

개인정보위, 「생성형 인공지능(AI) 개발·활용을 위한 개인정보 처리 안내서」 공개

- 기업의 AI 개발과 개인정보 보호의 균형점 모색을 위한 개인정보 처리 안내서 발표 -

개인정보보호위원회가 2025년 8월 「생성형 인공지능(AI) 개발·활용을 위한 개인정보 처리 안내서」를 발표하였습니다. 생성형 AI 기술의 급속한 확산과 함께 대규모 데이터 처리로 인한 개인정보 처리 및 보호 이슈가 중요한 법적 쟁점으로 대두되는 가운데, 본 안내서는 AI 수명주기 전 단계에서 고려해야 할 개인정보 보호 기준과 안전조치를 체계화하여 제시하였습니다. 특히 모델개발자와 모델이용자의 역할을 구분하여 각 단계별 구체적인 가이드라인을 제공함으로써, 기업들이 AI 혁신과 개인정보 보호 사이의 균형점을 찾을 수 있는 실무적 지침을 마련하였다는 점에서 중요한 의의를 갖습니다.

1. 개요

2. 생성형 AI 개발·활용시 단계별 고려사항

3. 기업의 대응 전략

4. 시사점

1. 개요

생성형 AI 기술의 급속한 발전과 일상생활로의 확산은 새로운 기회와 함께 복잡한 법적 과제를 제시하고 있습니다. 특히 AI 모델의 사전학습 및 추가 학습 과정에서 필수적으로 수반되는 대규모 데이터 처리는 개인정보 보호 관점에서 다양한 이슈를 야기하고 있습니다.

개인정보보호위원회는 이러한 현장의 법적 불확실성을 해소하고 AI 신기술·신서비스 개발에 개인정보가 안전하게 활용될 수 있도록 지원하기 위해 본 안내서를 마련하였다고 발표했습니다. 이는 그간 위원회에서 축적한 분야별 AI 데이터 처리 안내서, 사전실태점검, 규제 샌드박스 및 사전적정성 검토제 등의 정책적 경험을 종합한 결과물입니다.

본 안내서는 생성형 AI를 개발하거나 혹은 이를 활용하여 개인정보를 처리하는 기업 및 기관을 대상으로 하며, 특히 LLM을 개발하고 제공하는 모델 개발자와 모델을 이용하여 AI 서비스를 개발 또는 제공하는 모델 이용자가 참고하여 생성형 AI의 수명주기 단계별로 개인정보 보호 사항을 검토할 수 있도록 구성되었습니다. 현재는 주로 언어모델 기반 생성형 AI 중심으로 작성되었으나, 향후 개인정보위는 안내서에서 다루는 내용을 음성·이미지·영상 등 영역까지 점진적으로 확대·보완해 나갈 예정입니다.

2. 생성형 AI 개발·활용시 단계별 고려사항

안내서는 생성형 AI의 개발·활용을 5단계로 구분하여 각 단계별 개인정보 보호 고려사항을 제시하고 있습니다.

가. 목적 설정 단계

생성형 AI 개발·활용의 첫 단계에서는 개인정보 처리 목적을 구체적이고 명확하며 합법적으로 설정해야 합니다. 범용성을 전제로 하는 생성형 AI의 경우 개인정보 처리 목적을 사전에 구체화하는 데 일정한 한계가 있을 수 있으나, 기업이나 기관은 개인정보 보호 원칙을 고려하여 처리 목적을 최대한 구체화할 것이 권장됩니다. 이를 통해 발생할 수 있는 리스크를 예측하고 미리 리스크를 완화할 수 있습니다.

특히 생성형 AI의 사전추가학습에 필요한 데이터에는 개인정보가 포함될 수 있으므로, 개인정보의 수집 출처별로 적법근거를 확보해야 합니다. LLM 등 기초 모델을 개발하는 경우 광범위한 지식을 확보하기 위하여 공개된 개인정보를 수집하는 경우와 기존 이용자 개인정보를 재사용하여 서비스 개선 또는 신규 개발하는 경우에 따라 각각 다른 적법근거 검토가 필요합니다. 만약 수집한 목적과 별개의 신규 AI 서비스를 개발하는 경우라면, 가명·익명처리하여 이용하거나 별도의 적법한 근거가 있어야 합니다.

나. 전략 수립 단계

전략 수립 단계에서는 개인정보 안심설계(PbD, Privacy by Design) 원칙을 반영해야 합니다. 개인정보 영향평가(PIA)는 PbD 원칙을 AI 제품·서비스에 반영하는 데 유용한 실천적 수단이 될 수 있습니다.

