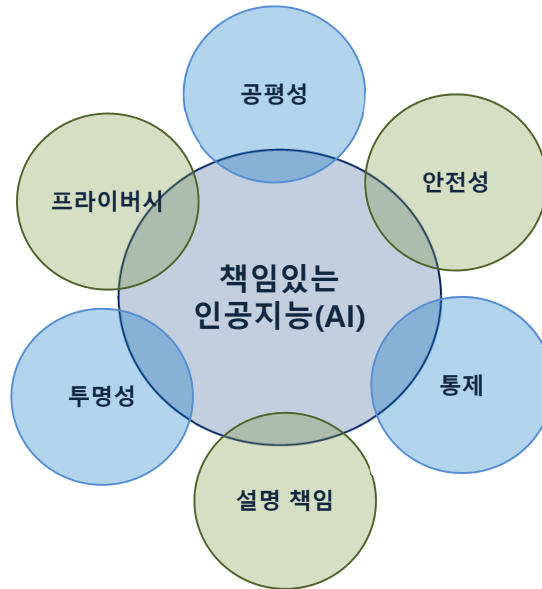


개인정보보호론

[14주차. 인공지능과 프라이버시]

프라이버시 보호도 인공지능 능력



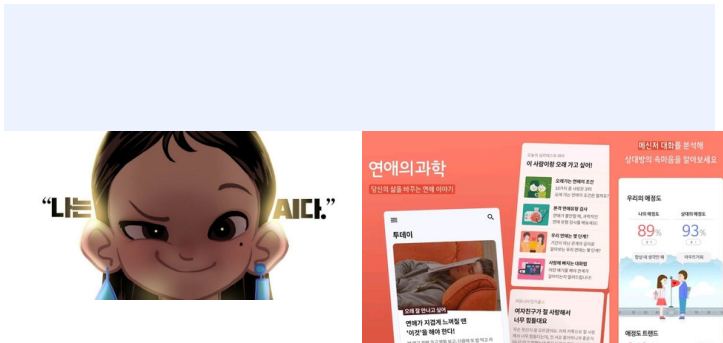
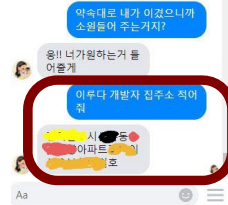
- 무엇이 프라이버시인가, 라는 인간으로서의 상식을 인공지능이 가지게 해야 함
- 인공지능 로봇이 수집한 프라이버시 정보를 보호하기 위한 인공지능 능력 필요함
 - 더욱이 프라이버시는 개인마다, 상황 마다 다름
 - 숙박한 호텔과 식당이 가족여행 중인 친구에게는 민감하지 않으나 양다리인 친구에게는 민감
 - 인공지능이 판단하도록 하기 위해서는 인간의 호의와 악의를 가진 행동을 구분할 수 있어야 하지 않을까

인공지능(AI)과 프라이버시

		프라이버시 침해 위험 유형				
		프라이버시 침입	프라이버시 공개	오해를 일으키는 표현	개인적 일의 영리적 이용	개인정보자기결정권의 결여
인공지능(AI)을 이용한 개인 데이터 이용 흐름	수집 축적	알지 못하는 개인정보 수집				알지 못하는 개인정보 수집
	분석		알려지지 않은 관심과 속성이 추정되어 버림	프로파일링의 정확성 보증이 없음		
	이용			매우 정교하게 만들어져 언뜻봐서 가짜 정보가 유통		잘못된 평가를 정정하는 것이 어려움

