




PIS FAIR 2025

인공지능과 개인정보처리의 주요쟁점 및 대응방안

고려대학교 정보보호대학원
권헌영 교수





01

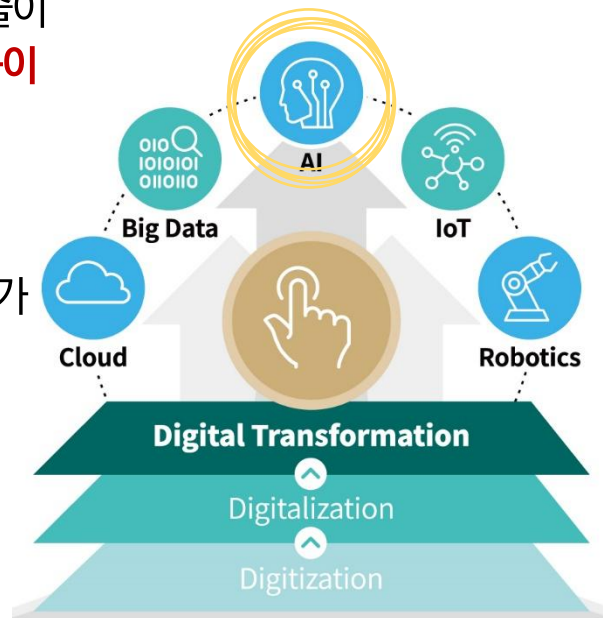
인공지능과 개인정보처리의 주요쟁점

01 AIX 시대의 도래

전 산업 분야가 조직 전반의 혁신을 위한 AI 기반의 전환 기술과 전략을 모색하고 있음

DX를 넘어 AIX로의 패러다임 전환

- 디지털 전환은 클라우드, 빅데이터, 사물인터넷, 로봇 등 다양한 기술이 견인하고 있으며, 현재 **인공지능이 두각**을 드러냄
- 2020년대 **생성형 AI 기술의 상용화**를 기점으로 인공지능을 업무의 도구로서 활용하는 사례가 증가
- 이에 따라 산업 전 분야가 **AI 기반의 전환**(AI-Based Transformation)을 통한 혁신의 가능성을 모색



AIX의 정의

AI 전환(AI Transformation, AIX)

“AI 전환은 단순히 AI 기술을 도입하는 것이 아니며, 조직의 체계·문화·전략에 **AI를 통합**하여 혁신을 이루는 과정”

01 AIX 시대의 도래

기술의 개발과 변화의 속도가 빠른 만큼, 인공지능 시대에서 개인정보처리 관련 보안 위협과 대응 방안 마련이 필요

인공지능 학습에 따른 프로파일링



인공지능이 방대한 데이터를 기반으로 학습 할 때, 단일 정보로는 식별이 어려우나 여러 데이터를 결합함으로써 개인을 식별하는 “**프로파일링(Profiling)**” 혹은 “**재식별(Re-identification)**” 문제가 발생함

인공지능 시스템에 대한 침해 및 오남용



인공지능 시스템의 경우 점점 더 활용 가치가 늘어나고 있으며, 각종 연구와 실험에도 동원되고 있고, 사용자들의 **민감한 데이터**들이 상대적 분량 뿐 아니라 절대적 분량에서도 많은 양을 구성하고 있어, 시스템 침해 및 오남용 시 미치는 **영향이 매우 큼**

I 02 인공지능 학습에 따른 프로파일링

단일 데이터로는 식별이 불가하나 여러 데이터들이 결합되며 개인을 재식별 할 수 있는 문제는, 인공지능을 학습시키기 위하여 방대한 데이터를 제공하였을 때 더 빈번히 나타날 수 있음

〈개인정보 재식별 관련 사고사례〉



2006년 8월, 미국 인터넷 기업인 AOL이 자사 검색 엔진의 로그 데이터 일부를 연구 목적으로 공개
개인정보 보호를 위해 사용자 이름이나, IP주소 등 직접 식별자를 제거하고, 각 사용자에게 임의의 숫자ID를
부여하는 가명처리를 거쳤으나, **복수의 검색어를 조합하여 개인을 재식별** 할 수 있는 사고 발생