생성형 AI의 개발 방식은 크게 세가지, 즉 ①상용 AI 서비스를 활용하여 신속하게 개발을 하는 서비스형 LLM, ②공개된 사전학습 모델을 활용하여 개발하는 기성 LLM 활용방법, ③모델을 처음부터 직접 사전학습 시키는 자체개발로 구분할 수 있습니다. 각 방식별로 다른 개인정보 보호 고려사항이 적용됩니다. 서비스형 LLM의 경우 외부 LLM으로 전송되는 데이터의 처리 목적·범위, 보관·파기 정책 등을 확인해야 하며, 기성 LLM 활용 시에는 사전학습 모델의 상업적 이용 가능 여부, 개인정보 안전장치 내재 여부 등을 검토해야 합니다. LLM을 자체개발하는 경우라면 AI수명주기의 전 과정을 책임지게 되므로, 사전학습부터 배포 및 운영, 사후 관리에 이르는 전 단계에서 개인정보 리스크 요인을 경감시켜야 합니다.

다. AI 학습 및 개발 단계

이 단계에서는 데이터, 모델, 시스템 수준에서의 프라이버시 안전조치가 필요합니다. 먼저 데이터 수준에서는 데이터 오염(data poisoning) 방지를 위한 출처 검증과 전처리가 중요합니다. 특히 주민등록번호와 그 밖의 고유식별정보, 계좌번호, 신용카드번호 등은 AI 학습 전 삭제하거나 가명·익명화해야 하며, 명시적으로 스크래핑 거부 의사를 표시한 경우는 학습용도에서 제외되어야 합니다.

모델 수준에서는 미세조정(fine-tuning) 및 정렬(alignment) 등 기법을 활용한 추가적인 안전조치가 권장됩니다. SFT(Supervised Fine-tuning), RLHF(Reinforcement Learning from Human Feedback), DPO(Direct Preference Optimization) 등의 기법을 통해 모델의 안전성을 강화할 수 있습니다. 아울러 적대적 공격에 대한 대응을 고려하여, 모델 수준에서 안전장치를 보강하고 학습데이터에 포함된 개인정보를 추출하는

것을 막도록 해야 합니다.

시스템 수준에서는 접근제어 관리, 입출력 필터링 적용 등이 필요합니다. 입력 필터는 악의적이거나 민감정보 생성을 유도하는 프롬프트를 분석차단하고, 출력 필터는 AI 출력값에 포함된 개인정보를 감지하여 제거하거나 치환하는 역할을 합니다.

또한 개발 전 과정에서 학습·평가를 병행하여 안전성을 지속적으로 점검하고 보완할 수 있는 평가 체계도 포함되어야 합니다.

라. 시스템 적용 및 관리 단계

배포 전 테스트를 통한 프라이버시 리스크 점검·문서화가 필요합니다. 허용되는 이용방침(AUP, Acceptable Use Policy)을 작성·공개하여 생성형 AI의 사용 목적, 금지 행위 등을 명확히 해야 합니다.

정보주체의 열람, 정정·삭제 등의 요청에 대해서는 시간·비용·기술적 측면에서 합리적으로 실현 가능한 범위에서 권리 보장 방안을 마련해야 합니다. 학습데이터셋의 특성상 전통적인 권리행사 보장이 어려운 경우에는 그 사유를 정보주체에게 알기 쉽게 알리고, 대체 수단 등을 통해 최대한 성실하게 응하는 것이 권장됩니다.

마. AI 프라이버시 거버넌스 구축

개인정보 보호책임자(CPO) 중심의 내부 관리체계를 구축·운영해야 합니다. CPO는 생성형 AI의 목적 설정부터 배포·관리에 이르기까지 전 과정에서 개인정보 처리의 적법성과 안전성을 확보하기 위한 관리·감독 책임을 수행해야 합니다.

3. 기업의 대응 전략

가. 단계별 체크리스트 구축

기업들은 본 안내서에서 제시한 5단계 프로세스에 따라 자사의 생성형 AI 개발·활용 현황을 점검하고, 각 단계별 개인정보 보호 요구사항을 충족하는 체크리스트를 구축해야 합니다. 특히 모델개발자인지 모델이용자인지에 따라 다른 의무사항이 적용되므로, 자사의 역할을 명확히 정의하고 해당하는 요구사항을 파악하는 것이 중요합니다.