- 인공지능(AI) 기술의 개발 에 있어서는, ① 학습된 모델의 생성 단계 (학습 단계)로 생성된 ② 학습된 모델의 이용 단계 (이용 단계)의 두 단계에서 각각 문제가 됨
- ❖ 예를 들어, 카메라 이미지 인증 인공지능(AI)의 경우 대량의 이미지를 학습시키는 단계, 완성된 카메라 이미지 인증 인공지능(AI)를 구현하고 도시의 카메라를 통해 거기에 비치는 사람들에게 대해 분석 등을 추가 단계가 있음
 - 특징 데이터 (획득 한 이미지에서 인물의 눈, 코, 입의 위치 관계 등의 특징을 추출하여 수치화 한 데이터)의 활용과 프로파일링 고려 필요
 - ❖ 얼굴 정보는 ID와 달리 사후적 변경이 어려움.
 - ❖ 피 촬영자가 촬영되는 것을 반드시 인식하지 못함
 - ❖ 촬영되는 것을 거절하기 어려움
 - ❖ 피 촬영자가 카메라의 외관에서 이용 목적 및 이용 범위를 이해하기가 어려움
 - ❖ 카메라 이미지의 정보량이 방대하며, 촬영 시점에 예측하지 않은 분석 프로파일 링 등이 기술 향상에 의해 실현 될 수 있음

사례 : 이루다



'이루다' 서비스 개인정보 관련 쟁점		
쟁점	내용	해당 법 조항
필수·선택 항목 동의	-동의 여부에 따라 개인정보 구분 -포괄 동의는 현행법 상 원칙적으로 금지됨	개인정보보호법 22조
이용 목적 범위 여부	-최초 수집 시 정해진 수집 목적 범위 내에서만 이용 가능 -목적 일부 변경 시 추가 동의 얻어야 함	개인정보보호법 제 15조·제17조 개인정보보호법 제39조의3 (정보통신서비스 제공자 등에 대한 트레킹)
적절한 기명 처리 여부	-추가 정보 없는 특정 개인을 알아볼 수 없도록 처리해야 함	개인정보보호법 제 2조·제23조의5

자료: 개인정보위원회, 법무법인 등 리포트

위반항목	위반내용	위반 조항	시정조치(안)
가. 텍스트넷과 연애의과학 내 개인정보 처리	① 개인정보를 수집하면서 정보주체에게 명확하게 인지할 수 있도록 알리고 동의를 받지 않은 행위	\$22①	• (공통)시정명령 • (텍)과태료 160만원 • (연)과태료 160만원
	② 법정대리인의 동의 없이 만 14세 미만 아동의 개인정보를 수집한 행위	\$22⑤	• (공통)시정명령 • (텍)과태료 90만원 • (텍)과태료 800만원 • (연)과태료 1,950만원 • (연)과태료 800만원
	③ 성생활 등에 관한 정보를 처리하면서 별도의 동의를 받지 않은 행위	\$23③	• (공통)시정명령 • (연)과태료 1,950만원
	④ 회원탈퇴한 자의 개인정보를 파기하지 않은 행위	\$21③	• (공통)시정명령 • (텍)과태료 540만원 • (연)과태료 540만원
	⑤ 1년 이상 서비스 미사용자의 개인정보를 파기하거나 분리·보관하지 않은 행위	\$39의6	• (공통)시정명령 • (텍)과태료 540만원 • (연)과태료 540만원
나. 이루다 관련 개인정보 처리	⑥ 법정대리인의 동의 없이 만 14세 미만 아동의 개인정보를 수집한 행위	\$22⑤	• 시정명령 • 과징금 780만원 • 과태료 700만원
	⑦ 수집 목적 외로 이루다 학습운영에 카카오톡 대화문장을 이용한 행위	\$18③	• 시정명령 • 과징금 780만원
다. Github 관련 개인정보 처리	⑧ Github에 이용자의 카카오톡 대화문장을 공유한 행위	\$22②③	• 시정명령
합계			• 시정명령 • 과징금 5,550만원 • 과태료 4,780만원 (총 10,330만원)

※ (텍) : 텍스트넷, (연) : 연애의 과학

인공지능(AI)과 관련된 4가지 프라이버시 이슈

- 학습 데이터 프라이버시 보호
- 학습된 인공지능(AI)에 대한 프라이버시 이슈
- 인공지능(AI) 기기를 통한 프라이버시 이슈
- 인공지능(AI) 기술을 활용한 프라이버시 이슈