2006년 10월, 넷플릭스는 자사의 추천 시스템인 알고리즘보다 뛰어난 알고리즘을 개발한 팀에게 상금을
수여하는 프로그램을 개최하고, 기존에 고객들이 평가한 평점 데이터를 익명화 하여 공개하였으나, 평가목록,
평가날짜 등을 이용한 **교차분석으로 개인을 재식별** 할 수 있는 사고 발생



2014년, 미국 MIT와 하버드대에서 MOOC학습 데이터 플랫폼인 edX에서 익명화 된 데이터를 공개하고
개인정보 보호를 위하여 비식별화 조치를 수행하였으나, 2021년 해당 데이터들에 대하여 익명화 기법의
취약성을 입증 및 개인을 재식별 할 수 있음을 증명

I 02 인공지능 학습에 따른 프로파일링

현행 개인정보보호법 상, 개인정보 수집은 ① 수집 목적 범위에서 이용 ② 사전 동의 ③ 최소한으로 수집 하도록 되어 있음

개인정보 보호법 제15조(개인정보의 수집 · 이용)

① 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 개인정보를 수집할 수 있으며 **그 수집 목적의 범위**에서 이용할 수 있다.

1. 정보주체의 동의를 받은 경우

2. 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우
3. 공공기관이 법령 등에서 정하는 소관 업무의 수행을 위하여 불가피한 경우

(이하 생략)

개인정보 보호법 제16조(개인정보의 수집 제한)

① 개인정보처리자는 제15조제1항 각 호의 어느 하나에 해당하여 개인정보를 수집하는 경우에는 **그 목적에 필요한 최소한의 개인정보를 수집**하여야 한다. 이 경우 최소한의 개인정보 수집이라는 입증책임은 개인정보처리자가 부담한다.

② 개인정보처리자는 정보주체의 동의를 받아 개인정보를 수집하는 경우 필요한 최소한의 정보 외의 개인정보 수집에는 동의하지 아니할 수 있다는 사실을 구체적으로 알리고 개인정보를 수집하여야 한다.

(이하 생략)

02 인공지능 학습에 따른 프로파일링

정보주체로부터 인공지능 학습의 목적으로 주체로부터 자료수집에 대한 동의, 최소한의 정보만 수집하여서 학습을 한다고 하여도 주체를 “재식별” 할 수 있게 되는 경우, 개인정보 유출의 위험 존재

개인정보 보호법 제2조(정의)

1. “개인정보”란 살아 있는 개인에 관한 정보로서 다음 각 목의 어느 하나에 해당하는 정보를 말한다.

가. 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보

나. 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보. 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다.

(이하 생략)

03 인공지능 시스템에 대한 침해 및 오남용

인공지능 시스템의 경우, 데이터 중 민감한 것이 많은 부분을 차지하고 있으나 사용자의 입력 데이터가 어떻게 처리되는지 등 “블랙박스” 형태로 구현되어 있는 경우가 많아 침해 및 오남용 사고 발생 시 사회적 영향이 일반 시스템 보다 큼



2025년 1월, **딥시크의 데이터베이스가 외부에 노출되어 있음**을 발견
 데이터베이스에 100만 건이 넘는 데이터가 인터넷에 노출되었으며 사용자 채팅기록, 비밀키, 백엔드 시스템 정보, API인증키 등 **매우 민감한 정보**들이 포함
 만약 사용자가 대화 중 자신의 이름, 주소, 연락처 등 개인정보를 입력했다면 해당 내용이 데이터베이스에 그대로 평문으로 저장되었을 것

Plain-Text chat messages from DeepSeek

```
td><td class="left">["disable_cache"]</td><td class="left">[""]</td><td class="left">2025-01-06 21:52:59.000000000</td><td class="left"></td><td class="left">otel-traces</td><td class="left"></td><td class="left">usage-checker</td><td class="left">{"JaegerTag":{"completion_tokens":745,"cost":"0.000247940","disable_cache":true,"finish_reason":"stop","input_len":521,"model":"deepseek-coder","msg":"介绍一下固体火箭助推器，可以包括其发明或发现、历史发展、历史意义、组成结构、工作原理、作用、未来发展等等。分段写，多写一点。","otel.library.name":"usage-checker","output_len":1359,"prompt_cache_hit_tokens":0,"prompt_cache_miss_tokens":281,
```

Which translates to

"Introduce solid rocket boosters, including their invention or discovery, historical development, historical significance, components, working principle, functions, and future developments. Write in sections with more details."