나. 개인정보 영향평가 실시

대규모 또는 민감한 개인정보 처리가 수반되거나 정보주체 권리에 중대한 영향을 줄 수 있는 AI 시스템의 경우 개인정보 영향평가를 실시하는 것이 권장됩니다. 민간 기업이 개인정보 영향평가를 자율적으로 수행하는 경우 과태료·과징금이 감경될 수 있어, 적극적인 활용을 검토해야 합니다.

다. 데이터 처리 기준 정립

공개된 개인정보의 처리에 있어서는 '정당한 이익' 조항의 적용 요건인 ①목적의 정당성, ②처리의 필요성, ③이익형량을 충족하기 위한 체계적인 기준을 마련해야 합니다. 특히 이익형량 단계에서 정보주체 권리 침해 가능성을 최소화하기 위한 기술적·관리적 안전조치와 정보주체 권리보장 방안을

구체적으로 수립해야 합니다.

라. 투명성 확보 방안 마련

AI 학습데이터에 본인의 정보가 포함되어 있는지, 개인정보가 AI에서 어떻게 처리되는지 등을 정보주체에게 명확하게 안내할 수 있는 체계를 구축해야 합니다. 데이터셋 수집 사실, 주요 출처, 처리 목적 등과 함께 AI 시스템 개인정보 처리 과정을 개인정보 처리방침, 기술문서, FAQ 등에 투명하게 공개하는 것이 권장됩니다.

마. 프라이버시 강화기술(PET) 도입 검토

차분 프라이버시, 연합학습, 동형암호화, 합성데이터 등 프라이버시 강화기술의 도입을 적극 검토해야 합니다. 특히 의료데이터와 같이 민감성이 높은 데이터의 경우 연합학습 등의 기법을 통해 개인정보를 직접 공유하지 않으면서도 AI 모델의 성능을 향상시킬 수 있습니다.

4. 시사점

생성형 AI가 산업 전반에 미치는 영향력이 확대되는 가운데, 개인정보 보호와 기술 혁신 사이의 균형점을 찾는 것이 기업들의 중요한 과제로 대두되고 있습니다. 본 안내서는 이러한 과제 해결을 위한 구체적이고 실무적인 지침을 제공한다는 점에서 큰 의의를 갖습니다. 특히 규제샌드박스과 사전적정성 검토제 등 혁신지원제도와의 연계를 통해 강화된 안전조치 하에서 개인정보 처리 근거를 확보할 수 있는 방안을 제시한 점은 주목할 만합니다. 이는 기업들이 혁신적인 AI 기술 개발과 개인정보 보호 요구사항 준수를 동시에 달성할 수 있는 실질적인 경로를 제공합니다.

앞으로 기업들은 본 안내서를 참고하여 생성형 AI 개발·활용 전략을 재점검하고, 개인정보 보호 체계를 강화해야 할 것입니다. 특히 AI 기본법 시행을 앞두고 있는 시점에서, 선제적인 컴플라이언스 체계 구축이 경쟁우위 확보의 핵심 요소가 될 것으로 예상됩니다.

향후 개인정보보호위원회가 예고한 음성·이미지·영상 등 영역으로의 확대와 함께, AI 기술 발전에 따른 지속적인 가이드라인 업데이트에 대한 면밀한 모니터링이 필요할 것입니다. 이를 통해 기업들은 급변하는 규제 환경에서도 안정적이고 지속가능한 AI 사업을 영위할 수 있을 것입니다.

화우 AI센터는 AI와 관련한 지식재산, 개인정보, 정보보안, 공정거래, 제조물책임, 입법컨설팅, 쟁송 등 모든 법적 영역에서 축적된 경험과 노하우를 기반으로 기업 고객을 위한 최적의 솔루션을 안내해드리고 있습니다. AI와 관련하여 문의사항이 있으신 경우 언제든지 화우에 연락하여 주시기 바랍니다.

Contacts

이광욱

파트너변호사

T. (+82) 2 6003 7535

E. kwlee@yoonyang.com

이근우

파트너변호사

T. (+82) 2 6003 7558

E. klee@yoonyang.com

정호선

변호사

T. (+82) 2 6182 8548

E. junghs@yoonyang.com

구소현

변호사

T. (+82) 2 6182 8745

E. shkoo@yoonyang.com

정종현

변호사

T. (+82) 2 6182 8716

E. jhjeong@yoonyang.com

경기룡

변호사

T. (+82) 2 6003 7082

E. kykyeong@yoonyang.com