생성형 AI의 위협과 개인정보 자기통제권 보호 방안

장재영
한국인터넷진흥원 개인정보안전활용본부 연구위원
jangjy@kisa.or.kr

생성형 AI의 위협

2025년 1월, 중국의 인공지능(Artificial Intelligence, AI) 스타트업 DeepSeek가 대규모 언어 모델(Large Language Model, LLM)인 DeepSeek V3 기반의 R1 모델을 출시했다. 이 모델은 ChatGPT보다 저비용·고학습 방식을 적용해 ChatGPT의 o1과 유사한 성능을 낼 수 있다고 평가받고 있다. 개발사에 따르면, DeepSeek V3 모델의 개발 비용은 약 557.6만 달러로, ChatGPT-4의 추정 개발 비용인 1억 달러 대비 17.9% 수준이라고 한다(김지수, 홍민지, 2025.03.01.). DeepSeek의 성공은 막대한 투자 없이는 경쟁력 있는 LLM 개발이 어려울 것이라는 업계의 기존 인식을 뒤집어 놓기에 충분했다.

DeepSeek는 시장에 신선한 충격을 주었지만 동시에 심각한 공포도 함께 불러 일으켰다. DeepSeek의 R1 모델은 보안 취약성이 높아 AI 모델 탈옥(Jail Breaking[1]) 발생률이 상대적으로 높은 것으로 평가받고 있다. GPT-4, Gemini-1.5 pro, Claude-3.5, Sonnet 등의 주요 경쟁 모델에 비해 보안 공격에 취약하다는

1 Jail Breaking: 탈옥을 의미하는 영어 단어로 컴퓨터 시스템이나 장치의 제한을 극복하기 위한 표현(출처: 위키피디아)

분석도 있다(김관영 등, 2025; Paul Kassianik, Amin Karbasi, 2025.01.31.). 또한, DeepSeek는 OpenAI의 데이터 무단 수집으로 인한 약관 위반 의혹도 받고 있다(Stephen Nellis et al., 2025.01.30.). 더 큰 문제는 DeepSeek의 과도한 이용자 정보 수집이다. DeepSeek는 이용자의 계정 정보, IP 주소, 기기 모델, 운영 체제뿐만 아니라 키보드 입력 패턴까지 수집하는 것으로 알려졌다. DeepSeek가 수집하고 있는 정보의 종류는 이용자가 생각하는 것보다 매우 광범위하고 구체적이며, 다른 생성형 AI 모델과 비교해도 지나치게 많다(<표 1> 참조). 특히, 수집된 정보가 바이트댄스 등 특정 기업에 이용자의 동의 없이 제공된다는 지적도 있다(송혜리, 2025.02.19.). 또한, DeepSeek가 중국 기업이라는 점에서 중국 '국가정보법' 제7조[2]에 따라 이용자의 정보가 언제든지 중국 정부에 제공될 수 있다는 우려도 있다(김태성, 정호준, 2025.02.06.; 박민숙, 이효진, 2022). DeepSeek의 이러한 개인정보 수집 이슈는 프라이버시 및 개인정보 자기통제권 침해, 불필요한 광고 및 타깃 마케팅 등 기업의 데이터 오남용은 물론 국가의 감시 등 개인의 평온을 깨뜨리는 민감한 위협 요인이 될 수 있다.

표 1 - DeepSeek 데이터 수집 종류

구분		주요 내용
귀하가 제공하는 정보	프로필 정보	생년월일, 사용자 이름, 이메일 주소, 전화번호, 비밀번호 등
	사용자 입력	텍스트, 오디오 입력, 프롬프트, 업로드된 파일, 피드백, 채팅 기록, 기타 콘텐츠 등
	당사 연락 정보	신원 또는 연령 증명, 서비스 이용에 대한 피드백 또는 문의, 당사 서비스 약관 또는 기타 정책의 잠재적 위반에 대한 정보 등
자동으로 수집된 정보	기술 정보	장비 모델, 운영 체제, 키 입력 패턴 또는 리듬, IP 주소 및 시스템 언어, 충돌 보고서 및 성능 로그, 서비스 관련 진단 및 성능 정보 등
	사용 정보	사용하는 기능 및 수행하는 작업과 같은 서비스 사용 정보 등
	쿠키	쿠키, 페이지가 조회된 시간 및 날짜, 픽셀 태그가 배치된 페이지에 대한 설명 및 컴퓨터 또는 기기의 유사한 정보 등
	결제 정보	주문 배치, 결제, 고객 서비스, 애프터 서비스 등
다른 출처 정보	로그인, 가입, 연결 서비스	Apple, Google 등 타사 서비스 연결을 위한 액세스 토큰 등
	광고 측정 및 기타 파트너	광고용 모바일 식별자, 해시된 이메일 주소 및 전화번호, 쿠키 식별자 등

출처: DeepSeek의 프라이버시 정책(2025.02.14.), 김관영 등(2025)

현재 우리는 중국의 DeepSeek에 우려하고 있지만 생성형 AI의 개인정보 침해 문제는 ChatGPT의 서비스 초기에도 유사하게 발생했다. OpenAI의 ChatGPT 3.5 유료 버전 사용자 정보가 타인에게 공개돼 유럽의 일반 개인정보 보호 규정(General Data Protection Regulation, GDPR[3]) 위반 이슈가 발생한 적이

2 국가정보법 제7조: 모든 조직과 시민은 국가의 정보 활동을 지지, 지원, 협력하여야 한다.