〈딥시크 데이터베이스에 **평문 형태로 저장된** 채팅 기록〉

03 인공지능 시스템에 대한 침해 및 오남용

인공지능 시스템의 경우, 데이터 중 민감한 것이 많은 부분을 차지하고 있으나 사용자의 입력 데이터가 어떻게 처리되는지 등 “블랙박스” 형태로 구현되어 있는 경우가 많아 침해 및 오남용 사고 발생 시 사회적 영향이 일반 시스템 보다 큼



2025년 2월, 딥시크에서 수집된 정보가 바이트댄스(제3자)로 제공되고 있음이 확인됨
딥시크는 바이트댄스 계열사의 클라우드 플랫폼을 이용하였고, 딥시크 앱 내에 바이트댄스 서비스와 연동되는 모듈을 포함시켜 바이트댄스 측으로 데이터를 전송함
사용자로부터 제3자에게 정보가 제공된다는 점에 대하여 동의를 받지 않았으며, 옵트아웃(Opt-out) 기능 역시 부재한 상황으로 인공지능 시스템을 오남용

03 인공지능 시스템에 대한 침해 및 오남용

인공지능 시스템의 경우, 데이터 중 민감한 것이 많은 부분을 차지하고 있으나 사용자의 입력 데이터가 어떻게 처리되는지 등 “블랙박스” 형태로 구현되어 있는 경우가 많아 침해 및 오남용 사고 발생 시 사회적 영향이 일반 시스템 보다 큼

2023년 3월, 오픈AI 내부 시스템에 대한 해킹 사고 발생
직원들이 기밀정보를 주고 받는 내부 메시징 시스템이 대상 시스템이었고, 이 사고의 경위와 파악된 사실들을 직원들에게는 내부적으로 공개하였음
그러나 고객 데이터나 사용자 개인정보가 유출되지 않은 것으로 판단하여 대외적으로는 공개하지 않았고, 2024년 7월 경 해당 사건이 외부로 알려지게 됨



2023년 7월, ChatGPT에서 한국 이용자들의 이름과 신용카드 번호 4자리까지 노출되는 사고 발생
서비스 속도를 올리기 위해 사용한 오픈소스 기반 캐시 솔루션에서 오류가 발생
오픈AI 측에서 보호조치를 하였음에도 사고가 난 사례로, 개인정보보호위원회는 안전조치 의무 위반으로 처분하지 않았으나, 유출 사실 인지 후 24시간 이내 신고하지 않아 과태료 부과



02

| 주요쟁점들에 대한 대응방안

01 공개된 개인정보 처리

「인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」를 활용한 법 규제 준수 및 위협 최소화

공개된 개인정보를 수집·이용 할 수 있는 기준

- 인공지능 개발을 위하여 공개된 개인정보가 수집, 이용될 수 있는 근거로서 개인정보 보호법 상의 ‘정당한 이익’ 적용 기준을 구체화

- 개인정보 보호법 제15조1항6호

① 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 개인정보를 수집할 수 있으며 그 수집 목적의 범위에서 이용할 수 있다.

6. 개인정보처리자의 **정당한 이익을 달성하기 위하여 필요한 경우**로서 명백하게 정보주체의 권리보다 우선하는 경우. 이 경우 개인정보 처리자의 정당한 이익과 상당한 관련이 있고 합리적인 범위를 초과하지 아니하는 경우에 한한다.



