개인정보보호위원회 보도자료

보도시점 2025. 8. 6.(수) 13:00 (2025. 8. 7.(목) 조간) 배포 2025. 8. 6.(수) 07:30

생성형 인공지능(AI) 개발·활용 위한 개인정보 처리 기준 제시한다

- 개인정보위, 「생성형 인공지능(AI) 개발·활용을 위한 개인정보 처리 안내서」 공개
- 실제 정책·집행 사례에 기반한 생성형 인공지능 생애주기 단계별 법적 고려 사항, 안전성 확보 기준 제시로 현장 불확실성 해소 기대

생성형 인공지능(AI) 서비스를 개발·활용하는 기업·기관을 위한 개인정보 안전 처리 기준이 마련되었다.

개인정보보호위원회(위원장 고학수, 이하 '개인정보위')는 8월 6일 서울 중구 소재 정동 1928 아트센터에서 「생성형 인공지능과 프라이버시」 오픈 세미나를 개최하고 **「생성형 인공지능(AI) 개발·활용을 위한 개인정보 처리 안내서」**(이하 '안내서')를 공개했다. 개인정보위는 안내서가 **생성형 인공지능 개발과 활용의 전 과정**에서 **개인정보 보호법** 적용의 **불확실성을 해소**하고, 기업·기관의 **자율적 법준수 역량을 높이는 데 큰 역할을 할 것으로 보고 있다.** 예를 들어 ChatGPT 등 **상용 대규모언어모델(LLM) 서비스**를 **활용**하거나 Llama 등의 **오픈소스 LLM**을 **미세조정**하여 **서비스를 개발·제공하는 사업자 등**이 본 안내서를 활용할 수 있다.

우리나라에 축적된 **의료·공공·금융분야 등의 데이터는 세계 최고 수준이다. 이들 데이터는 생성형 인공지능 발전**을 견인하는 **핵심 재료**로 활용된다. 그러나 **생성형 인공지능의 발전**은 한편 **프라이버시 침해 위험을 동반**할 **우려**가 있어 개인정보 처리에 관한 **명확한 기준 마련**이 **필수적**이다. **일선 현장**에서도 생성형 인공지능 활용과 관련한 **개인정보 보호법 적용의 불확실성**을 해소할 수 있는 체계적 안내가 필요하다는 요구가 지속되어 왔다.

※ (현장 목소리) "쉽게 이해하고 적용할 수 있는 안내서 시급", <산업계 개인정보 정책간담회, '25.3.>
"AI 사업 추진과정에서 개인정보 보호법 관련 불확실성으로 애로", <AI 스타트업 간담회, '25.3.>

이에 개인정보위는 올 초부터 **내부 검토**와 **외부 전문가 의견 수렴**을 거쳐 안내서 초안을 마련하고, **「AI 프라이버시 민·관 정책협의회*」**(7.8.) 및 개인정보보호위원회 제16회 전체회의(7.23.)를 거쳐 **공개본을 확정**했다.

* 학계, 법조계, 산업계, 시민단체 등 인공지능(AI) 분야 차세대 전문가 32명으로 구성

개인정보위가 이번 안내서에서 크게 중점을 둔 부분은 다음 세가지이다.

첫째, 생성형 인공지능을 **개발·활용**하는 **생애주기**를 **4단계로 분류**하고 단계별로 확인할 **최소한의 안전조치**를 **체계적**으로 **제시**했다. 또한, 인공지능 시스템이 실제로 **개발·활용**되는 **방식**과 **맥락***을 **유형화**하고, 각 유형에 따른 **법적 기준**과 **안전성 확보 기준**을 제시하였다.

* ▲서비스형 LLM 활용(예: ChatGPT API 연계) ▲기성 LLM 활용(예: LLama 오픈소스 모델 활용) ▲자체개발(예: 경량모델(SLM) 자체 개발) 등