3 REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

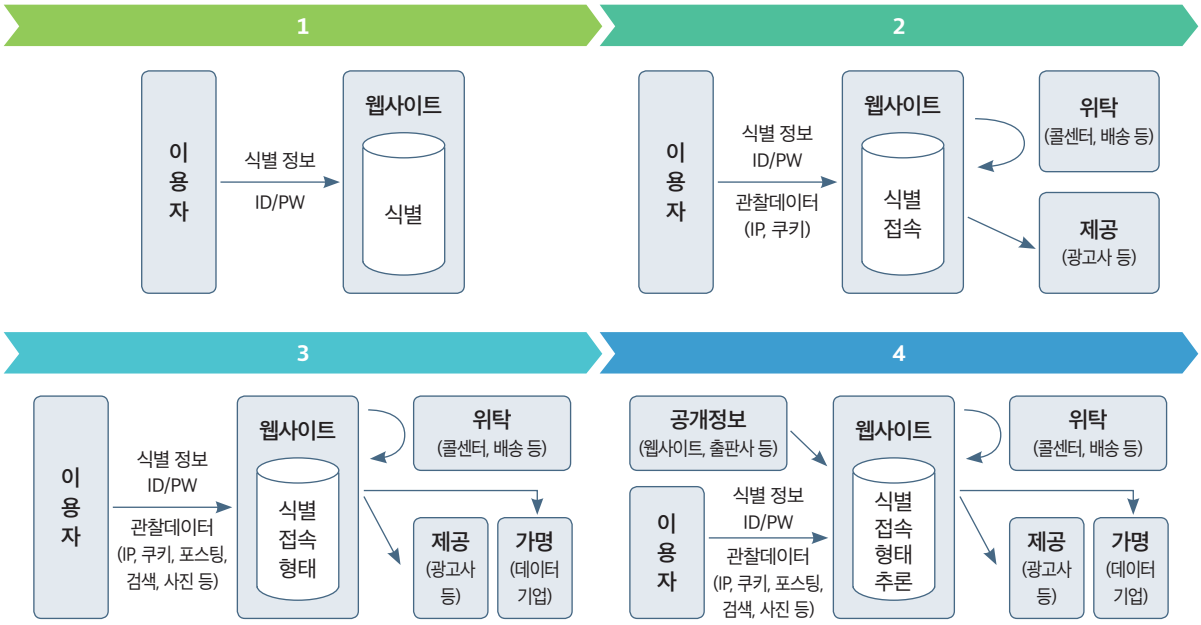
있다(김도원 등, 2023). 우리나라도 개인정보 유출 신고 의무 위반으로 OpenAI에 과태료를 부과한 사례가 있다(개인정보위, 2023.07.27.)

이처럼 전 세계가 생성형 AI의 혁신에 열광하고 있지만, 동시에 개인정보 침해 우려도 함께 제기되고 있다. 앞으로도 LLM을 기반으로 한 생성형 AI가 발전하면 할수록, 새로운 혁신이 창조되면 될수록, 생성형 AI로 인한 개인의 식별 가능성은 증가할 것이기 때문에 개인정보 보호 이슈가 부각될 수밖에 없다. AI가 일반화되는 시대에 왜 생성형 AI에서 개인정보가 이슈가 될 수밖에 없고 어떻게 하면 개인정보가 적절히 보호될 수 있을지에 대해 함께 고민해 볼 필요가 있다.

개인정보 이용의 발전 단계

정보통신 기술의 발전은 개인정보의 이용 방식에도 변화를 가져왔다. 지난 수십 년간 개인정보 활용 방식은 처리되는 개인정보의 종류와 양이 증가하는 방향으로 발전해 오고 있다. 개인정보 처리의 궁극적인 목표는 개인을 정확히 프로파일링(profiling)해 조직의 이익을 극대화하는 것이다. 조직의 이러한 요구는 생성형 AI를 비롯한 AI가 가장 효과적으로 수행할 수 있는 영역 중 하나이다. 따라서 앞으로도 AI는 개인정보 처리의 효율성을 높이기 위해 다양한 분야에서 활용될 것으로 기대된다. 개인정보 처리의 발전 단계는 <그림 1>에 도식화해 놓았다.

그림 1 - 개인정보 처리의 발전 단계



초기 단계: 개인 식별을 위한 활용

정보통신 서비스 도입 초기에는 개인정보가 주로 웹사이트 등에서 개인 확인 용도로 활용됐다. 이용자는 웹사이트, 포털, 이메일, 온라인 카페 등의 서비스 이용을 위해 이름, 주민등록번호, 전화번호 등과 같은 개인 식별 정보를 제공하고, 아이디와 비밀번호를 생성해 로그인했다. 국내 포털 서비스인 네이버, 다음 등은 회원가입 시 실명 인증을 요구했으며, 일부 웹사이트는 주민등록번호를 활용한 본인 확인 절차를 거쳤다. 이 시기에는 기업이 개인정보를 활용하는 방식이 제한적이었다.

개인정보의 내부 활용 및 제3자 제공 단계

인터넷 서비스가 발전하면서 기업은 개인정보를 단순한 식별 수단을 넘어 내부 업무 효율화 및 비즈니스 확장을 위해 활용하기 시작했다. 이 시기에 개인정보 취급위탁[4]과 제3자 제공[5]이 본격화됐다. 예를 들어, 이동통신 3사는 고객의 가입 정보 및 사용 패턴을 분석해 금융사, 보험사와 제휴한 맞춤형 마케팅을 제공하고 있으며, 카드사와의 협업을 통해 통신비 자동이체 시 추가 혜택을 제공하는 마케팅을 진행하고 있다. 이 시기부터 개인정보는 단순한 고객 관리용에서 벗어나 비즈니스 자산으로 활용하는 전환점이 됐다.