01 공개된 개인정보 처리

「인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」를 활용한 법 규제 준수 및 위협 최소화

공개된 개인정보를 수집·이용 할 수 있는 기준

○ 정당한 이익의 요건

구 분	주요 내용
목적의 정당성	개인정보처리자의 정당한 이익의 존재 → 공개된 개인정보 처리를 통해 개발하려는 AI의 목적, 용도를 구체화하여 정당한 이익을 명확화 (ex) 의료진단보조, 신용평가, 텍스트 생성 분류 번역 등을 수행하는 LLM 등
처리의 필요성	공개된 개인정보 수집, 이용의 필요성과 상당성, 합리성이 인정될 것 (ex) 의료진단보조 AI 개발 시 개인의 소득, 재산 등 관련 없는 정보는 학습 배제
구체적 이익 형량	개인정보처리자의 정당한 이익이 정보주체 권리에 명백히 우선 → ‘명백성’ 요건 충족을 위해 (i) 정보주체 권익침해 방지를 위한 안전성 확보조치 (ii) 정보주체 권리보장 방안 마련, 시행을 통해 개인정보처리자 이익이 우선하도록 조치



01 공개된 개인정보 처리

「인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」를 활용한 법 규제 준수 및 위협 최소화

공개된 개인정보를 수집·이용 할 수 있는 기준

- 공개된 개인정보처리 관련 대법원 판례 : 목적의 정당성



대법원 2016. 8. 17. 선고 2014다235080 판결(“로앤비 판결”)

이 사건 개인정보와 같이 국민 누구나가 일반적으로 접근할 수 있는 정보원(情報源)에 공개된 개인정보의 경우에는 이를 수집할 수 있는 ‘알 권리’가 정보처리자나 그로부터 정보를 제공받는 정보수용자에게 인정됨은 물론, 이러한 ‘알 권리’를 기반으로 하는 정보수용자들의 표현의 자유도 이 사건 개인정보의 처리행위로 보호받을 수 있는 법적 이익에 포함된다고 볼 수 있다. 또한 영업의 자유가 직업수행의 자유의 일환으로 헌법상 보장되므로, 기업이나 사인이 영리추구를 위하여 개인정보를 활용하는 것도 영업의 자유에 의하여 당연히 보장되는 영리활동에 해당한다고 할 것이고 나아가 피고 로앤비가 집적인 데이터베이스에서 정보를 편리하게 제공받고자 하는 사회적 수요가 존재한다면 그 정보제공으로 인하여 이러한 수요가 충족됨으로써 사회 전체의 경제적 효율성도 증가하게 될 것이다.

01 공개된 개인정보 처리

「인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」를 활용한 법 규제 준수 및 위협 최소화

공개된 개인정보를 수집·이용 할 수 있는 기준

- 공개된 개인정보처리 관련 대법원 판례 : 구체적 이익 형량



대법원 2016. 8. 17. 선고 2014다235080 판결(“로앤비 판결”)

개인정보자기결정권이라는 인격적 법익을 침해·제한한다고 주장되는 행위의 내용이 이미 정보주체의 의사에 따라 공개된 개인정보를 그의 별도의 동의 없이 영리 목적으로 수집·제공하였다는 것인 경우에는, 그와 같은 정보처리 행위로 침해될 수 있는 정보주체의 인격적 법익과 그 행위로 보호받을 수 있는 정보처리자 등의 법적 이익이 하나의 법률관계를 둘러싸고 충돌하게 된다. 이때는 정보주체가 공적인 존재인지, 개인정보의 공공성과 공익성, 원래 공개한 대상 범위, 개인정보 처리의 목적·절차·이용형태의 상당성과 필요성, 개인정보 처리로 인하여 침해될 수 있는 이익의 성질과 내용 등 여러 사정을 종합적으로 고려하여, 개인정보에 관한 인격적 보호에 의하여 얻을 수 있는 이익과 정보처리 행위로 얻을 수 있는 이익 즉 정보처리자의 ‘알 권리’와 이를 기반으로 한 정보수용자의 ‘알 권리’ 및 표현의 자유, 정보처리자의 영업의 자유, 사회 전체의 경제적 효율성 등의 가치를 구체적으로 비교 형량하여 어느 쪽 이익이 더 우월한 것으로 평가할 수 있는지에 따라 정보처리 행위의 최종적인 위법성 여부를 판단하여야 하고, 단지 정보처리자에게 영리 목적이 있었다는 사정만으로 곧바로 정보처리 행위를 위법하다고 할 수는 없다.