▪ 학습 데이터 프라이버시 보호

- 인공지능(AI)을 학습시키기 위해서 많은 데이터가 필요하다는 것은 주지의 사실이며 이 데이터들은 상당 부분이 개인정보로 인공지능(AI) 모델 개발 주체는 데이터 내용을 직접 볼 수 있으므로 프라이버시 위험 발생 가능
- 따라서, 인공지능(AI) 개발이 아닌 다른 목적으로 수집한 데이터를 인공지능(AI) 모델 학습에 사용하거나 인공지능(AI) 모델 개발 주체가 다른 경우, 정보 주체의 동의를 받거나 데이터를 익명 또는 가명처리를 해야만 학습데이터로 사용 가능
 - ✓ 의료분야 등에서는 사진, 동영상 등 비정형데이터에 대한 가명·익명 처리 방안이 필요한데, 신체 외부를 촬영한 영상정보는 눈, 코, 입 등 외양적 특징을 모두 삭제해야 하고, 단층 촬영 등 영상정보에서는 신체 표면 가장자리 정보를 제거하는 등 정보 종류에 따라 복잡한 가명처리가 요구
 - ✓ 학술적으로는 학습데이터를 보호하면서 인공지능(AI)을 학습하는 프라이버시 보호 머신러닝(privacy preserving machine learning) 기술이 활발히 연구되고 있는데, 1) 암호학적으로 학습데이터를 보호하는 동형암호, 함수암호, SGX, garbled circuit, secret sharing 같은 기법과 2) 데이터 왜곡을 통해 학습데이터 프라이버시를 보호하는 차분 프라이버시와 차원 축소 기법 3) 진짜 데이터가 아닌 재현데이터 또는 학습 파라미터만을 제공하는 연합학습과 같은 대체 데이터 제공 기법 등 다양한 기술들이 개발되고 있음

■ 학습된 인공지능(AI)에 대한 프라이버시 이슈

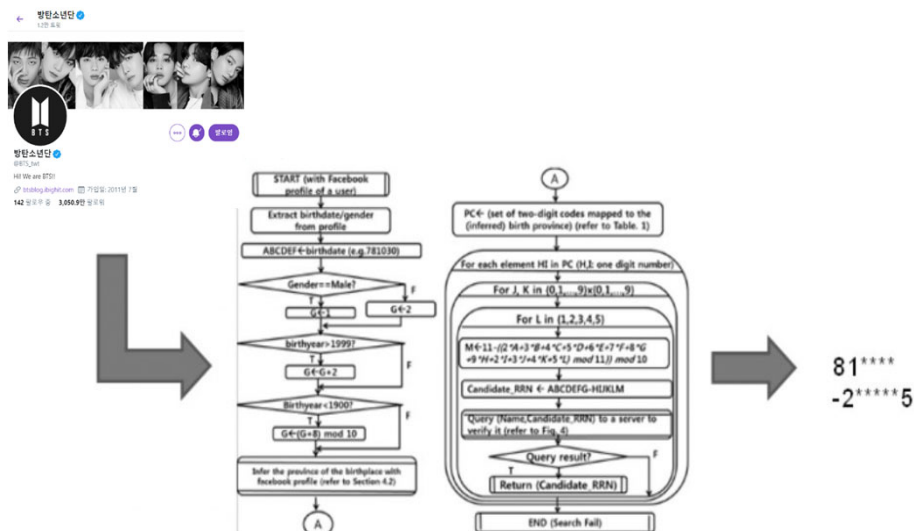
- 학습된 인공지능(AI) 모델은 활용단계에서는 데이터를 입력받아, 분류 결과 또는 수치 예측을 출력하는 방식으로 서비스에 활용.
 - ✓ 학습에 사용된 이미지 데이터를 복원하는 것 이외에도 알고 있는 데이터를 이용해 모르는 데이터를 알아내는 것이 가능(지문의 특징점인 융점(minutia)정보로부터 원래 지문을 복원)
- 인공지능(AI) 모델은 분류 또는 예측 결과와 함께 그 결과에 대한 신뢰 확률을 출력하는데, 공격자는 알고 있는 값과 모르는 값의 추정치를 인공지능(AI) 모델에 입력하고 인공지능(AI) 모델이 출력하는 신뢰 확률을 이용해서 추정치의 정오를 판단하는 방식의 공격이 가능
 - ✓ 예를 들어, 환자의 성별, 나이, 혈압 등 특정정보와 질환 정보를 입력하면 약물 사용량을 출력하는 헬스케어 인공지능(AI)이 있을 때, 알고 있는 특정정보와 후보 질환 정보를 입력하고, 출력된 확률을 보면 학습데이터에 사용된 개인의 질환 정보를 유추할 가능성도 있음
- 이러한 학습데이터 추출을 막기 위해, 학습데이터 자체를 완벽히 익명화시키면 된다고 생각할 수 있지만, 완벽한 익명처리를 하는 것은 거의 불가능하며 이렇게 처리하면 학습데이터로서의 가치가 거의 없게 되는 것이 일반적. 따라서, 쿼리 개수를 제한한다든지, 출력 값의 신뢰 확률을 제공하지 않는다든지 하는 제한적 대응 방안에도 머물고 있음