온라인 행태 분석 및 프로파일링 단계

디지털 광고 기술이 발전하면서 기업은 이용자의 행태 데이터를 수집·분석해 맞춤형 광고를 제공하기 시작했다. 이 시기에 온라인 행동 기반 광고(Online Behavioral Advertising)가 본격적으로 등장했다. 구글의 애드워즈(AdWords, 현재의 구글 애즈)가 대표적인 사례다. 기업은 사용자의 웹사이트 방문 기록, 클릭 패턴, 검색 키워드 등의 데이터를 기반으로 개인별 관심사를 분석했다. 이커머스 기업 또한 사용자의 구매 이력을 분석해 추천 알고리즘을 고도화했다. 이 단계의 중요한 차별점은 기업이 이용자가 직접 제공한 정보뿐만 아니라, 사용자의 행동 데이터를 기반으로 새로운 개인정보를 생성하기 시작했다는 점이다.

다양한 데이터를 결합한 개인정보 생성·예측 초고도화 단계

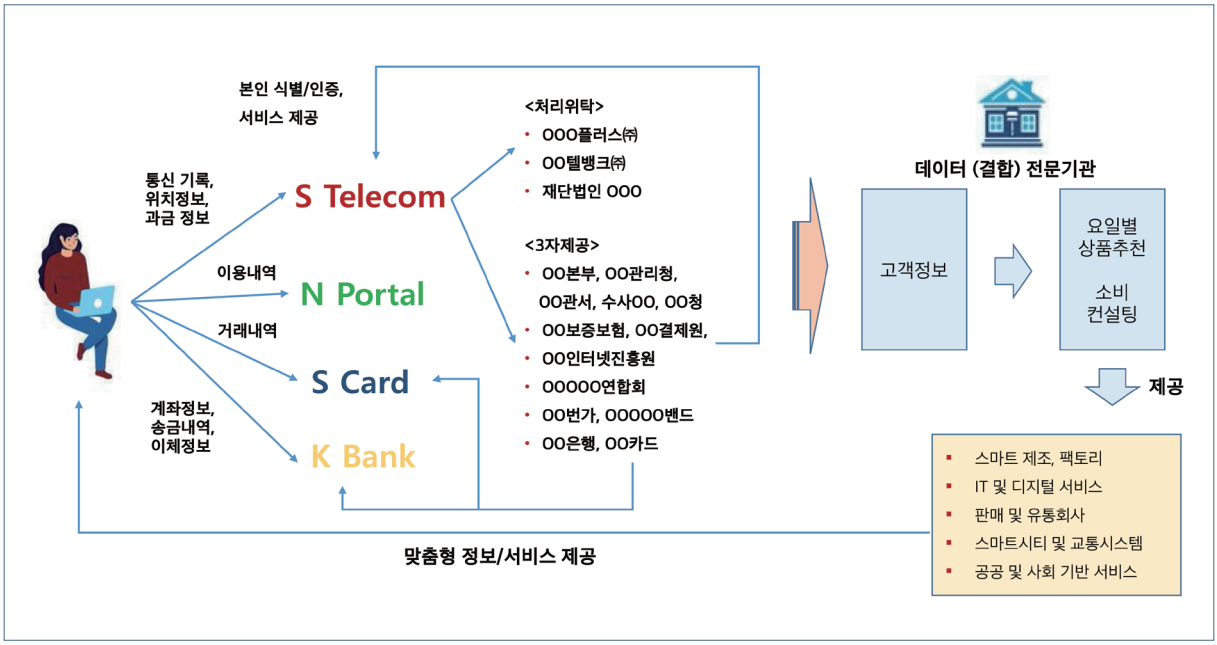
현재 기업은 개인이 제공하지 않은 정보까지 수집해 개인 식별 가능성을 고도화하고 있다. 대표적인 것이 LLM이다. LLM을 만드는 생성형 AI 기업은 데이터 기업이나 데이터 결합 기관 등에서 수집한 정보뿐만 아니라, 온라인에서 확보한 다양한 정보까지 무차별적으로 수집해 새로운 (개인)정보를 생성하고 있다. 또한, 데이터 결합 전문기관의 출현으로 기업은 개인의 동의 없이도 합법적으로 개인 데이터를 활용할

4 취급위탁: 위탁 기관의 이익을 위해 고객 데이터 관리, 고객센터 운영 등의 업무를 외부 업체(콜센터, IT 아웃소싱 업체 등)에 맡기는 방식

5 제3자 제공: 제3자의 이익을 위한 기업 간 제휴, 공동 마케팅, 공동 이벤트 등을 위해 개인정보를 타 기업에 이전하는 방식

수 있게 됐다. 이를 통해 AI 기업은 정보주체[6]보다 개인을 더 잘 이해하는 시대가 됐다. 개인을 보다 더 잘 이해하려는 욕구는 조직의 생존과 직결된 문제이다. 따라서 기술이 발전하면 할수록 개인에게 최적화된 맞춤형 서비스를 제공하려 할 것이고 그러려면 개인정보를 보다 구체적으로 많이 수집할 수밖에 없다(<그림 2> 참조).

그림 2 - 이동통신 서비스 이용자의 개인정보 처리 흐름



출처: 저자 작성(SKT의 개인정보 처리방침을 고려하여 작성)

기업 등이 개인의 정보를 활용할 수 있는 것은 개인정보를 제공받은 기업이 정보주체와 약속한 목적 내에서 안전하게 사용할 것이라는 신뢰에 기인한다(권영준, 2024). 이러한 신뢰 기반 위에 개인정보 활용 방식은 단순한 식별 수단 → 제3자 제공 → 프로파일링 → 생성형 AI를 통한 개인정보 생성·예측 초고도화 단계로 진화해 왔다. 그러나 현재는 정보주체가 제공하지 않은 정보까지 생성형 AI 학습에 이용함으로써 정보주체의 권리인 자기정보통제권이 약화될 것이라는 우려가 그 어느 때보다 높다. 따라서 LLM을 기반으로 한 생성형 AI에서 개인정보 보호의 핵심인 정보주체의 자기정보통제권이 왜 문제가 되는지를 살펴볼 필요가 있다.

