생성형 인공지능(AI) 개발·활용을 위한

개인정보 처리 안내서

2025. 8. 6.

개인정보보호위원회

목차

발간 배경

AI 기술 발전 급속화

기업 · 기관의 개발 · 활용 활발

현장의 어려움 해소 필요 + **위원회에 축적된 경험 반영**

- **정책 수립**
  - AI 시대 안전한 개인정보 활용 정책방향('23. 8.)
- **지침 · 안내서**
  - 비정형 데이터 가명처리 기준('24. 2.)
  - 공개된 개인정보 처리 안내서('24. 7.)
  - AI 프라이버시 리스크 관리 모델('24. 12.)
- **AI 분야 혁신지원제도**
  - 자율주행AI학습('24. 2.)
  - 첨단바이오 국제공동연구('24. 6.)
  - 보이스피싱 예방 AI 학습('24. 10.)
- **사전실태점검**
  - 국내외 주요 6개 LLM 사업자 대상('24.3.)
  - AI 응용 서비스 대상('24.6.)

생성형 AI의 수명주기 각 단계별
개인정보 보호 원칙 준수를 위해 고려할 적법성, 안전성 확보 조치 등을 소개

적용 범위 및 성격

적용 범위

생성형 AI 개발 · 활용 시 개인정보를 처리하는 기업 · 기관 등

- (모델 개발자) 대형언어모델(LLM)을 개발하고 제공·판매하는 기업·기관 등
- (모델 이용자) ①외부 상용 모델을 API 연계하거나,
  ②공개 모델을 다운로드하고 개발해 AI 서비스 제공하는 기업·기관 등

성격

생성형 AI 개발 · 활용 단계별 개인정보 보호 준수 위한 고려사항 안내

- 다른 법령상 의무에 대해서는 다루지 않음
- 향후 AI 관련 법 · 제도 · 기술이 발전함에 따라 주기적으로 수정 · 보완 예정

Ⅱ. 생성형 AI 개발 · 활용 단계

목적 설정

생성형 AI 기술로 달성하려는 구체적 목적 설정 단계

- ✓ **목적 적합성**, **최소수집 원칙** 등 고려, **적법근거** 확보
- ✓ **수집 출처별*** 처리근거 검토사항 안내

* 공개된 개인정보, 이용자 수집 정보 등

전략 수립

AI 개발·활용 방식 및 리스크 관리 등 핵심 전략 수립 단계

- ✓ **개인정보 안심설계**(PbD) 위한 **개인정보 영향평가** 실시
- ✓ LLM 개발·활용 방식별* 유의사항 안내

* ① 서비스형 LLM, ② 기성 LLM 활용, ③ 자체개발

시스템 적용 및 관리

실제 서비스 환경에 적용하여 사용자와 상호작용하는 단계

- ✓ **배포 전 테스트** 통한 프라이버시 리스크 점검 · 문서화
- ✓ **허용되는 이용방침**(AUP) 작성·공개
- ✓ **정보주체 권리 보장 방안** 마련

AI 학습 및 개발

데이터 (추가)학습, 모델 미세조정·정렬 등 수행 단계

- ✓ **(데이터) 출처 검증, 전처리, 가명·익명처리** 등
- ✓ **(모델) 미세조정, 정렬** 등 안전장치 추가
- ✓ **(시스템) 접근권한 통제, 입출력 필터링** 적용

AI 프라이버시 거버넌스 구축

생성형 AI 목적 설정부터 적용·관리까지 이어지는 전 프로세스 관리·감독 필요

- ✓ (구성) **개인정보 보호책임자**(CPO) 중심의 내부관리체계 구성
- ✓ (역할) AI 프라이버시 리스크 관리 **정책 마련·문서화**, 취약점 **상시 모니터링·평가 및 보고**, **권리행사 지원**

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

① 목적 설정 : 생성형 AI 기술을 통해 달성하고자 하는 목적 설정

‘어떤 유형’의 개인정보를 ‘어떤 목적’으로 처리할지 개인정보 처리 목적 구체화

AI 디지털 교과서(AIDT) 사례

- AIDT 통합포털은 **학생별 학습이력 데이터를 수집 · 저장** (학습시간, 성취수준, 진도율, 접속시간, 커뮤니티 참여도 등)
- 학생의 학습이력 데이터의 오남용 우려가 있는 상황에서 학습데이터로 활용하는 **처리 목적이 불분명**

➡ 통합포털 DB에 관리되는 데이터 처리 **항목 및 목적을 명확히 할 것**을 시정 **권고**(‘25.5.)