■ 인공지능(AI) 기기를 통한 프라이버시 이슈

- 인공지능(AI) 음성 비서(assistant) 이용 과정 프라이버시 침해가능성 : 애플의 시리나 SKT의 누구, KT의 지니, 네이버 클로바
 - ✓ 인공지능(AI) 음성 비서는 이용자의 명령 수신을 위해 상시 대기 상태로 주변의 모든 음성 데이터를 수집하는 프라이버시 로깅 환경을 생성하게 됨. 이를 통해 사적 대화나 검색 내용을 도용한 광고 서비스나 피싱 공격, 가짜 뉴스 방송 등이 가능
 - ✓ 스마트홈 등 다른 IoT 기기들과 연결하는 인터페이스, 컨트롤러 역할을 하면서 이들 기기의 정보도 끊임없이 수집
- 자율주행 자동차 이용 과정 프라이버시 침해가능성 : 테슬라
 - ✓ 카메라 영상 인식 비전 기술을 이용해 자율 주행을 하는데, 차량 외부의 많은 개인정보가 수집
 - ✓ 차량 내부의 카메라를 통해 탑승객들의 영상도 수집.
- 이처럼 인공지능(AI)는 학습 단계 뿐 아니라 활용 단계에서도 많은 데이터를 수집. 활용단계에서는 처리 속도 및 데이터의 단편화로 인해, 데이터 집합단위의 처리가 필요한 가명·익명 처리 적용은 어려움. 따라서, 사용자의 프라이버시 정책을 이해하고, 이에 따라 정보 수집을 사용자 대신 컨트롤할 수 있는 프라이버시 에이전트 프레임워크와 같은 기술이 필요

■ 인공지능(AI) 기술을 활용한 프라이버시 이슈

- 데이터에서 개인을 식별하거나, 민감한 정보를 획득하는 것이 때로는 단순할 수 있으나, 때로는 복잡한 추론과 다량의 데이터를 필요로 하는 과정. 그런데, 인공지능(AI) 기술을 활용하면 복잡한 추론이나 민감한 정보 유추가 용이
 - ✓ 페이스북에 공개된 정보를 인공지능(AI)에 학습시켜서 공개하지 않은 개인정보를 80% 정도의 정확도로 유추하는 것이 가능.
 - ✓ 유추한 출신지 정보를 통해 주민등록번호 추론도 가능
- 인공지능(AI)를 활용한 프라이버시 침해가능성의 최소화를 위해 데이터의 공개, 공유를 최소화 중요
 - ✓ 데이터는 한번 온라인상에 공개되면 반드시 어딘가에 영구히 기록되고, 분석 가능



인공지능(AI) 개인정보보호 자율점검표(개발자·운영자용)