6 정보주체: 처리되는 정보에 의하여 알아볼 수 있는 사람으로서 그 정보의 주체가 되는 사람(출처: 개인정보 보호법 제2조(정의) 제3호)

생성형 AI 시대, 개인정보 자기통제권의 위기

생성형 AI 시대는 개인정보 보호 패러다임의 근본적인 변화를 요구하고 있다. 현행 개인정보 보호 체계는 정보주체가 제공한 자신의 개인정보를 효과적으로 통제할 수 있다는 믿음을 전제로 하고 있다. 그러나 생성형 AI는 이용자가 제공하지 않은 정보도 학습하고, 이용자가 모르는 이용자에 관한 정보까지 생성하고 있으며, 이러한 정보의 존재 자체를 이용자가 직관적으로 인식하지 못할 가능성이 높기 때문에 개인정보 자기통제권의 실효적 행사가 본질적으로 제한될 위험이 높다. 이러한 환경에서 정보주체가 신뢰를 갖고 자신의 개인정보 관리 권한을 제3자에게 넘긴다는 것은 우려되는 일일 수밖에 없다. 또한 LLM의 기술적 특성상 이용자의 통제권을 즉시 행사하기 어려운 기술적 한계도 존재한다.

생성형 AI가 생성한 개인정보에 대한 정보주체의 통제권 약화

생성형 AI는 LLM의 학습을 위해 온라인에서 수집한 데이터와 제3자로부터 확보한 대량의 데이터를 학습해 개인정보를 생성할 가능성이 있다. 예를 들어, AI가 학습한 텍스트 데이터에서 특정인의 직업, 소득 수준, 정치적 성향 등을 추론해 제공할 경우, 해당 정보주체는 이러한 정보가 생성되었다는 사실조차 인지하지 못할 수 있다. 또한, 당사자는 자신에 대한 정보를 제공한 적이 없다면 정보주체에 관한 정보임에도 AI가 생성한 정보에 대해 소유권을 주장하기 어려울 수 있다(윤수영, 여정성, 2021). 더욱이 생성형 AI가 이미지와 같은 비정형 데이터를 처리하면서 내 개인정보를 사용하지 않았는데도 우연히 나와 매우 유사한 이미지가 생성되었다면 해당 이미지를 내 것이라고 주장할 수 있을지도 의문이다. 이렇듯 AI가 생성한 정보는 정보주체도 존재 자체를 인식하기 어려울 수 있고, 해당 개인정보가 누구의 소유인지 모호하고, 그 정보가 나에 관한 정보인지 판단이 애매할 수 있기 때문에 생성형 AI 시대에는 정보주체가 자신에 관한 개인정보라 하더라도 통제권을 행사하기 어려울 수 있다.

생성형 AI의 학습 특성과 정보주체 권리 행사 제약

현행 개인정보 보호법은 정보주체에게 개인정보의 열람(제35조), 정정·삭제(제36조), 처리 정지(제37조), 파기(제21조) 등의 권리를 보장하고 있다. 그러나 생성형 AI가 학습한 데이터에 대해서는 LLM의 기술적 특성상 정보주체의 권리를 실질적으로 보장하기 어렵다. AI가 특정 개인에 대한 정보를 학습해 생성한 경우, 이를 수정하거나 삭제하기 위해서는 전체 모델을 재학습해야 하고 많은 비용과 시간이 소요된다(장재영, 김종민, 2024). 따라서 개인정보 보호법에서 요구하고 있는 '즉시' 열람, 정정·삭제, 처리 정지, 파기 등이 용이하지 않다. 실제 장재영(2024)의 연구에 따르면, 현행 개인정보 보호법에서 LLM의 학습 특성은 주로 정보주체의 자기결정권 보장에 도전이 되고 있다고 한다.

이처럼 생성형 AI 시대에는 개인정보 자기통제권이 약화될 가능성이 크며, 기존 개인정보 보호 체계로는 생성형 AI의 학습 특성으로 인해 발생한 문제를 충분히 해결하기 어렵다. 개인정보 자기통제권이 약화되면

소비자 주체성 침해, 소비자와 기업 간 데이터 처리 관련 정보 격차 심화, 자동화된 의사결정(Automated decision-making)으로 인한 이용자 차별과 같은 문제가 발생할 수 있다. 따라서 AI가 생성한 정보주체의 통제권 약화와 정보주체 권리 행사 제약 문제를 해결하기 위해 새로운 법적·기술적 대응이 필요하다.

정보주체의 권리 보장 관련 국내외 현황

국제기구 및 주요 국가별 현황

세계 각국은 생성형 AI의 활성화에 따른 부정적 영향을 최소화하기 위해 다양한 정책적, 기술적 노력을 기울이고 있다. 2017년에 AI에 대한 최초의 국제적 윤리 지침 중 하나인 아실로마 AI 23 원칙(The Asilomar AI Principles)이 발표된 이후 2019년 경제협력개발기구(The Organisation for Economic Co-operation and Development, OECD)가 국제기구 수준에서는 최초로 'AI 권고안(Recommendation of the Council on Artificial Intelligence; the AI Principles)'을 채택했다(유화선 외, 2024.05.28.). OECD의 AI 권고안에는 포용적 성장, 지속적 개발 및 복지 기반 하에 AI 개발 및 활용 과정에서의 문제를 줄이기 위한 투명성, 설명 가능성, 인간 중심 가치 등의 접근법을 제시했다. 2024년 3월에는 생성 AI 모델의 잘못된 정보와 허위 정보 해결, 상호 운용 가능한 AI 거버넌스와 정책 환경 촉진을 포함한 개정안을 채택했다(한국인터넷진흥원, 2024.07.).

