이터 결합 가능성, 프로파일링 위험, 개인정보 추론 가능성, 재식별 가능성, 정보주체의 권리 침해 가능성 등이 있는데 GDPR은 개인정보가 유출되었는지 여부만이 아니라 개인정보가 재식별될 수 있는지 여부까지 평가대상으로 삼고 있다.

4. 인공지능 데이터 영향평가(재식별 영향평가) 제도의 도입방안

GDPR 제35조는 생성형 AI 시대 개인정보 보호를 위한 중요한 시사점을 제공한다. Sweeney 사건, Netflix 사건, AOL 사건, Strava 사건 등은 개인정보 유출이 아니라 데이터 결합과 재식별이 문제였다는 점에서 GDPR 방식의 접근이 더욱 적절하다고 볼 수 있다. 현행 개인정보 영향평가와 AI 기본법상 영향평가는 각각 개인정보파일과 AI 시스템을 평가하는 제도이다. 그러나 생성형 AI 시대에 가장 중요한 위험요소인 데이터 자체의 재식별 가능성은 충분히 평가하지 못하고 있다. 특히 공공데이터 개방과 AI 학습데이터 활용이 확대되는 상황에서 데이터 자체에 대한 독립적인 영향평가 제도가 요구된다.

생성형 AI는 데이터 결합과 추론 능력을 통하여 과거에는 존재하지 않았던 새로운 위험을 발생시키고 있다. 대표적으로 건강정보, 위치정보, 검색기록, 공공데이터가 결합될 경우 특정 개인의 건강상태, 정치성향, 종교, 경제수준 등을 추론할 수 있다. 또한 Strava 사건에서 보듯이 조직정보와 국가안보정보까지 재식별될 가능성이 존재한다. 재식별 위험은 생성형 AI 시대 개인정보 침해의 핵심 요소이다. 과거에는 개인정보 유출 여부가 중요한 문제였다면, 오늘날에는 데이터가 적법하게 제공되었더라도 AI가 이를 분석하여 특정 개인을 식별할 수 있는지가 중요한 문제로 등장하고 있다. 따라서 공공데이터 개방 이전 단계에서 재식별 가능성을 평가하고 위험을 완화할 수 있는 제도적 장치가 필요하다. AI 데이터 영향평가는 GDPR 제35조의 DPIA를 참고하되 생성형 AI 시대의 특수성을 반영한 새로운 영향평가 제도로 설계될 필요가 있다.

현행 개인정보보호법 제33조는 개인정보 영향평가 제도를 규정하고 있으나, 이는 개인정보파일 구축에 따른 침해요인을 평가하는 제도이다. 생성형 AI 시대의 재식별 위험을 반영하기 위해서는 별도의 AI 데이터 영향평가 규정을 신

설할 필요가 있다. 따라서 개인정보보호법에 제33조의2를 신설하여 공공데이터 개방, 가명정보 결합, 생성형 AI 학습데이터 구축 등의 경우 AI 데이터 영향평가를 의무화하는 방안을 검토할 수 있다. AI 데이터 영향평가의 적용대상은 재식별 위험이 높은 데이터 처리행위로 한정하는 것이 바람직하다. 대표적으로 다음과 같은 경우가 포함될 수 있다. 첫째, 생성형 AI 학습을 위한 대규모 데이터 구축, 둘째, 공공데이터와 민간데이터의 결합, 셋째, 건강정보·위치정보·생체정보·유전정보 활용, 넷째, 가명정보 결합전문기관을 통한 데이터 결합, 다섯째, 국가중요시설 및 군 관련 데이터 활용이다. 이와 같은 고위험 데이터 처리행위에 대하여 의무적으로 AI 데이터 영향평가를 실시하도록 함으로써 재식별 위험을 사전에 통제할 수 있을 것이다. 평가주체는 원칙적으로 데이터를 활용하는 개인정보처리자로 하되, 일정 규모 이상의 고위험 사업은 외부 전문기관의 검증을 받도록 할 필요가 있다. 특히 공공기관의 경우 개인정보보호위원회가 지정하는 전문평가기관의 검토를 거치도록 하고, 민간기업의 경우에도 일정 규모 이상의 생성형 AI 개발사업에 대해서는 외부 전문가의 평가를 의무화하는 방안을 고려할 수 있을 것이다. 그리하여 AI 데이터 영향평가는 ① 데이터 활용계획 수립 ② 재식별 위험분석 ③ AI 추론 가능성 분석 ④ 조직 및 국가안보 영향분석 ⑤ 위험완화조치 마련 ⑥ 개인정보보호위원회 또는 전문기관 심사 ⑦ 승인·조건부 승인·보완·불승인 결정 과 같은 절차를 통하여 공공데이터 개방과 AI 산업 발전을 저해하지 않으면서도 개인정보 침해 위험을 효과적으로 관리할 수 있을 것이다.

