개인정보보호위원회 보도자료 다시 대한민국! 새로운 국민의 나라

보도시점 2024. 7. 17.(수) 14:00 배포 2024. 7. 17.(수) 07:30

인공지능(AI) 개발·서비스에 이용되는 '공개 데이터' 처리 기준 제시

- 개인정보위, 「인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」 공개
- 기업의 불확실성 해소 및 국민의 프라이버시 보호 효과 기대

생성형 인공지능(AI) 모델 개발 시 활용되는 **'인터넷상 공개 데이터'**를 안전하게 처리할 수 있는 **정부 차원의 기준**이 나왔다.

개인정보보호위원회(위원장 고학수, 이하 '개인정보위')는 **인공지능(AI) 개발에 필수적인 공개 데이터**가 현행 개인정보 규율체계 내에서 **적법하고 안전하게 처리**될 수 있도록 **「인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」**(이하 '안내서')를 마련해 **공개**했다.

공개 데이터는 인터넷상 누구나 합법적으로 접근할 수 있는 데이터로서, **챗지피티(ChatGPT) 등 생성형 인공지능(AI)을 개발**하기 위한 **학습데이터의 핵심원료**로 쓰인다. 인공지능(AI) 기업들은 커먼크롤*(common crawl), 위키백과(wikipedia), 블로그, 웹사이트 등에 있는 공개 데이터를 **웹 스크래핑**** **등의 방식으로 수집**해 **인공지능(AI) 학습데이터로 활용**하고 있다.

* 인터넷상의 데이터를 자동으로 수집하여 누구나 접근하고 분석할 수 있도록 저장·유지·관리하는 공개 저장소(http://commoncrawl.org)

** 웹사이트에서 필요한 데이터를 자동으로 추출하는 기법

이러한 **공개 데이터에는** 주소, 고유식별번호, 신용카드번호 등 **여러 개인정보**('공개된 개인정보')가 **포함**될 수 있어, 국민의 프라이버시가 침해될 우려가 크다. 하지만 현행 **개인정보 보호법**(이하 '보호법')에는 이러한 공개된 개인정보 처리에 적용될 수 있는 **명확한 기준이** 없다.

예를 들어, 인공지능(AI) 학습에 공개 데이터가 **대규모로 처리되는 상황**에서 현행 보호법 상의 **정보주체 개별 동의**나 **계약 체결** 등의 조항을 적용하는 것은 사실상 어렵다. 또한 인공지능(AI) **학습**이 **전통적인 개인정보 처리방식**과 다르다 보니, **보호법상의 안전성 확보조치** 등의 규정도 **그대로 적용**하는 것에 **일정한 한계**가 존재했다.

이에 개인정보위는 **공개된 개인정보 수집·활용의 법적기준을 명확화**하고 **인공지능(AI) 개발 및 서비스 단계**에서 **어떤 안전조치를 취하는 것이 적정**한지에 대해 기업이 참고할 수 있는 최소한의 기준을 제시하는 **안내서**를 **마련**하게 되었다고 밝혔다. 공개된 개인정보를 활용하는 기업들의 **개인정보 침해 이슈를 최소화**하는 동시에 **법적 불확실성을 해소**해 **기업의 혁신성장**을 **돕겠다**는 취지이다.

개인정보위는 지난해 8월 「인공지능(AI) 시대 안전한 개인정보 활용 정책 방향」을 발표한 이후 **「인공지능(AI) 프라이버시 민·관 정책협의회*」**를 중심으로 안내서에 관한 논의를 진행하는 한편, **학계·산업계·시민단체**와도 **소통하며 광범위한 의견수렴**을 병행하였다.

* 학계, 법조계, 산업계, 시민단체 등 인공지능(AI) 분야 차세대 전문가 30명으로 구성, 3개 분과(데이터 처리기준, 리스크 평가, 투명성 확보) 운영

특히, **유럽연합(EU), 미국** 등 인공지능(AI) 혁신과 안전의 균형을 꾀하는 해외 주요국에서도 최근 **공개 데이터 처리를 포함**한 **인공지능(AI)·데이터 처리 전반**에 대하여 **개인정보 보호 규율체계를 형성**해나가고 있는 점을 고려해, **국제적으로 상호운용성 있는 기준**을 마련하는 데 중점을 두었다.