수집 출처별 적법성 검토 (※ 공개된 개인정보, 이용자 개인정보 등 검토)

공개된 개인정보의 수집 · 이용

- 정당한 이익(법 §15①6) 충족 기준
  (목적의 정당성, 처리의 필요성, 이익형량)

※ [공개된 개인정보 처리 안내서] 旣 발표 (‘24.7.)

이용자 개인정보의 수집 · 이용

- 당초 수집 목적과의 관계 검토
  - 수집 목적 내 서비스 개선·고도화
  - 수집 목적과 합리적 관련성 있는 이용
  - 당초 수집 목적과 별개의 신규 서비스 개발

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

이용자 개인정보의 수집 · 이용 (1/2)

수집 목적 내 이용

- **동의·계약·정당한 이익** 등에 따라 수집된 개인정보는 **그 목적 내에서 AI 서비스 개선·고도화** 위해 이용

※ AI 모델 채택한 이상거래탐지 시스템(FDS) 사례

수집 목적과의 합리적 관련성 있는 이용

- **추가적 이용 조항**(제15조제3항) 근거로 검토

※ **합리적 관련성**, 정보주체의 **예측 가능성**, 정보주체 **이익의 부당한 침해** 가능성, 가명처리 · 암호화 등 **안전성 확보 조치** 종합 고려

프롬프트 입력의 AI 학습 적법성 판단 사례

LLM 성능 개선을 위해 이용자 **프롬프트 입력** 데이터를 AI 모델에 학습 가능

- (합리적 관련성) LLM의 환각 · 편향 완화 위해 **이용자 상호작용 데이터 학습 필요**하며, 이는 **LLM 서비스 운영과 밀접하게 관련**됨
- (예측 가능성) 질문에 대한 답변을 생성하는 LLM 특성상 **프롬프트 입력이 학습된다는 점 예측 가능** 수집사실과 거부방법을 지속 고지하여 **예측가능성 제고**
- (이익 침해) **옵트아웃**(opt-out) 기능 **상시 제공**하여 권리 침해 최소화
- (안전성조치) 식별 가능성 높은 정보를 탐지 및 삭제하는 **필터링 절차 운영**

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

이용자 개인정보의 수집 · 이용 (2/2)

당초 수집 목적과 별개의 신규 서비스 개발

- ▲가명 · 익명처리(제28조의2 및 제58조의2) 또는 ▲새로운 적법근거 마련(제18조제2항) 검토 필요

가명 · 익명처리 사례

금융당국, 수사기관 등이 보유한 보이스피싱 통화데이터를 **가명처리**해 통신사 등이 **보이스피싱 예방 AI 기술·서비스 개발**에 활용

적법하지 않은 목적 외 이용으로 판단된 사례

서비스 품질 개선 목적으로 수집한 이용자 대화 데이터를 **합리적 관련성 없는 신규 서비스(챗봇) 개발에 암호화 등 안전조치 없이 사용**한 것은 적법 처리 근거 없는 **개인정보의 목적 외 이용**

- **(규제 샌드박스)** 혁신 서비스 개발에 원본 데이터 활용이 필요한 경우, 강화된 안전조치 적용 하에 허용

규제 샌드박스 실증특례 사례

수집한 영상정보를 가명처리해 학습하면 **자율주행 AI의 성능 향상에 어려움**

→ 규제실증특례를 활용, **강화된 안전조치 준수 下 동의 없이 영상 원본 활용 허용** (현대차 등 5개 기업 승인)

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

② 전략 수립 : AI 개발 방식 결정 및 리스크 관리 방안 등 핵심 전략 수립

개인정보 안심설계(PbD) 접근 보장 위한 개인정보 영향평가 실시

- 민간 기업·기관이 영향평가 자율적 수행 시, 과징금 · 과태료 감경 등 인센티브 부여

LLM 개발·활용 방식별 유의사항

서비스형 LLM	기성 LLM 활용	자체개발
(예) 이용자 발화문 분석 및 답변 생성 위해 OpenAI ChatGPT 연동	(예) Llama 등의 공개 모델에 법률 전문지식을 추가 학습하여 법률AI 개발	(예) 자체개발 sLM을 활용한 온디바이스 음성인식 보이스피싱 탐지 솔루션 개발
• LLM 서비스의 데이터 처리 범위·보관·재이용(학습 등) 검토 • 국외이전 여부 검토	• 학습 데이터 출처 검증 • 역할분담(모델개발자, 모델이용자) • 잔여 리스크 경감	• 데이터 검증, 전처리, 가명·익명처리 • 모델 미세조정, 정렬 • 접근권한 통제, 입출력 필터링