국가 단위에서는 유럽 연합이 'AI 법(AI Act)[7]'을 제정해 위험에 기반한 규제 접근법(A risk-based approach to regulation)을 도입했다. 이 법률에는 AI 공급자(provider)와 배포자(deployer)에게 위험관리 시스템 구축, 적정성 평가, 기본권 영향 평가, 투명성 의무 등을 부여하고 AI의 효율적 규제를 위한 AI 사무국 및 AI 이사회 조직 구성을 담고 있다. 유럽은 이미 GDPR을 통해 AI 시스템의 전체 생명 주기(lifecycle)에 대한 개인정보 보호 근거 규정을 가지고 있고, GDPR 제22조 프로파일링 등 자동화된 개별 의사결정(Article 22 GDPR. Automated individual decision-making, including profiling)을 통해 AI를 규율할 수 있다. 그러나 AI 법은 개인에 대한 위험 통제와 신뢰 구축에 중점을 두고 있으므로 이 조항만으로는 AI의 다양한 개인정보 처리와 정보주체의 개인정보 자기결정권을 보호하기에는 한계가 있다. 따라서 향후 AI 법과 GDPR 간의 개인의 권리를 보호를 위한 관계 설정이 중요한 이슈로 부각할 가능성이 있다(한국인터넷진흥원, 2023).

미국의 경우 트럼프 정부는 2019년 미국의 AI 분야 선도를 위한 행정명령(Executive Order 13859)[8]인 'Maintaining American Leadership in Artificial Intelligence'에 서명했고, 바이든 정부에서는 2023년 10월 혁신 및 경쟁 촉진, AI 안전 및 보안 강화, AI 관련 위험 평가 및 관리 강화 등을 포함한 8대 원칙을 담고 있는

7 Artificial Intelligence Act(Regulation (EU) 2024/1689)

8 Executive Order on Maintaining American Leadership in Artificial Intelligence

새로운 행정명령(Executive Order 14110)[9]인 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence'에 서명했다. 이 외에 미국은 Algorithmic Accountability Act of 2023 등의 입법을 추진하고 있고, 미국 국립표준기술연구소에서는 AI Risk Management Framework를 개발해 AI의 위협의 최소화를 위해 노력하고 있다. 전반적으로 미국은 AI의 부정적 영향을 최소화하면서도 AI 혁신을 우선하는 정책을 펼치고 있다.

중국 또한 2023년 8월에 '생성형 AI 서비스 잠정관리 방법(生成式人工智能服务管理暂行办法[10])' 등을 마련해 AI 기술의 발전을 촉진하면서 동시에 사회주의 핵심 가치, 국가안보, 개인정보 보호를 추진하고 있다. 이외에는 '인터넷 정보 서비스 알고리즘 추천 관리 규정(互联网信息服务算法推荐管理规定[11])', '인터넷 정보 서비스 딥페이크 관리 규정(互联网信息服务深度合成管理规定[12])' 등을 마련했다(한국인터넷진흥원, 2024.07.). 중국은 AI를 사회주의 핵심 가치 준수 및 안보 수단으로 인식해 적극적인 진흥 정책을 펼치고 있으며 이 과정에서 개인정보 자기통제권과 저작권 등은 일부 유보될 수 있다는 방향으로 정책을 추진하고 있는 것으로 보인다. 이들 국제기구와 주요 국가들의 AI 정책과 개인정보 통제권 제도의 특징을 요약하면 <표 2>와 같다.

표 2 - 세계 주요 기구 및 국가별 인공지능 규제와 개인정보 자기통제권 특징

구분	주요 내용	개인정보 자기통제권
OECD	- 비구속적 원칙 중심 - 투명성, 책임성, 인간 중심 - 회원국 간 정책 조화 및 국제 협력 기반	- 프라이버시 보호 및 데이터 자율성 - 정보주체의 권리 보장 강조 - 자기결정권 보호를 위한 기술적·제도적 설계 원칙(Privacy by Design) 반영
EU	- 법제화 중심 접근 - 위험 기반 규제 - 투명성 확보, 범용 AI 별도 규제 대상, GDPR 및 저작권법 준수	- GDPR과 연계해 정보주체 권리 실현 - 자동화된 결정에 대한 거부권, 설명 받을 권리, 데이터 이동권 등 도입
미국	- 시장 중심, 기업의 자율 규제 강조 - AI의 혁신과 경쟁력 확보 우선 - 안전하고 책임감 있는 AI 사용, 고위험 AI 기업 안전성 테스트 의무화 - 연방 차원 법률 부재, 행정명령 및 권고 중심	- 알고리즘 책임법 등으로 점진적인 개인보호 - 자기결정권 요소 포함 시도 및 프라이버시 강화 기술 도입 장려 - 법적 구속력 약하고, 보호 수준 제한적
중국	- 중앙 집중적 모델 - 사회주의 핵심 가치 준수, 데이터 안보 및 통제 우선, 사회 안정 강조 - 알고리즘 투명성 및 윤리성 규정화	- 자기결정권 보장 명시하고 있으나, 국가 개입 여지 커 실질적 보장 제한적 - AI 규제에 국가 감시 체계 병행 운영