< 참고: 해외 정책 동향 >

- **(영국)** 웹 스크래핑을 통해 수집한 데이터를 생성형 인공지능(AI) 학습에 사용하는 것에 '정당한 이익'이 인정될 수 있음을 밝히고 의견수렴 중('24.1.~)
- **(프랑스)** 인공지능(AI) 학습 목적으로 개인정보 처리시 '정당한 이익'이 인정되기 위한 기준 제시('23.10.~)
- **(미국)** 공개된 정보를 개인정보 범위에서 제외하는 연방 개인정보보호법 제정안(APRA) 발의('24.4.)

먼저, **보호법 제15조**에 따른 **'정당한 이익'** 조항*에 의해 **공개된 개인정보를 인공지능(AI) 학습·서비스 개발에 활용**할 수 있다는 점을 **분명**히 했다.

* (보호법 제15조제1항제6호) 개인정보처리자의 정당한 이익을 달성하기 위하여 필요한 경우로서 명백하게 정보주체의 권리보다 우선하는 경우. 이 경우 개인정보처리자의 정당한 이익과 상당한 관련이 있고 합리적인 범위를 초과하지 아니하는 경우에 한한다.

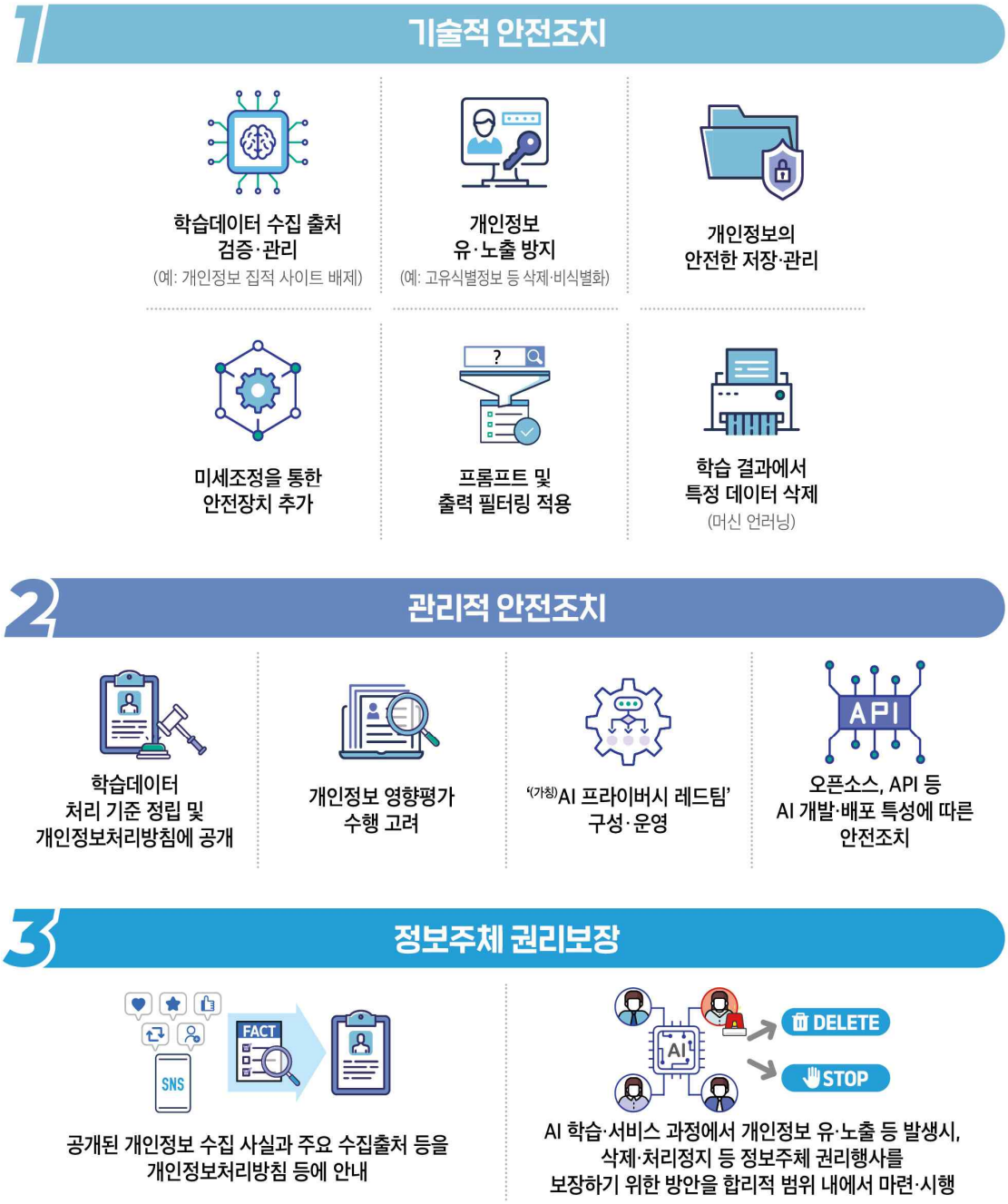
또한 이러한 **'정당한 이익' 조항이 적용**되기 위해서는, **인공지능(AI) 개발 목적의 정당성, 공개된 개인정보 처리의 필요성, 구체적 이익형량**이라는 **세 가지 요건을 충족**하여야 한다. 개인정보위는 이번 안내서를 통해 **세 가지 요건의 내용**과 **적용사례**도 안내했다.