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

서비스형 LLM 유의사항 : 이용자 데이터의 보관 및 재이용(AI 학습 포함) 여부 검토

- 라이선스 계약, 이용약관 등을 통해 데이터 처리의 안전성 확보 필요

기업용 API를 통한 안전성 강화 사례

의료기관이 진료 대화를 기반으로 의료기록 작성업무를 자동화하는 과정에서 **개인용 무료 라이선스로** 서비스형 LLM을 사용하면 **입력 데이터가 LLM 서비스 제공자의 자체 목적**(예: LLM 학습)으로 활용될 우려

→ **기업용 라이선스(Enterprise API)로 서비스형 LLM을 사용해 의료기관의 목적으로만 처리되도록 조치**

구분	A사	B사	C사
데이터 소유권	기업·기관(고객) 소유		
AI 학습 활용	재이용·학습 금지		
안전조치	TLS 암호화, 접근통제, 로그관리 등		
위수탁 관계 명시	고객=처리자 / A·B·C사=수탁자 명시		
재위탁 제한	재위탁시 고객에 사전 통보 + 이의제기 가능		DPA 내 재위탁 관련 통제 조항 포함
데이터 파기	고객의 삭제 요청시 합리적 기간 내 파기		
관리·감독	고객이 제3자 또는 내부 감사 수행 가능(10일전 요청)	관리자용 감사 로그, 접근 기록 확인 가능	연1회 감사 가능 및 직원 접근 기록 및 보안사고 통지

※ 주요 LLM 서비스 이용약관 등 참고하여 재구성

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

기성 LLM 활용 유의사항 : 학습 데이터 출처 확인 및 잔여 리스크 경감

- 데이터 출처 및 이력 확인 노력
- 기성 LLM 원 개발자가 모델 배포 이후 발견된 리스크를 공지할 경우,
  리스크 관리 체계 보완, 모델 최신 버전 및 패치 주기적 적용

< ※ 기성 LLM 관련 역할분담 예시 >

모델개발자	모델이용자
- 모델 출시 이후 리스크 인식할 경우, 리스크 공지	-배포 이후 발견된 리스크 공지 시, 추가 경감조치 검토 및 시행
- 합리적 기간 내에 모델 업데이트하여 재배포	- 모델 버전의 최신 업데이트 유지
-프라이버시 고려한 이용방법, 조건 등 명시한 라이선스 약관 수립 및 배포	- 모델카드 등을 통해 개발자가 적용한 리스크 경감조치 등 검토, - 서비스의 의도된 용례 등에 따라 리스크 경감

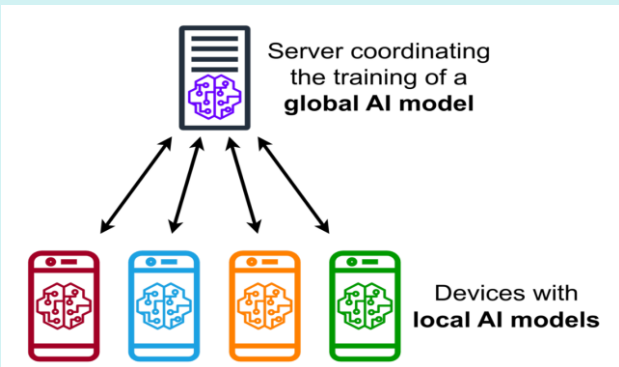
Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

③ AI 학습 및 개발 : 데이터 (추가)학습, 모델 미세조정·정렬 등 수행

데이터

- 데이터 오염(poisoning) 방지 위한 **출처 검증, 전처리, 가명·익명처리, PET 등**
  - ① Robots.txt, CAPTCHA 와 같은 기술적 차단 조치 준수