02 가명정보 제도 활용

안전하게 가명처리한 가명정보를 활용하고, 지속적인 모니터링을 통하여 재식별의 위험 최소화

가명정보 제도의 적극 활용, 모니터링을 통한 재식별 위험 최소화

〈개인정보 보호법 제28조의2, 제28조의3〉

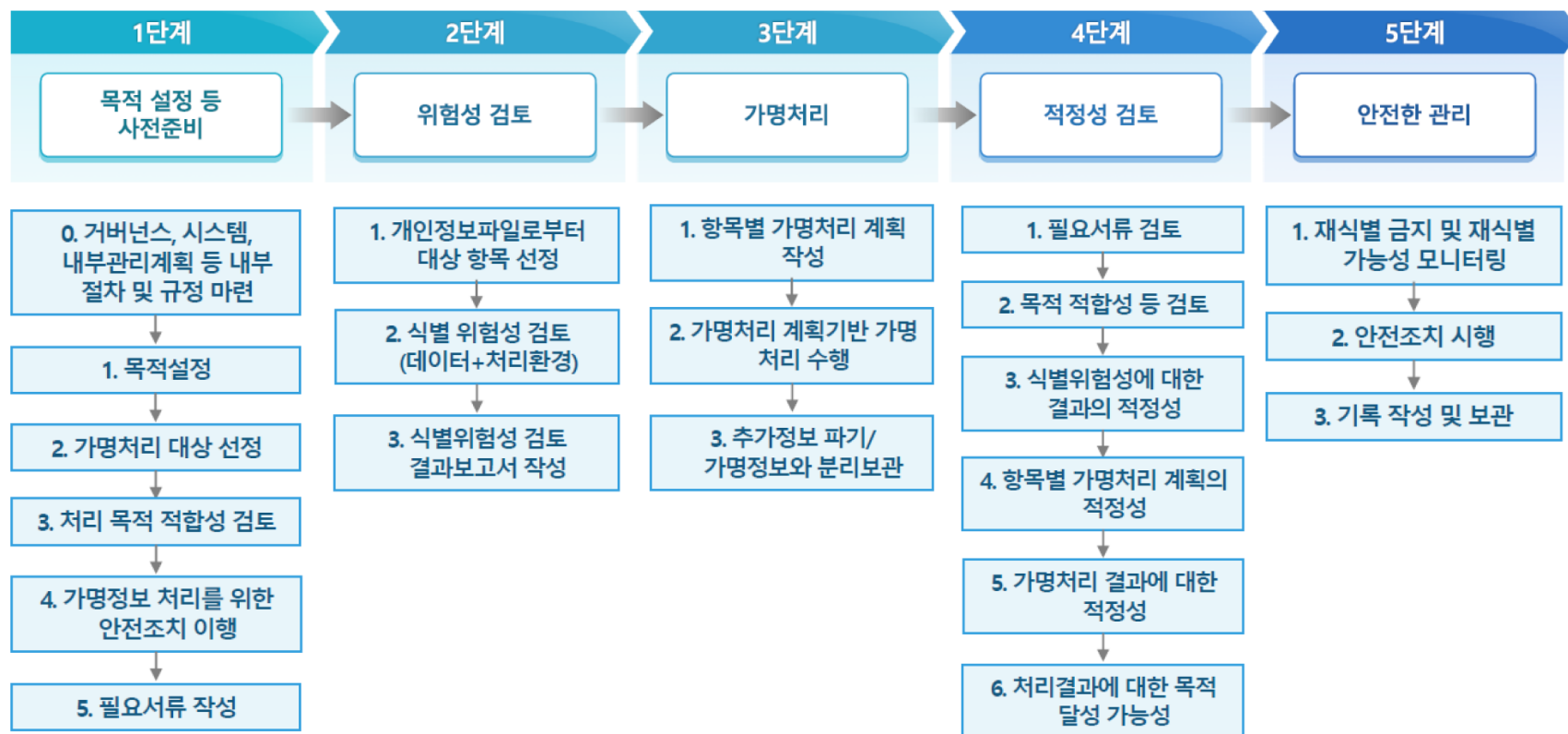
구 분	주요 내용
가명정보 처리	개인정보처리자는 “통계작성, 과학적 연구, 공익적 기록보존 등” 목적으로 정보주체의 동의 없이 가명정보 처리 가능
가명정보 제3자 제공	개인정보처리자는 가명정보를 제3자 제공하는 경우 “특정 개인을 알아보기 위하여 사용될 수 있는 정보”를 포함하여서는 안됨
가명정보 결합	서로 다른 개인정보처리자 간 가명정보를 결합하는 경우, 결합전문기관에서만 결합이 가능함