요건	주요내용
목적의 정당성	◆ 개인정보처리자의 정당한 이익의 존재 • 공개된 개인정보 처리를 통해 개발하려는 AI의 목적·용도를 구체화하여 정당한 이익을 명확화 ※ (예) 의료진단보조, 신용평가, 텍스트 생성·분류·번역 등을 수행하는 LLM 등
처리의 필요성	◆ 공개된 개인정보 수집·이용의 필요성과 상당성·합리성이 인정될 것 ※ (예) 의료진단보조 AI 개발시 개인의 소득·재산 등 관련없는 정보는 학습 배제
구체적 이익형량	◆ 개인정보처리자의 정당한 이익이 정보주체 권리에 명백히 우선 • '명백성' 요건 충족을 위해 (i) 정보주체 권익침해 방지를 위한 안전성 확보조치 및 (ii) 정보주체 권리보장 방안 마련·시행을 통해 개인정보처리자 이익이 우선하도록 조치 ※ 상세내용 후술

따라서, '정당한 이익' 조항의 합리적 해석기준을 마련하는 것은 **유럽연합 일반 개인정보보호법**(EU GDPR*)이나 최근 **인공지능(AI) 안전성 규범 논의** 등 **글로벌 스탠다드와의 상호 운용성을 높이게 되는** 측면이 있다.

* 영국, 프랑스 등 유럽연합(EU) 주요국도 '정당한 이익'이 공개된 개인정보의 처리 근거가 될 수 있다는 입장

또한 개인정보위는 안내서를 통해 인공지능(AI) 기업이 '정당한 이익'을 근거로 공개된 개인정보를 처리하기 위해 고려할 수 있는 **기술적·관리적 안전성 확보조치**와 **정보주체 권리보장 방안**을 구체적으로 안내했다.



다만, **빠른 인공지능(AI) 기술변화 등을 고려**하여 세부적 안전조치 등을 유연하게 **도입·시행**할 수 있도록 했다. 인공지능(AI) 기업은 모든 안전조치를 **의무적으로 시행해야 하는 것은 아니며**, 안내서에 제시된 **여러 안전조치의 순기능**과 **인공지능(AI) 성능저하, 편향성 등 부작용**과, **기술 성숙도**를 고려하여 기업의 **특성에 맞는 「안전조치의 최적 조합」**을 스스로 **선택**하여 **이행**할 수 있다.

개인정보위는 ’24.3. 인공지능(AI) 사전실태점검※을 통해 파악한 **주요 대규모 언어모델(LLM) 사업자**의 **실제 안전조치 이행사례**를 안내해, 기업이 **「최적 조합」을 판단하는 데 참고**할 수 있도록 하였다.