출처: 저자 작성

9 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

10 https://world.moleg.go.kr/web/wli/lgslInfoReadPage.do?CTS_SEQ=2642&AST_SEQ=53&nationReadYn=Y&ETC=1&searchNtnl=CN

11 https://world.moleg.go.kr/web/wli/lgslInfoReadPage.do?CTS_SEQ=367&AST_SEQ=53&ETC=0

12 http://27.101.212.134/web/wli/lgslInfoReadPage.do?CTS_SEQ=22891&AST_SEQ=53

국내 현황

국내에서는 AI를 규율하기 위한 다양한 입법 활동이 진행됐다. 22대 국회에서만 10여 개의 법률이 제안되었고, 2025년 1월 21일에는 국내 최초의 법률인 '인공지능 발전과 신뢰 기반 조성 등에 관한 기본법'이 제정됐다. 이 기본법은 인공지능의 건전한 발전과 안전한 활용을 위한 국가 차원의 종합적 기준을 마련하여, 국민의 권익과 존엄성 보호, 삶의 질 향상, 국가경쟁력 강화를 도모하는 데 목적이 있다(과학기술정보통신부, 2024.12.27.).

국내의 개인정보를 규율하는 '개인정보 보호법'은 2023년 3월 14일 개정된 법률을 통해 생성형 AI를 포함한 자동화된 의사결정을 규율하기 위해 AI 등 완전 자동화된 결정에 대해 정보주체의 거부권과 설명권 보장 조항을 마련했다. 이 법은 유럽의 GDPR과 유사한 내용을 담고 있다. 그러나 입법 방식의 차이에 따라 GDPR은 보호조치와 개인정보 영향평가를, 개인정보 보호법은 처리 방식의 공개와 거부권 제공에 중점을 두고 있다. 또한 개인정보 보호법이 최근에 신설된 조항이므로 AI의 특성을 보다 구체적으로 다루고 있다는 특징이 있다(장재영, 2024). <표 3>에 GDPR과 개인정보보호법상의 자동화된 결정 규정을 비교했다.

표 3 - GDPR과 개인정보보호법상의 자동화된 결정 관련 규정 비교

항목	GDPR	보호법
규제 대상	- 프로파일링(개인의 사적인 측면의 평가)	- 완전히 자동화된 시스템으로 개인정보를 처리한 의사결정
정보주체 권리	- 자동화된 결정에 대한 참여 권리 - 자동화된 결정의 논리 및 결과 설명 요구권 - 자동화된 결정에 이의 제기 권리	- 자동화된 결정에 대한 설명 요구권 및 거부권 - 인적 개입 요구권 및 재처리 요청권
처리 방식의 공개	- 처리의 투명성 보장	- 자동화된 결정의 기준, 절차, 개인정보 처리 방식의 공개
개인정보 영향평가 의무	- 의무 조항 있음	- 명시적 의무 없음
하위 법령 위임	- 위임 조항 없음	- 자동화된 결정 관련 절차와 방법 등을 대통령령으로 위임

출처: 장재영(2024)

현재 국내·외에서는 AI의 발전에 따른 안전과 이용자 신뢰 및 정보주체의 자기결정권 보호를 위해 많은 노력을 기울이고 있다. 그러나 AI의 특성 및 현행 법률의 입법 취지, 규제 영역의 차이 등에 따라 이용자 등의 정보주체를 효율적으로 보호하는 데에는 한계가 있다. 또한 AI가 최근에 급격히 발전하고 있는 기술이므로 고려할 때 정보주체의 권리 보장의 실효성이 담보되기 위해서는 보다 많은 노력이 필요해 보인다. 따라서 다양한 측면에서 생성형 AI 환경에서 정보주체의 권리 보장 방법을 고민할 필요가 있다.

생성형 AI 환경에서 정보주체의 권리 보장 방법

정보주체의 개인정보 자기통제권 보장 강화

생성형 AI 시대에 개인정보 보호의 핵심은 정보주체의 자기정보통제권 보호이다. 따라서 AI가 생성한 개인정보의 출처, 생성 방식, 활용 목적을 정보주체가 명확히 확인할 필요가 있다. 현행 GDPR이나 개인정보 보호법에서는 자동화된 의사결정 조항으로 인공지능을 규율하고 있다. 그러나 해당 조항만으로 개인정보를 효과적으로 규율하기에는 한계가 있다(주민호, 2023). 따라서 공개된 개인정보의 성격, 공개의 대상 범위, 공개된 개인정보의 처리 방식, 정보주체의 예견 가능성을 기준으로 공개된 정보의 활용과 개인이 제공하지 않은 정보를 이용한 개인정보의 생성에 대한 규율 근거를 정비할 필요가 있다(개인정보위, 2024a; 장재영, 2024). 또한 생성형 AI의 생성 데이터 추적(Traceability) 및 설명 가능성(Explainability)을 확보하고 AI가 생성한 정보가 원본 데이터와 어떤 연관성이 있는지를 분석하는 AI 해석 가능성 강화 기술(Interpretable AI)을 적극 개발할 필요가 있다(조남용, 안동욱, 2023.09.06.; Schneider, 2024). 이를 활용해 개인과 관련해 어떠한 정보가 생성될 수 있는지를 검토하면 개인정보 전체에 대한 정보주체의 자기정보통제권이 강화될 수 있을 것이다.