02 가명정보 제도 활용

안전하게 가명처리한 가명정보를 활용하고, 지속적인 모니터링을 통하여 재식별의 위험 최소화

가명정보 제도의 적극 활용, 모니터링을 통한 재식별 위험 최소화

- 개인정보의 가명처리 시, 5단계의 처리 절차를 통하여 안전하게 가명처리를 진행
- 개인정보처리자는 가명정보 처리 과정에서 특정 개인이 식별될 위험이 있는지 여부를 **지속적으로 모니터링**을 통하여 안전하게 관리해야 함



02 가명정보 제도 활용

안전하게 가명처리한 가명정보를 활용하고, 지속적인 모니터링을 통하여 재식별의 위험 최소화

가명정보 활용 관련 보호조치

관리적 보호조치

- 가명정보 및 추가정보를 안전하게 관리하기 위한 **내부관리계획**을 수립 및 시행
- 가명정보 처리 외부 위탁 시 위탁업무 수행 목적 외 **가명정보의 처리 금지에 관한 사항** 등을 포함한 문서작성 및 수탁자 관리감독
- 가명처리와 관련하여 **개인정보 처리방침**에 포함하여 공개 필요

기술적 보호조치

- 추가정보와 가명정보를 **분리하여 별도로 저장**, 관리하고 추가정보와 가명정보가 불법적으로 결합되어 재식별에 악용되지 않도록 접근통제 강화
- 가명정보 또는 추가정보에 접근할 수 있는 담당자에 대하여 **최소권한**으로 엄격히 통제 및 **접근 차등부여**
- 가명정보 처리 관련 기록 작성 및 보관

물리적 보호조치

- 가명정보 또는 추가정보를 전산실이나 자료보관실에 보관하는 경우 **비인가자의 접근으로부터 출입 통제**
- 가명정보 또는 추가정보가 보조저장매체 등에 저장되어 있는 경우 **잠금 장치가 있는 안전한 장소**에 보관 및 반입, 출입 등 통제를 위한 보안대책 마련

I 02 가명정보 제도 활용

안전하게 가명처리한 가명정보를 활용하고, 지속적인 모니터링을 통하여 재식별의 위험 최소화

Q. 개인정보 중 민감정보나 고유식별정보도 가명처리하여 활용할 수 있는가?

주민등록번호를 제외한 다른 고유식별번호와 민감정보는 가명처리하여 활용할 수 있으며, 주민등록번호는 법률, 대통령령 등의 구체적 근거가 있는 경우에 한하여 가능

Q. 가명처리 과정에서 일시적으로 가명정보와 추가정보가 같은 서버에 존재할 수 있는데, 이 경우 안전조치의무 위반인가?

가명정보와 추가정보를 분리, 보관하고 각 접근권한을 분리하도록 한 것은 가명처리 이후 재식별을 방지하기 위한 것으로, 가명처리 과정에서 일시적으로 가명정보와 추가정보가 동일 서버에 존재하는 것은 안전조치의무 위반에 해당하지 않음