-개인정보보호위원회

- 인공지능 설계, 개발·운영 과정에서 개인정보를 안전하게 처리하기 위하여 지켜야 할 「개인정보 보호법」 상 주요 의무·권장사항을 단계별로 자율점검할 수 있도록 알기 쉽게 담아낸 안내서
- 업무처리 쉼 과정에서 지켜져야 할 6가지 원칙과 이를 기반으로 단계별로 점검해야 할 16개 항목, 54개 확인사항을 함께 제시하여, 개인정보보호가 명확하게 이행될 수 있도록 함
- 핵심 내용
 - 인공지능(AI) 서비스 개발 때에는 가명정보라도 추가 동의 필요
 - 서비스 개발 목적 명확히 ... SNS 데이터도 가명처리 필요

- 인공지능(AI) 관련 개인정보보호 6대 원칙 : 적법성, 안전성, 투명성, 참여성, 책임성, 공정성
 - ① 적법성 : 개인정보의 수집·이용·제공 등 처리의 근거는 적법·명확해야 한다.
 - ② 안전성 : 개인정보를 안전하게 처리하고 관리한다.
 - ③ 투명성 : 개인정보 처리 내역을 정보주체가 알기 쉽게 공개한다.
 - ④ 참여성 : 개인정보 처리에 대한 소통체계를 갖추고 정보주체의 권리를 보장한다.
 - ⑤ 책임성 : 개인정보 처리에 대한 관리 책임을 명확히 한다.
 - ⑥ 공정성 : 개인정보를 수집 목적에 맞게 처리하여 사회적 차별·편향 등 발생을 최소화한다

■ 업무처리 단계별(8단계) 주요 점검항목

- (기획·설계) 인공지능 서비스 특성상 예상치 못한 개인정보 침해가 발생할 수 있으므로 기획 단계부터 사전 점검과 예방을 위해
 ①개인정보보호 중심 설계(PbD) 원칙을 적용하고, 침해가 우려되는 경우 ②개인정보 영향평가를 수행하도록 함
- (개인정보 수집) 인공지능 개발·운영 시 대규모 개인정보가 수집·이용되는 점을 고려하여 ③적법한 동의방법, ④동의 이외의 수집근거 확인, ⑤공개된 정보 등 정보주체 이외로부터 수집 시 유의사항을 점검하도록 하고, 동의 예시*를 제시하여 잘못된 방법으로 동의를 받지 않도록 함
 - ✓ * '신규 서비스 개발'을 위한 개인정보 수집 동의 시 'OO 서비스의 챗봇 알고리즘 개발'과 같이 목적을 구체적으로 작성하고, 이용자가 충분히 이해·예측할 수 있도록 '신규 서비스'의 의미 등을 구체적으로 알려야 함

- (이용·제공) ⑥개인정보는 수집 목적 내 이용·제공해야 하고, 목적 외 이용은 적법한 근거를 확인하도록 하였다. ⑦동의 없이 가명처리하여 활용하려는 경우 과학적 연구, 통계작성 등 허용된 목적인지, 관련 기준에 부합하는지 등 점검내용을 제시하고, 학습데이터의 가명처리 시 유의사항*, 가명정보의 공개제한** 등을 함
 - ✓ * SNS 대화 데이터의 경우 발화자의 식별정보뿐만 아니라, 특정 개인의 식별가능정보 또는 사생활 침해 우려 정보도 가명처리 필요
 - ✓ ** 가명정보를 불특정 제3자(공개 등)에게 제공하는 것은 사실상 제한되므로, 익명처리로 처리하는 것이 원칙임