학습 데이터에 개인정보 최소화

생성형 AI로부터 개인정보를 보호하기 위해서는 AI에 학습되는 데이터에 개인정보 이용을 최소화해야 한다. 이를 위해 생성형 AI의 학습에는 가명 정보(Pseudonymization)[13]나 익명 정보(Anonymization)[14]를 적극 활용할 필요가 있다. 해당 기술로는 차등 프라이버시(Differential Privacy)[15] 또는 합성 데이터(Synthetic Data)[16]와 같은 프라이버시 강화 기술(Privacy Enhancing Technology, PET)[17]이 있다(장재영, 김범수, 2024). 차등 프라이버시를 사용하는 대표적인 기관은 미국의 인구조사국(US Census Bureau)이다. 미국 인구조사국은 인구조사 데이터를 공개할 때 차등 프라이버시 기술을 적용해 개별 응답자의 신원을 보호하면서도 통계적 유용성을 유지하고 있다. 또한, 합성 데이터를 활용해 실제 개인정보 대신 유사한 패턴을 가지는 가상의 데이터를 만들어 생성형 AI 학습에 활용할 필요가 있다(개인정보위, 2024b). 현재

13 가명 정보: 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리한 것(개인정보 보호법 제2조제1항의2)

14 익명 정보: 시간·비용·기술 등을 합리적으로 고려했을 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보(개인정보 보호법 제58조의2)

15 차등 프라이버시: 데이터에 노이즈를 추가해 개별 데이터의 기여도를 숨기면서도 전체적인 데이터 패턴을 유지하는 기술

16 합성 데이터: 컴퓨터 시뮬레이션 또는 알고리즘을 이용해 원본 데이터의 구조적 및 통계적 속성을 재현한 데이터

17 PET: 개인 데이터에 대한 기밀성을 보호하면서 정보의 수집·처리·분석·공유를 가능하게 하는 기술의 집합으로 차등 프라이버시, 합성 데이터, 영지식 증명, 동형암호, 안전한 다자간 연산, 신뢰 실행 환경 등이 해당(출처: 이종혁, 2025.01.22.)

합성 데이터는 민감한 개인정보를 취급하는 금융권, 자율주행 업계 또는 데이터 부족에 시달리는 기업 현장에서 도입을 적극 추진하고 있다(트랜D, 2025.03.05.). 개인정보를 은폐할 수 있는 기술을 적극적으로 사용하면 불필요한 개인정보의 사용이나 생성을 줄일 수 있어 개인정보의 침해 예방에 효과가 있을 것이다.

정보주체 권리 보장 기술 적용

LLM은 한 번 학습된 정보를 쉽게 삭제할 수 없다는 점에서 기존의 개인정보 삭제 요구권 등을 효과적으로 행사하기 어렵다. 이를 해결하기 위한 방법으로는 필터링(Filter), 미세조정(Fine-Tuning), 언러닝(Unlearning) 기술이 있다(장재영, 2024). 필터링 기술은 AI 모델이 특정 개인정보를 학습하지 못하도록 사전에 차단하는 방식으로, 입력 데이터(프롬프트 필터링)와 출력 데이터(출력 필터링)에 적용할 수 있다(Ohm, 2024). 예를 들어, AI 모델이 주민등록번호나 신용카드 번호 등 민감한 정보를 생성하지 못하도록 특정 패턴을 탐지해 자동으로 차단할 수 있다. 미세조정은 AI 모델을 특정 규칙에 맞춰 다시 학습시키는 방식으로, 정보주체의 권리 보호를 위한 기술적 조치로 활용될 수 있다. 대표적인 방법으로는 파라미터 효율 미세조정(Parameter-Efficient Fine-Tuning, PEFT), 지도학습 기반 미세조정(Supervised Fine-Tuning, SFT), 직접 선호 최적화(Direct Preference Optimization), 사람 피드백 기반 강화 학습(Reinforcement Learning from Human Feedback, RLHF) 등이 있다(개인정보위, 2024a). 또한, 언러닝 기술을 적용해 특정 개인정보를 포함하는 학습 데이터를 제거하는 방법도 개발할 필요가 있다. 이 기술은 특정 정보가 학습된 이후에 정보주체가 개인정보 삭제를 요청할 경우 학습한 알고리즘에서 해당 정보만 삭제할 수 있는 기술이다(개인정보위, 2024a). 언러닝은 현재 연구가 진행되고 있는 기술이지만 궁극적으로 현행 개인정보 관련 법률 준수에 핵심인 기술이다. 따라서 향후 개인정보 보호에 중요한 역할을 할 것으로 기대된다.

결론

DeepSeek와 ChatGPT 사례에서 알 수 있듯이 생성형 AI의 발전은 다양한 분야에서 혁신적인 변화를 가져오고 있지만, 개인정보 보호에 대한 새로운 위협 또한 초래하고 있다. 본 글에서는 생성형 AI가 개인정보를 침해할 가능성과 그에 따른 위험성을 분석하고, 개인정보 자기결정권 보호 측면에서 해결 방안을 제시하고자 했다.

개인정보 보호는 단순히 기술적인 문제를 넘어, 개인의 권리와 자유를 보장하는 핵심적인 요소이다. 생성형 AI 환경에서 개인정보 보호를 위해서는 정보주체가 자신의 정보를 스스로 통제할 수 있는 권리 보장 체계 구축이 필수적이다. 이를 위해 우선 공개된 정보의 활용과 개인이 제공하지 않은 정보를 이용한 개인정보의 생성에 대한 규율 근거 마련이 필요하다. 또한, LLM 학습 과정에서 차등 프라이버시와 합성 데이터를

활용해 개인정보 사용 최소화 원칙이 준수되도록 해야 한다. 생성형 AI가 추론한 정보가 원본 정보와 어떤 연관성이 있는지를 분석하는 AI 해석 가능성 강화 기술도 제공해야 한다. 마지막으로 생성형 AI에 학습된 개인정보의 자기통제권 부여가 용이하도록 언러닝 기술을 적극 도입하고, 언러닝 기술의 도입이 활성화되기 이전까지는 필터링과 미세조정 기술을 적극 적용해 개인정보 보호법에서 요구하는 정보주체 권리 보장 항목이 준수되도록 노력해야 한다.