Q. 가명정보를 제공하였을 시, 가명정보 활용 과정에서 생긴 문제에 대해서 제공자도 법적 책임이 있는가?

개인정보를 보호법에서 정한 처리 목적에 따라 가명처리하고 관련 안전조치 등 법률에서 정한 사항을 모두 준수하여 가명정보를 제공한 경우, 가명정보를 제공받은 자가 의도치 않게 특정 개인을 알아볼 수 있는 정보가 생성되었다는 사실만으로는 가명정보를 제공한 자에 대하여 개인정보 보호법상 행정처분을 하지 아니함

또한, 가명정보를 제공받은 자가 안전조치 미이행 등으로 가명정보를 유출하였거나 고의로 재식별 행위를 하는 경우, 해당 행위자만 제재함

03 인공지능 시스템의 관리적, 기술적 리스크 관리

[해외] 인공지능의 개발 및 활용에 대한 안전 확보를 위한 국제적인 규제 프레임워크 논의 활성화



[미국] '24년 하반기, 주정부 차원의 AI 규제 논의 활성화

- 2024년 기준 45개 주에서 약 700건의 AI 관련 법안이 발의됨
- 캘리포니아주 | AI 투명성법(SB942) 등 다수 AI 규제 법안에 서명
- 콜로라도주 | AI 소비자 보호법 제정



[일본] '24년 4월, 연성규범 방식의 AI 규제를 위한 가이드라인 발표

- 법적 구속력이 없는 자율 규제 형태의 'AI 사업자 가이드라인'을 발표함으로써 안전한 AI 개발 및 활용을 위한 지침을 사업자에게 제공



有 법적 구속력 無



[EU] '25년 2월, 유럽의회의 AI Act 단계적 시행

- 금지된 AI 시스템 관련 조항이 발효 및 집행되어
 - ▲ 의사결정 저해
 - ▲ 사회적 약자 차별
 - ▲ 소셜 스코어링
 - ▲ 생체정보 불법 수집
 등에 해당하는 인공지능 기술의 사용이 전면적으로 금지됨



[영국] AI 보안 연구소 도입 (前 AI 안전 연구소) 등 유연한 규제 방식 채택

- 유관 기관이 부문별 · 상황별 지침을 채택하는 유연한 규제 방식 채택
- AI 보안 연구소를 설립하여 AI에 대한 위험 평가 실시, 정보 공유 채널 구축 등의 목표를 달성하고자 함

03 인공지능 시스템의 관리적, 기술적 리스크 관리

[해외] 2025 Stanford AI Index Report, Responsible AI 관련 학회 제출 논문 키워드

● 분석 대상 학회

AAAI (Association for the Advancement of Artificial Intelligence)

AIES (AI, Ethics, and Society)

FACCT (Fairness, Accountability, and Transparency)

ICML (International Conference on Machine Learning) 등

● 개인정보 관련 카테고리에 포함된 키워드 목록

설계에 의한 프라이버시(Privacy by Design)

차등프라이버시(Differential Privacy)

데이터 거버넌스(Data Governance)

데이터 보호(Data Protection) 등



03 인공지능 시스템의 관리적, 기술적 리스크 관리

[해외] 인공지능 시스템에서의 개인정보처리 및 보호를 강화 시킨 해외 주요 사례



애플 : **차등프라이버시(Differential Privacy) 기법을 도입**

사용자 기기(iPhone)에서 수집한 정보에 임의의 노이즈를 추가함으로써, 개인을 식별할 수 없도록 함
시리(Siri)의 음성인식의 경우 기기 내에서 처리하여 네트워크를 통하지 않고 요청을 수행하도록 함
시리(Siri) 서버 측에서는 사용자 대신 무작위 식별자를 사용하여 데이터를 처리하고, 6개월이 지나면 식별자를 분리하여 사용자 식별을 방지



구글 : **연합학습(Federal Learning)과 차등프라이버시(Differential Privacy) 기법 도입**

스마트폰 키보드 앱 Gboard에서 다음 단어 추천, 자동완성 등의 AI기능을 높이기 위해서는 사용자들의 실제 타이핑 데이터를 학습해야 하나, 민감한 정보가 포함될 수 있어 프라이버시 침해 위협 존재
연합학습을 통해 사용자 기기에서 모델이 학습되고, 학습된 모델 업데이트만 중앙서버로 전송되며 원본 데이터는 서버로 보내지 않도록 함
중앙서버로 모아지는 데이터의 경우에도 차등프라이버시 기법을 적용하여 사용자 식별을 방지