Ⅳ. 결론

생성형 AI의 발전은 개인정보 보호에 관한 기존 법체계가 전제하였던 위험 구조를 근본적으로 변화시키고 있다. 과거에는 개인정보 유출이나 해킹과 같은 보안침해가 주요 위험요소였다면, 오늘날에는 적법하게 제공된 데이터가 다른 데이터와 결합되거나 인공지능에 의해 분석·추론되는 과정에서 새로운 개인정보 침해가 발생할 수 있다. 특히 공공데이터 개방정책이 확대되고 생성형 AI의 학습데이터 수요가 증가함에 따라 가명정보나 비식별정보조차 재식별될 가능성이 높아지고 있다.

그러나 현행 개인정보보호법상 개인정보 영향평가는 개인정보파일의 안전성

확보를 중심으로 설계되어 있으며, AI 기본법상 고영향 AI 영향평가는 인공지능 시스템의 사회적 영향을 중심으로 평가하고 있다. 따라서 데이터 결합에 의한 재식별 가능성이나 생성형 AI의 추론 위험을 직접 평가할 수 있는 제도는 부재한 상황이다.

이에 본 연구는 GDPR 제35조의 데이터보호영향평가(DPIA)를 참고하되 생성형 AI 시대의 특성을 반영한 새로운 제도로서 「AI 데이터 영향평가(AI Data Impact Assessment)」 또는 「재식별 영향평가(Re-identification Impact Assessment)」 제도의 도입을 제안하였다. AI 데이터 영향평가는 개인정보파일 자체나 AI 시스템만을 평가하는 것이 아니라, 데이터가 생성형 AI와 결합되는 과정에서 발생할 수 있는 재식별 위험을 평가하는 제도이다. 특히 데이터 결합 가능성, 재식별 가능성, AI 추론 가능성, 민감정보 노출 가능성, 조직정보 및 국가안보정보 재식별 가능성 등을 종합적으로 검토함으로써 생성형 AI 시대에 적합한 새로운 개인정보 보호체계를 구축하는 것을 목표로 한다.

구체적으로 본 연구는 개인정보보호법에 「제33조의2(AI 데이터 영향평가)」를 신설하여 공공데이터 개방, 가명정보 결합, 생성형 AI 학습데이터 구축 등 일정한 고위험 데이터 처리행위에 대하여 영향평가를 의무화하는 방안을 제안하였다. 또한 평가항목으로 데이터 결합 가능성, 재식별 가능성, AI 추론 가능성, 조직 및 국가안보 위험 등을 포함하는 표준 평가모델을 제시하였다. 이러한 제도는 공공데이터 개방과 AI 산업 발전을 저해하기 위한 규제가 아니라, 데이터 활용과 개인정보 보호가 조화를 이루도록 하기 위한 예방적 안전장치로 이해되어야 한다.