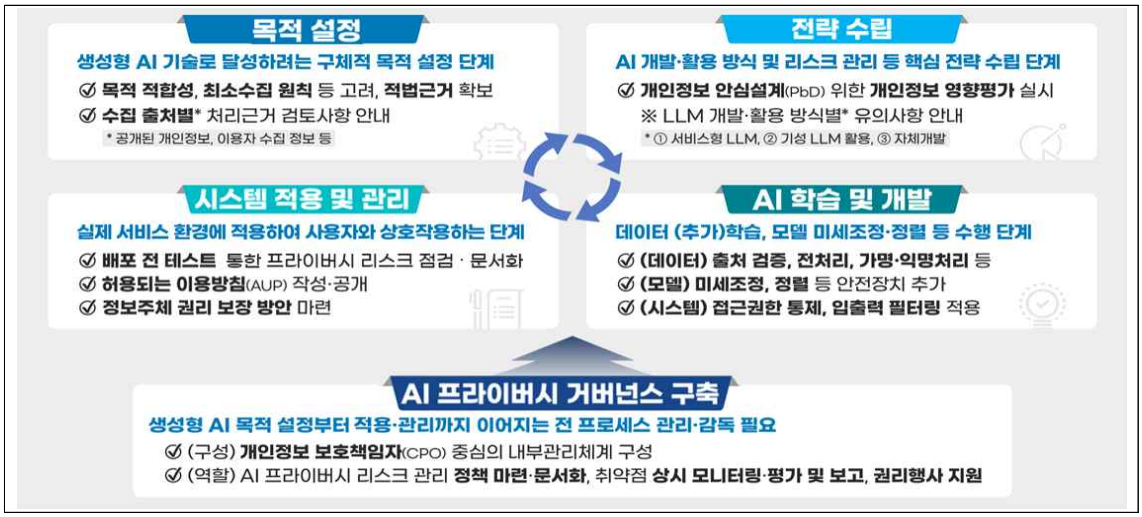
구체적으로 ①목적 설정 단계에서 인공지능 **개발 목적**을 명확히 하고, 개인정보 종류·출처별 인공지능 학습에 필요한 **적법 근거**를 다룬다. **②전략 수립** 단계는 개발 방식을 나누어 **유형별 리스크 경감 방안**을 안내한다. **③학습 및 개발** 단계는 **데이터오염, 탈옥** 등 리스크를 고려한 **다층적 안전조치**를 제시하고 **인공지능 에이전트**의 관리 방안도 포함했다. **④적용 및 관리** 단계는 **정보주체 권리 보장** 등을 중점적으로 다룬다. **⑤마지막으로 전체 과정**에서의 **개인정보 보호 관점**을 **내재화**하기 위해 필수적인 **개인정보보호 책임자(CPO) 중심의 거버넌스 구축 방안을 제시**한다. 이러한 기반에서 기업·기관은 위 과정을 **반복**하며 **시스템**을 **고도화·개발**하는 것이 권장된다.

※ 안내서 요약본 : 참고 자료 참조

둘째, 이용자 개인정보를 인공지능에 학습할 수 있는 법적 기준 등 생성형 인공지능 개발·활용 과정에서 **불확실성이 높은 이슈**들에 대해 개인정보위의 **정책** 및 **집행 사례**를 바탕으로 **구체적인 해결방안**을 제시했다. 개인정보위는 그간 생성형 인공지능에 대한 **▲안내자료 마련 ▲사전실태점검** 등 집행 사례 **▲규제샌드박스** 및 **사전적정성 검토** 등을 통해 다양한 **정책적 경험을 축적**해왔는데, 이들 경험에서 얻은 **구체적 사례**에 기반한 **법 해석 기준**과 **안전조치 기준**을 **안내서에 반영**하여 유용성을 높였다.

셋째, 인공지능 에이전트, 지식증류, 머신 언러닝 등 생성형 AI 개발·활용과 관련한 최신 기술 동향과 연구 성과 등을 반영하였다. 안내서는 향후 급속한 **기술 발전**과 **국내·외** 개인정보 보호 **정책** 변화에 발맞춰 **지속적**으로 **업데이트**될 예정이다.

<생성형 인공지능 개발·활용 단계별 고려 사항>



고학수 개인정보위 위원장은 "명확한 안내서를 통해 **실무 현장의 법적 불확실성이 해소**되고, **생성형 인공지능 개발·활용**에 **개인정보 보호 관점이 체계적으로 반영**될 수 있을 것으로 기대된다."라며, "앞으로도 개인정보위는 **'프라이버시'와 '혁신' 두 가치가 상호 공존**할 수 있도록 **정책 기반을 마련**해 나가겠다."고 말했다.

담당부서	개인정보정책국 인공지능프라이버시팀	책임자 담당자	팀 장 구민주 (02-2100-3073) 사무관 전이루 (02-2100-3078)

참고 「생성형 인공지능(AI) 개인정보 처리 안내서」 요약

❶ 목적 설정

먼저, 생성형 인공지능을 통해 달성하고자 하는 **목적을 구체화**하고, 이를 토대로 **'어떤 유형의' 개인정보**를 **'어떤 목적'**으로 **처리**할지 구체적으로 설정해야 한다. **목적 설정**은 프라이버시 **리스크 관리를 위한 출발점**이자, 개인정보 처리의 **적법성을 판단하는 핵심 기준**이 된다.