03 인공지능 시스템의 관리적, 기술적 리스크 관리

[해외] 인공지능 시스템에서의 개인정보처리 및 보호와 관련한 해외 연구 내용

nature research

npj | Digital Medicine

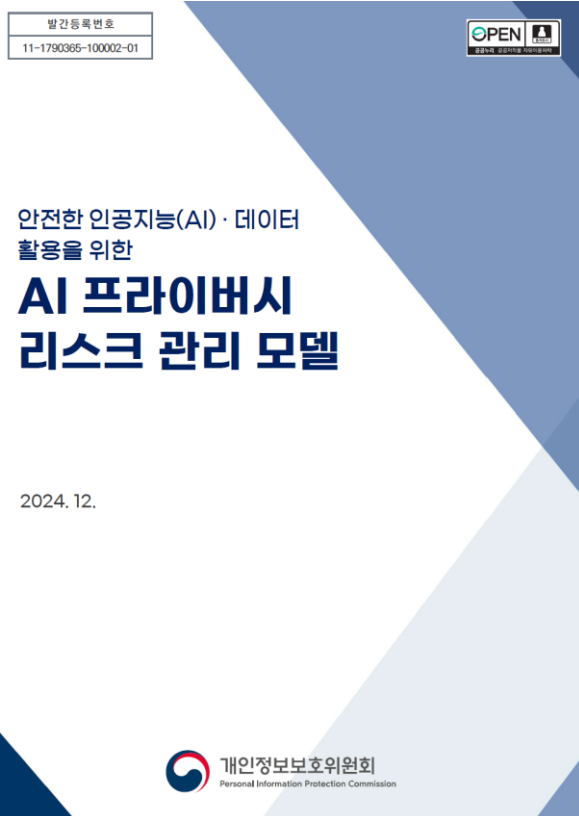
- 「Privacy-first health research with federated learning」 (Sadilek et al., 2021, npj Digital Medicine)

- 의료와 관련된 연구에서, 연합학습과 차등프라이버시 기법을 도입하여 환자의 로컬 장치에서 데이터를 보관한 채로 모델을 학습
- 중앙화된 데이터 저장소에 이를 전송하지 않고도 학습이 가능하다는 것을 입증
- 8개의 실제 임상, 역학 연구 데이터를 사용하여 기존 중앙집중 모델과 연합학습 모델을 비교
- 연합학습 모델과 중앙집중 모델이 유사한 성능을 보였으며, 차등프라이버시 기술 적용 시에도 성능 저하가 없었음을 입증

03 인공지능 시스템의 관리적, 기술적 리스크 관리

2024.12. 개인정보보호위원회, 「AI 프라이버시 리스크 관리 모델」 발표

인공지능 시스템 개발자 · 운영자는 **개별 시스템 환경 및 맥락에 따라** 아래 사항들을 기반으로 **최적의 안전조치 조합 마련**이 가능



인공지능 시스템 관리적 안전조치

- 학습데이터 출처, 이력관리
- 안전한 보관, 파기 방안 마련 및 실행
- 인공지능시스템 참여자 간 역할 명확화
- 허용되는 이용방침 작성, 공개
- 자동화된 결정에 대한 개인정보처리자의 조치, 기준 준수
- AI프라이버시 레드팀 구성, 운영
- 정보주체 신고방안 및 조치방안 마련

인공지능 시스템 기술적 안전조치

- 학습데이터 전처리
- 학습 시, 합성데이터 사용 고려
- 모델 미세조정을 통한 안전장치 추가
- 입출력 필터링 적용
- 차분 프라이버시 기법 적용
- 출처데이터 추적 및 합성 콘텐츠 탐지방안 마련
- 생체정보 가명, 익명처리



THANK YOU