결국 생성형 AI 시대의 개인정보 보호는 더 이상 개인정보 유출 방지에만 머물러서는 안 된다. 앞으로의 개인정보 보호정책은 데이터가 AI와 결합될 때 발생하는 새로운 위험을 사전에 예측하고 통제하는 방향으로 발전하여야 한다. 특히 공공데이터 개방정책이 지속적으로 확대되는 상황에서 개인정보 보호와 AI 산업 발전의 균형을 확보하기 위해서는 데이터 중심의 새로운 영향평가 체계가 필수적이다. 본 연구에서 제안한 AI 데이터 영향평가 제도는 이러한 변화에 대응하기 위한 하나의 입법적 대안으로서 의의를 가지며, 향후 개인정보보호법과 AI 관련 법제의 발전 과정에서 중요한 논의의 출발점이 될 수 있을 것으로 기대한다.

궁극적으로 생성형 AI 시대의 데이터 거버넌스는 “얼마나 많은 데이터를 개방할 것인가”의 문제가 아니라 “어떻게 안전하게 개방할 것인가”의 문제로 전환되어야 한다. 공공데이터의 개방과 활용은 지속적으로 확대되어야 하지만,

그 과정에서 국민의 개인정보 자기결정권과 정보인권이 침해되어서는 안 된다. 따라서 공공데이터 개방과 AI 산업 발전, 그리고 개인정보 보호의 조화를 실현하기 위한 새로운 법적 장치로서 AI 데이터 영향평가 제도의 도입이 적극적으로 검토될 필요가 있다.

[참고문헌]

1. 국내문헌

강태경 · 유진, 「코로나19 위기관리정책의 인권 형사정책적 개선방안 연구」, 한국형사정책연구원, 2021.

경 건, 「한국의 공공데이터 개방의 법제와 쟁점」, 「행정법연구」 제47권, 행정법이론실무학회, 2016.

권은정, 「공공데이터 영역 리스크 관리에 관한 법적 소고」, 「행정법연구」 제60권, 행정법이론실무학회, 2020.

김경대, 「생성형 AI의 개인정보 침해에 대한 법적 대응방안 - 개인정보보호법을 중심으로」, 「지식융합연구」 제8권 제2호, 2025.

김지희, 「보건의료 빅데이터의 활용과 개인정보보호」, 경인문화사, 2022.

남기연 · 정현, 「생성형 AI 음악물의 권리침해에 관한 법적 연구」, 「스포츠엔터테인먼트와 법」 제27권 제2호, 2024.

2. 국외문헌

Carlini, Nicholas et al., “Extracting Training Data from Large Language Models,” Proceedings of the USENIX Security Symposium, 2021.

Carlini, Nicholas et al., “Quantifying Memorization Across Neural Language Models,” International Conference on Learning Representations (ICLR), 2023.

Narayanan, Arvind & Shmatikov, Vitaly, “Robust De-anonymization of Large Sparse Datasets,” Proceedings of the 2008 IEEE Symposium on Security and Privacy, IEEE Computer Society, 2008.

Pandurangan, Vijay, “On Taxis and Rainbows: Lessons from NYC’s Improperly Anonymized Taxi Logs,” Medium, June 21, 2014.

Sweeney, Latanya, Simple Demographics Often Identify People Uniquely, Data Privacy Working Paper No. 3, Carnegie Mellon University, 2000.

Tockar, Anthony, “Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset,” Neustar Research Blog, September 15, 2014.

3. 법령

「공공데이터의 제공 및 이용 활성화에 관한 법률」

「개인정보 보호법」
「감염병의 예방 및 관리에 관한 법률」
「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」
Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation, GDPR).

4. 판례

헌법재판소 2019. 7. 25. 선고 2017헌마1329 결정. Landwehr v. AOL Inc., No. 1:11-cv-01014-CMH-TRJ (E.D. Va. May 24, 2013).