안내서는 생성형 인공지능 학습에 이용되는 학습데이터를 크게 **공개 데이터**와 **이용자 데이터**로 구분하여 각각에 대한 적법기준을 제시했다. **우선,** 인터넷 등에 **공개된 데이터**는 개인정보 보호법상 **'정당한 이익'** 조항(제15조제1항제6호)에 따라 적법하게 활용될 수 있다. 만약 기업·기관이 **이미 보유**하고 있는 **이용자 정보**를 인공지능 학습을 위해 **재이용**하고자 하는 경우에는 **당초 수집 목적과의 관계**에서 **▲목적 내 이용 ▲추가적 이용 ▲별개의 목적**에 해당할 수 있어, 각각에 대한 **판단 기준**과 **구체적 사례**를 담았다.

목적 내 이용	• 동의, 계약, 정당한 이익 등 적법근거에 기초해 수집한 개인정보는 수집 목적 범위 내에서 AI 서비스 개선 및 고도화를 위해 이용 가능 • (예) 출입국 관리 목적으로 수집한 안면정보를 AI기반 시스템 고도화에 사용
추가적 이용	• 수집 목적과 합리적 관련성 있는 AI 개발에 활용 가능 • (예) 적절한 안전조치 및 정보주체 선택권 보장을 전제로 이용자 프롬프트의 AI 학습 허용
별개의 목적	• 당초에 수집한 목적과는 명확히 구분되는 별개 목적 위한 신규 AI 서비스 개발로, 가명·익명 처리 또는 개인정보 보호법상 새로운 적법근거 마련 필요 • 강화된 안전조치 등 전제로 규제샌드박스 통한 개인정보 처리 근거 확보 가능 • (예) 규제샌드박스 통해 보이스피싱 통화데이터를 '성문분석' 위한 AI 학습에 허용

❷ 전략 수립

다음으로 **목적 달성**을 위한 **구체적 전략**을 **수립**하는 단계이다. 이 단계에서는 **생성형 AI 개발 방식, 미세조정 등 추가학습 여부** 등을 고려하여 개인정보와 관련한 **리스크 요인을 파악**하고 **전략을 수립**하게 된다. 목적 설정 및 전략 수립 단계를 거치면 실제 서비스 구현에 앞서 주요 분야의 기획·설계가 완료된다.

안내서는 생성형 AI 개발·활용 방식을 **▲서비스형 LLM**(예: ChatGPT API 연계) **▲기성 LLM 활용**(예: Llama 오픈소스 모델 추가학습) **▲자체개발**(예: 경량모델(SLM) 개발)의 **세가지 유형**으로 구분하고, 각 **유형별 고려사항**과 **대응전략**을 안내했다. 먼저, **서비스형 LLM**을 **API 연계**하여 활용하는 경우에는 이용자의 개인정보가 안전하게 처리될 수 있도록 **라이선스 계약, 이용약관** 등을 통해 **데이터 처리 목적과 범위 등에 대한 통제장치를 마련**하도록 안내했다. 개인정보위는 **의료기관**이 **서비스형 LLM**을 사용해 **의료기록 작성업무**를 **자동화**한 사안에서, 보다 높은 수준의 프라이버시 보호를 제공하는 **'기업용 라이선스(Enterprise API)*'**를 사용하도록 한 사례가 있다.

* 개인정보 저장 및 재이용(AI 학습 포함) 배제를 기본값으로 함

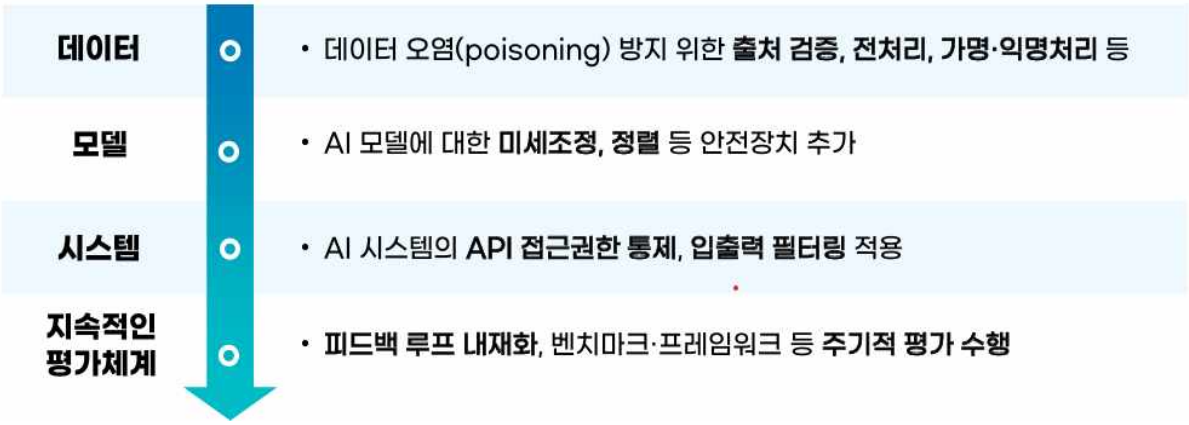
또한, **일선 현장**에서는 **공개 모델**에 전문 지식 등을 **추가학습**하여 **활용**하는 수요도 많다. 그런데 이러한 기성 모델의 **리스크**가 **사후에 발견**되기도 한다. 이때, **모델개발자**는 해당 사항을 **신속하게 공지**하고 **개선모델을 재배포**하는 등 조치를 취해야 하고, **모델이용자**는 **공지를 상시 확인**하며 **모델의 최신 버전**과 **패치**를 **주기적**으로 **적용**하는 등 시스템 **안전성과 신뢰성 확보**를 위해 **노력**할 것을 권장했다.