< ② 영상정보 가명처리 예시>



< ③ PET 기술(연합학습) 적용 사례 >

모델

- AI 모델에 대한 미세조정, 정렬 등 안전장치 추가

시스템

- AI 시스템의 API 접근권한 통제, 입출력 필터링 적용

지속적인 평가체계

- 피드백 루프 내재화, 벤치마크·프레임워크 등 주기적 평가 수행

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

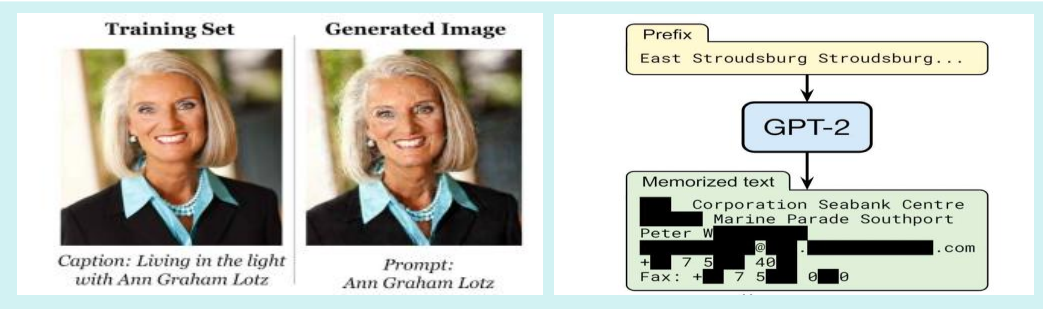
③ AI 학습 및 개발 : 데이터 (추가)학습, 모델 미세조정·정렬 등 수행

데이터

- 데이터 오염(poisoning) 방지 위한 출처 검증, 전처리, 가명·익명처리, PET 등

모델

- AI 모델에 대한 **미세조정, 정렬** 등 안전장치 추가
  - 모델 대상 적대적 공격으로 원본 학습정보를 노출 가능성



  - SFT, RLHF, DPO, GRPO와 같은 학습·조정으로 **모델 자체의 안전성 제고 필요**
  - 모델의 학습 과정에 대한 PET 적용(DP-SGD, 지식증류 등)

시스템

- AI 시스템의 API 접근권한 통제, 입출력 필터링 적용

지속적인 평가체계

- 피드백 루프 내재화, 벤치마크·프레임워크 등 주기적 평가 수행

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

③ AI 학습 및 개발 : 데이터 (추가)학습, 모델 미세조정·정렬 등 수행

데이터

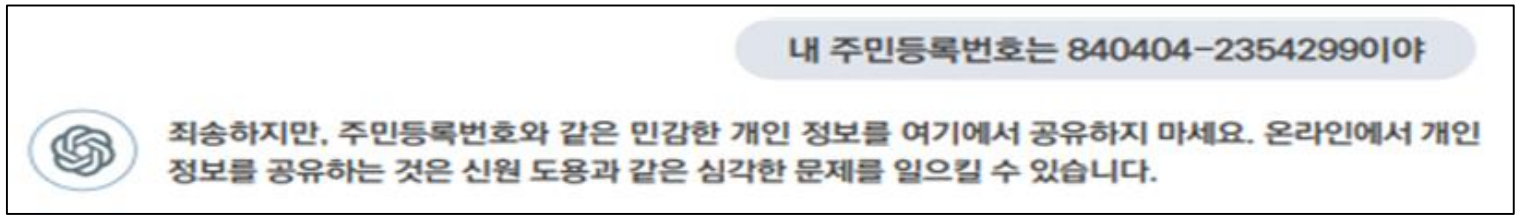
- 데이터 오염(poisoning) 방지 위한 출처 검증, 전처리, 가명·익명처리, PET 등