생성형 AI 시대에서 개인정보 자기통제권의 보호는 선택이 아닌 반드시 실현해야 할 과제이다. 개인정보 오남용을 알면서도 신뢰하고 자신의 정보를 제3자에게 넘길 정보주체는 제한적일 것이기 때문이다. 따라서 AI 개발자와 기업은 개인정보 보호를 우선적으로 고려하는 윤리적 책임을 다해야 하며, 개인 또한 자신의 정보가 어떻게 활용되는지에 대한 경각심을 가질 필요가 있다. 본 글이 개인정보 보호의 핵심 개념인 개인정보 자기통제권의 강화에 기여하고, 나아가 보다 안전하고 신뢰할 수 있는 디지털 환경을 구축하는 데 도움이 되기를 바란다.

참고문헌

- 과학기술정보통신부(2024.12.27.), 보도자료: 인공지능 시대의 새로운 서막, AI기본법 국회 본회의 통과
- 개인정보위(2023.07.27.), 보도자료: 오픈AI에 과태료 부과 및 개선권고 - 챗GPT 포함 국내외 주요 인공지능(AI) 서비스 대상 사전 실태점검 예정
- 개인정보위(2024a), 인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서
- 개인정보위(2024b), 데이터의 안전한 활용을 위한 합성데이터 생성·활용 안내서
- 김관영, 천곤궁, 김성훈(2025.03.), 딥시크(DeepSeek)의 등장과 인공지능(AI) 보안 이슈, KISA INSIGHT, Vol. 1, pp. 1-25
- 김도원, 김성훈, 이재광, 박정훈, 김병재, 정태인, 최은아(2023.05.), ChatGPT(챗GPT) 보안 위협과 시사점, KISA INSIGHT, Vol. 3, pp. 1-26
- 김지수, 홍민지(2025.03.10.), DeepSeek-R1 기술 분석, 한글과컴퓨터
- 김태성, 정호준(2025.02.06.), 韓 개인정보 中 서버에 저장 … 정부·기업 기밀 넘어갈수도, 매일경제, https://www.mk.co.kr/news/it/11234618
- 박민숙, 이효진(2022), 중국 개인정보 보호법의 주요 내용과 전망, 대외경제정책연구원
- 송혜리(2025.02.19.), 딥시크는 왜 이용자 정보를 틱톡 서버로 보냈을까, 뉴시스, https://www.joongang.co.kr/article/25318236
- 에스케이텔레콤(2025.03.26.), 개인정보 처리방침
- 윤수영, 여정성(2021), 데이터 경제에서 소비자주권의 확장 방향성과 소비자 데이터권리 연구, 소비자학연구, 32(5), 169-195
- 이종혁(2025.01.22.), 개인정보보호 강화기술(PET) 개발 동향, 주간기술동향, IITP
- 장재영(2024), 생성형 인공지능 모델의 개인정보 라이프 사이클에 따른 국내 개인정보 보호법 개선 고려 요소: GDPR과 개인정보 보호법의 비교·분석, 융합보안논문지, 24(3), 81-93

- 장재영, 김범수(2024), 프라이버시 보호 인공지능 개발 라이프 사이클 모델, 전자거래학회논문지, 29(4), 97-117
- 장재영, 김종민(2024), 인공지능의 학습 특성을 고려한 개인정보 라이프 사이클 모델, 융합보안논문지, 24(2), 47-53
- 조남용, 안동욱(2023.09.06.), 생성형 AI의 등장으로 더욱 중요해진 설명 가능한 AI (XAI), 인사이트 리포트, 삼성 SDS
- 주민호(2023), 자동화된 의사결정에 대한 기본권의 실효적 보장-EU AI 법 제54조와 GDPR 제22조의 관계를 중심으로, 법학논고, (82), 63-88
- 트랜D(2025.03.05.), 생성형 AI시대, 데이터 부족 해결사 '합성 데이터'가 뜬다, 중앙일보, https://www.joongang.co.kr/article/25318236
- 한국인터넷진흥원(2024.07.), 주요국(EU, 미국, 중국)의 AI 규제 현황과 시사점, 글로벌 개인정보 규제 심층분석
- 한국인터넷진흥원(2023), EU 인공지능법(안)과 GDPR의 상호작용 분석, 개인정보보호 월간동향분석 제1호, 1-10
- DeepSeek(2025.02.14.), DeepSeek Privacy Policy
- Johannes Schneider(2024), Explainable generative ai(genxai): A survey, conceptualization, and research agenda, Artificial Intelligence Review 57, (11), 1-38
- Morandín-Ahuerma, F.(2023), Twenty-three Asilomar principles for Artificial Intelligence and the Future of Life
- Paul Ohm(2024), Focusing on Fine-Tuning: Understanding the Four Pathways for Shaping Generative AI. Science and Technology Law Review, 25(2)
- Paul Kassianik and Amin Karbasi(2025.01.31.), Evaluating security risk in Deepseek and other frontier reasoning, CISCO
- Stephen Nellis, Krystal Hu, Jeffrey Dastin, Anna Tong and Katie Paul(2025.01.30.), Why blocking China's DeepSeek from using US AI may be difficult, Reuters, https://www.reuters.com/technology/artificial-intelligence/why-blocking-chinas-deepseek-using-us-ai-may-be-difficult-2025-01-29/