※ 개인정보위는 대규모 언어모델(LLM) 사업자의 **공개된 개인정보 처리 과정**에서 드러난 **중요한 문제점**에 대해 **안전조치를 강화**할 것을 **개선권고**한 바 있음('24.3월, 개인정보위 의결)

- 개인정보위가 **주요 개인식별정보가 노출된 도메인 정보(URL)***를 **주기적으로 탐지**하여 인공지능(**AI) 기업에 제공**하면, 기업이 해당 도메인 정보(URL)를 **학습데이터 수집에서 배제**할 것을 권고함

* 개인정보위와 한국인터넷진흥원(KISA)은 공공·민간 홈페이지를 대상으로 주민등록번호, 여권번호, 운전면허번호 등 보호법 상 노출이 금지된 개인정보의 노출 및 불법적 유통을 탐지하여 삭제하고 있음('23년 기준 총 20,999개 페이지 탐지)

마지막으로, 인공지능(AI) 개발을 위한 학습데이터 처리와 관련한 인공지능(AI) **기업과 개인정보보호책임자(CPO)의 역할**을 강조하였다. 개인정보보호책임자(CPO)를 구심점으로 하는 **‘(가칭)인공지능(AI) 프라이버시 담당조직’**을 자율적으로 구성·운영하고 **안내서에 따른 기준 충족 여부를 평가**하여 **그 근거를 작성·보관**하도록 권고하였다. 인공지능(AI) 성능 개선 등 중대한 기술적 변경이나 개인정보 침해 발생 우려 등 **위험 요인**을 **주기적으로 모니터링**하고, 개인정보 유·노출 등 침해사고 발생 시 **신속한 권리구제 방안**도 마련하도록 했다.

안내서는 추후 개인정보 관련 법령 제·개정, 인공지능(AI) 기술발전 추이, 해외 규제정비 동향 등을 고려해 **지속 업데이트**될 예정이다.

아울러, 공개된 개인정보와 함께 인공지능(AI) **학습데이터의 주요 원천**을 이루는 **이용자 개인정보**의 적법한 **처리 근거**와 **기준** 등에 대해서는 학계, 산업계, 시민단체 등의 의견수렴을 거쳐 구체화해 나갈 예정이다.

이 밖에 개인정보위는 사전적정성 검토제, 규제샌드박스, 개인정보 안심구역 등 혁신지원제도를 통해 **인공지능(AI) 기업과 수시로 소통**하면서 **기술발전과 시장상황을 모니터링**하고, 이를 통해 **축적된 사례**와 **경험**을 토대로 보호법**을 인공지능(AI) 시대에 맞게 정비**하는 작업도 추진할 예정이다.

이번 안내서 관련 논의에 참여한 **김병필 카이스트 교수**(민·관 정책협의회 데이터 처리기준 분과장)는 “본 안내서는 **개인정보를 충실히 보호**하면서도

인공지능(AI) **혁신을 장려**하는 **적정한 절충점**을 찾고자 하는 노력의 일환이다. **신뢰할 수 있는 인공지능(AI) 개발, 이용**을 위한 **좋은 참고 자료**가 될 것이라 생각한다"면서, "다만, 인공지능(AI) **기술이 빠르게 변화**하는 만큼 향후 본 안내서에 포함된 내용도 **지속적으로 발전**해 나갈 것이라는 점을 염두에 두었으면 한다"고 강조했다.

민·관 정책협의회의 공동의장인 **배경훈 엘지 에이아이 연구원장**은 "이번 안내서 공개는 **인공지능(AI) 기술 발전**과 **개인 데이터 보호**를 **동시에 달성**하기 위한 중요한 진전이자 첫걸음"이라며, "**공개 데이터**에서 **개인정보를 안전하게 처리**할 수 있도록 **기준**을 **제공함으로써 인공지능(AI) 기술 개발**에 있어 **법적 불확실성이 낮아져 안전하게 데이터를 활용**할 수 있게 되었고, 이는 곧 **국민들이 신뢰**할 수 있는 **데이터 처리 환경**에서 **인공지능(AI) 기술의 혜택**을 누릴 수 있는 기반이 될 것"이라고 전망했다.

고학수 개인정보위 위원장은 "인공지능(AI) 기술 진보가 빠르게 이루어지고 있지만 **인공지능(AI) 개발의 핵심 관건**인 **공개 데이터 학습**이 보호법에 비추어 **적법**하고 **안전**한지 여부는 **공백**인 상황이었다"면서 "이번 안내서를 통해 **국민이 신뢰**하는 **인공지능(AI)·데이터 처리 관행**을 **기업 스스로 만들어 나가고** 이렇게 축적된 **모범사례**가 **안내서에 지속적으로 반영**될 수 있기를 기대한다"고 말했다.

담당부서	개인정보정책국 인공지능프라이버시팀	책임자	(代)팀 장	태현수 (02-2100-3071)
		담당자	서기관	구민주 (02-2100-3073)