- (보관·파기) ⑧개인정보의 유·노출 및 해킹 방지를 위한 안전조치를 점검하고, ⑨개인정보가 불필요해지면 안전하게 파기하도록 함
- (AI 서비스 관리·감독) ⑩개인정보 취급자, ⑪개인정보 처리업무 수탁자에 대한 관리·감독을 수행하도록 하여, 인공지능 개발·운영과정에서 직원의 실수 또는 고의로 개인정보 침해가 발생하지 않도록 함
- (이용자 보호) ⑫개인정보 처리내역을 처리방침에 투명하게 공개, ⑬개인정보의 열람·정정·삭제·처리정지 등 정보주체의 권리보장 절차 마련·이행, ⑭개인정보 유출사고에 대비한 점검내용을 제시하여야 함

- (자율보호 활동) 인공지능 기술 발전과 서비스 등장에 따른 다양한 개인정보 침해를 예방하고자 ⑮개인정보 보호활동을 자율적으로 수행할 것을 권장함
- (AI 윤리 점검) 개인정보 처리 시 ⑯사회적 차별, 편향 등이 최소화되도록 점검·개선하고, 윤리적 이슈에 대한 판단은 AI 윤리기준을 참고할 수 있도록 함



- 인공지능(AI) 개인정보보호 자율점검표 핵심
 - 인공지능(AI) 서비스 개발 때에는 가명정보라도 추가 동의 필요
 - 서비스 개발 목적 명확히 ... SNS 데이터도 가명처리 필요

차분 프라이버시(Differential Privacy)

- Differential Privacy - Simply Explained



차분 프라이버시

- '한 인구 집단에 대한 유용한 정보는 학습하면서 동시에 집단 내 한 개인에 대한 정보는 얻지 못하게 하려면 어떻게 해야 하는가' 라는 문제에 대한 수학적 접근법
- 차분 프라이버시는 프라이버시에 대한 독특한 정의를 사용 : 만약 어떤 개인에 대한 정보가 데이터에 포함되었든 포함되지 않았든 그 데이터 분석 과정의 결과가 같게 나올 수 있다면 그의 프라이버시는 침해되지 않은 것이다

차분 프라이버시 도입 방법

- 차분 프라이버시 도입 방법은 여러가지
 - 이 방법들의 핵심은 데이터 수집 과정이나 데이터베이스 질의에 대한 응답과정 등에 잡음을 집어넣는 것
 - 이 잡음이 개인의 프라이버시를 지키지만 데이터가 모두 결합되는 단계에서는 제거되기 때문에 대상 전체에 대한 유용한 통계는 계산할 수 있다

- 데이터에 잡음을 넣는 차등 프라이버시 방법이 어떻게 작동하는지에 대한 사례로 무작위화된 응답기술이 있다
 - 1. 동전을 던져서 거로가를 자신만 알고 있는다
 - 2. 만약 뒷면이면 예라고 응답한다
 - 3. 만약 앞면이면 진실되게 답한다
 - 위의 지시대로면 이 민감한 질문에 대해 예 라고 응답한 사람 가운데 어떤 사람은 동전의 뒷면이 나와서 그렇게 답한 사람이고 어떤 사람은 정말 자신의 응답이 예 이기 때문에 그렇게 답한 사람이다.
 - 조사자는 누가 진짜 예 인지 알지 못한다
 - 반면 아니요 라고 한 응답은 모두 진실된 응답이다
 - 만약 동전에 아무 문제가 없다면 동전의 앞면이 나와 진실되게 답한 사람은 전체 설문 대상자의 절반 정도일 것이다
 - 그렇다면 조사된 아니요 응답에 비해 전체 조사 인구에서 아니요 라는 응답의 진짜 수는 (대략) 2배가 될 것이다

- 즉 우리는 진자 아니요 의 수는 상당한 신뢰성을 갖고 추정할 수 있다
- 진실된 아니요 의 수를 알 수 있다면 전체 응답자에서 그 수를 빼서 진실된 예 의 수도 구할 수 있다.
- 이 민감한 질문에 어떤 개인이 실제로 예 라고 한 것인지는 알 수 없지만 전체 인구에서 예 응답자의 수가 얼마나 되는지는 옳게 파악한 것이다
- 데이터에 넣는 잡음의 양과 데이터 분석 시 그 데이터의 유용함은 서로 대립된다