❸ AI 학습 및 개발

이 단계에서는 사전에 설정한 목적과 전략에 따라 **학습 및 개발을 수행**하게 된다. 생성형 인공지능은 **학습데이터를 그대로 발화**하는 **'암기' 위험**을 비롯해 **데이터 오염**(data poisoning), **탈옥**(jailbreak) 등 **다양한 리스크 요인**이 있어, 이를 **경감**하기 위한 **다층적인 안전조치** 마련이 필수적이다.

안내서는 학계, 산업계 등에서 연구되고 있는 **최신 리스크 경감방안**을 종합하여 **▲데이터 수준 ▲모델 수준 ▲시스템 수준**에서 고려할 수 있는 **기술적·관리적 안전조치 기준을 안내**했다. 이와 더불어, 생성형 인공지능 학습 및 개발 시 **지속적인 평가**를 통해 **프라이버시 안전성을 점검·보완**할 수 있는 **피드백 루프**(feedback loop)를 내재화할 것을 권장했다.

<학습·개발 시 리스크 경감 방안 주요 내용>



최근에는 자율적으로 외부 지식을 **검색**하거나 **기억** 능력(메모리)을 갖춘 **검색형·기억형** AI 에이전트가 등장하고 있다. **검색형 에이전트**는 **불법적**이거나 **유해한** 정보를 **활용하지 않도록 통제**할 필요가 있고, **기억형** 에이전트는 **이용자 프로파일링**과 같은 위험을 낮추기 위해 **사전적 통제 장치**를 적용할 것을 권장했다.

❹ 시스템 적용 및 관리

마지막으로 학습·개발을 완료한 생성형 인공지능 시스템을 **실제 활용 환경에 배포·적용**하여 **최종적으로 서비스 제공을 시작**하는 단계이다. 이 단계는 **배포 이전**과 **배포 이후**로 나누어 **리스크 관리 전략**을 세워야 한다.

배포 이전 단계에서는 **AI 결과값의 정확도, 안전조치 우회 시도에 대한 저항성, 학습데이터의 노출 가능성** 등을 **집중 점검**하고 **결과를 문서화**하는 것이 바람직하다. 한편, **배포 이후**에는 **적절한 모니터링**과 함께 **정보주체 권리 보장 방안**이 마련되어야 한다. 현 기술 수준에서 정보주체의 열람, 정정·삭제 등 **전통적인 권리보장**이 제약될 수 있는 **한계**가 있으나, 그 경우에도 **해당 사유**를 정보주체에게 **명확하게 알리고, 출력 필터링 등 대체 수단**을 통해 정보주체 요구에 충실히 응할 것이 권장된다.

❺ 거버넌스 구축

이상에서 살펴본 프라이버시 리스크 관리가 실효적으로 이루어지기 위해서는 기관·기업 내 **거버넌스 체계 마련이 중요**하다. 특히 **개인정보 보호책임자(CPO)를 중심**으로 **내부 관리체계를 구축·운영**할 필요가 있다. **개인정보 영향평가, 레드티밍** 등을 활용해 프라이버시 **리스크를 지속적으로 평가**하는 것이 바람직하다.

아울러, CPO가 조직 내 **AI 기획·개발 초기 단계**부터 **적극적으로 개입**하여 개인정보 처리에 관한 충분한 정보를 확보하고, 관련 부서에 **적시 피드백을 제공**하여 개인정보 **보호 원칙이 생성형 AI 서비스에 내재화**될 수 있는 체계를 마련할 것을 권장했다.