5. 인터넷 자료

BBC News, “ChatGPT Bug Exposed Users’ Chat Histories,” March 24, 2023.
BBC News, “Strava Fitness App Can Reveal Military Bases, Analysts Say,” January 29, 2018.
Barbaro, Michael & Zeller Jr., Tom, “A Face Is Exposed for AOL Searcher No.4417749,” The New York Times, August 9, 2006.
Nathan Ruser, X(Twitter) Post, January 27, 2018.
OpenAI, “March 20 ChatGPT Outage: Here’s What Happened,” March 24, 2023.
Reuters, “OpenAI Says ChatGPT Bug Exposed Some User Information,” March 24, 2023.
Royal Free London NHS Foundation Trust, “Royal Free - Google DeepMind Trial Failed to Comply with Data Protection Law.”
박수선, 「확진자 동선 보도 ‘성소수자 혐오’ 우려 큰데...교회언론회 ‘공익 보도였다’」, 「PD저널」, 2020. 5. 11.

[Abstract]

A Study on Personal Data Re-identification Risks in Public Data Disclosure and Generative AI:
Toward the Introduction of an AI Data Impact Assessment*

Park Jeong In** · Jung Hyun***

The rapid advancement of Generative Artificial Intelligence (Generative AI) has significantly increased the value of public data while simultaneously creating new legal risks that existing personal data protection frameworks are not fully equipped to address. In particular, as public data are increasingly utilized as training data for generative AI, pseudonymized or de-identified data may be combined with other datasets or analyzed through AI inference techniques, thereby enabling the re-identification of specific individuals or organizations. Consequently, the risk of re-identification has emerged as a critical legal issue in the AI era. However, the current Personal Information Protection Act primarily focuses on the security of personal information files through the Personal Information Impact Assessment (PIA), while the Framework Act on Artificial Intelligence mainly evaluates the impact of high-impact AI systems on fundamental rights and freedoms. Neither framework sufficiently addresses the risks arising from data linkage, re-identification, and AI-driven inference.

This study aims to analyze the characteristics and legal implications of re-identification risks associated with the disclosure of public data and the use of generative AI, examine the limitations of existing impact assessment systems, and propose the introduction of a new AI Data Impact Assessment framework. To this end, the study reviews the legal framework governing

* This paper is a result of the National Research Foundation of Korea's Humanities and Social Sciences Research Professor Support Program.
** Lead Author: Research Professor, AI DYNAINFO Institute, Duksung Women's niversity, Ph.D. in Law
*** Co-author: Visiting Professor, Department of Law, Dankook University, Ph.D. in Law

public data disclosure and personal information protection, analyzes representative cases—including the Sweeney case, the Netflix Prize case, the AOL Search Log case, the New York City Taxi Dataset case, the Strava Heat Map case, and Korean healthcare data utilization—and comparatively examines the Data Protection Impact Assessment (DPIA) under Article 35 of the General Data Protection Regulation (GDPR).

The findings indicate that privacy risks in the era of generative AI extend beyond conventional data breaches to encompass data linkage, AI-based inference, and the re-identification of individuals, groups, and organizations. These emerging risks cannot be adequately managed under the current Personal Information Impact Assessment or High-Impact AI Impact Assessment systems. As public data disclosure and AI-driven data utilization continue to expand, a more comprehensive ex ante risk management framework is required to protect not only personal information but also organizational information and national security interests.

Accordingly, this study proposes the introduction of an AI Data Impact Assessment (AI-DIA), or alternatively a Re-identification Impact Assessment (RIA), modeled on the GDPR's Data Protection Impact Assessment. The proposed framework would apply to high-risk data processing activities, including public data disclosure, the combination of pseudonymized data, and the construction of AI training datasets. It would assess re-identification risks, data linkage risks, AI inference capabilities, and the potential exposure of sensitive information prior to data processing. Such a framework is expected to establish an effective balance between public data utilization and privacy protection while contributing to the development of a governance model suitable for the generative AI era.

Keywords : Public Data, Generative Artificial Intelligence, Personal Information Protection, Re-identification, AI Data Impact Assessment (AI-DIA), Re-identification Impact Assessment (RIA), Data Protection Impact Assessment (DPIA), General Data Protection Regulation (GDPR), Pseudonymized Data, Data Governance