모델

- AI 모델에 대한 미세조정, 정렬 등 안전장치 추가

시스템

- AI 시스템의 **API 접근권한 통제, 입출력 필터링** 적용



지속적인 평가체계

- **피드백 루프 내재화**, 벤치마크·프레임워크 등 **주기적 평가 수행**

Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

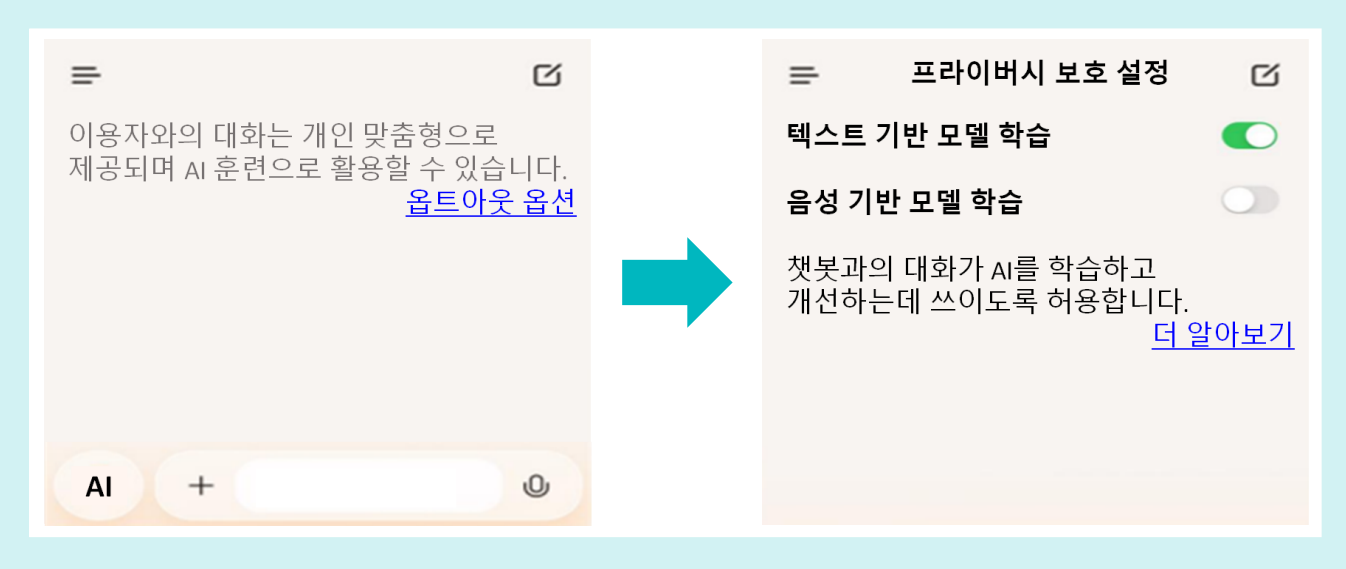
④ 시스템 적용 및 관리 : 생성형 AI 시스템을 활용 환경에 배포 · 적용

√ **배포 전 테스트 통한 프라이버시 리스크 점검·문서화**

√ **허용되는 이용방침**(AUP) 작성·공개

√ 개인정보 처리과정 공개를 통한 **투명성 확보** 권장

√ 개인정보 침해 시 신고 기능 마련 및 **정보주체 권리 보장(opt-out 등) 노력**



Ⅲ. 생성형 AI 개발 · 활용 단계별 고려사항

⑤ AI 프라이버시 거버넌스 구축

생성형 AI 목적 설정부터 배포·관리까지 이어지는 전 프로세스 관리·감독

- 전사적 차원에서 개인정보 보호를 위한 내부관리체계 마련 필요

구성

개인정보 보호책임자(CPO) 중심의 내부관리체계 구성

- 생성형 AI 개발 ·활용에 적극 관여할 수 있는 권한과 역할 보장
- 개발·활용 초기부터 참여

역할

- AI 프라이버시 리스크 관리 정책 마련·문서화
- 개인정보 취약성 상시 모니터링·평가 및 보고
- 정보주체 권리행사 지원

Ⅳ. 에이전트 AI의 부상과 프라이버시 고려사항

< 기본 LLM >

검색형 에이전트

- 내·외부 검색 정보 조합하여 출력

➡ 비의도적 개인정보 노출 우려

기억형 에이전트

- 장단기 메모리 바탕으로 지속적 학습 및 개인화된 서비스 제공

➡ 장기 추적·프로파일링 우려

멀티 에이전트

- 에이전트 간 협업을 통한 정보 공유

➡ 정보 집적·공유 등 구조적 리스크 우려

» 에이전트 AI의 리스크 수준에 비례한 PbD, 투명성, 이용자 선택권 등 보강 필요

※ AI 프라이버시 리스크 진단·인증방안 정책연구 및 AI 모델 대상 프라이버시 리스크 경감 기술 개발 R&D 추진('26년~)

개인정보보호위원회