발간등록번호
11-1790365-000029-01

가명정보 처리 가이드라인

2024. 2.

개인정보보호위원회
Personal Information Protection Commission

가명정보 처리 가이드라인

목차

가명정보 처리
가이드라인

제1장

가이드라인 개요

1 목적

◉ 빅데이터, AI 등 다양한 융·복합 산업에서의 **데이터 이용 수요가 급증하는 가운데,** 데이터 활용의 핵심인 가명정보 활용을 위한 법적 근거가 마련됨에 따라

◉ 가명정보 활용에 필요한 **가명정보 처리 목적, 처리 절차 및 방법, 안전조치에 관한 사항 등을 안내**하여 안전한 데이터 활용 환경을 마련하고자 함

☑ 4차 산업혁명 시대 신성장 동력인 '데이터' 활용에 대한 시대적 요구를 반영한 데이터3법*이 시행('20.8.5.)되어 개인정보처리자가 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 목적으로 개인정보를 가명처리하여 활용할 수 있는 기반이 새롭게 마련됨

* 개인정보 보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 '정보통신망법' 이라 함), 신용정보의 이용 및 보호에 관한 법률(이하 '신용정보법' 이라 함)

☑ 본 가이드라인은 「개인정보 보호법」(이하 '보호법'이라 함) 개정 및 시행('20.8.5.)으로 새롭게 도입된 '가명정보 처리에 관한 특례'(보호법 제3장 제3절)에 관한 설명과 구체적 사례를 제공함으로써 가명정보의 처리에 대한 이해를 돕고, 처리 과정에서 발생할 수 있는 개인정보 오·남용을 방지하여 안전한 가명정보 활용 방안을 안내하기 위해 작성하였음

※ '23년 3월 14일 개정되어 동년 9월 15일 시행된 '보호법' 등 법령의 개정 내용과 비정형데이터의 가명처리와 관련한 내용 등을 추가로 반영

▷ 개인정보처리자가 법에 따른 규정을 준수한 경우 가이드라인 미준수를 사유로 처벌받지 않음. 따라서, 개인정보처리자는 데이터의 관련 분야 및 특수성 등을 고려하여 상황에 따라 유동적으로 처리 가능함

2 적용 대상

☑ 본 가이드라인의 적용 대상은 보호법(제3장 제3절 가명정보 처리에 관한 특례)에 근거한 가명정보 처리이며 개인정보보호위원회(이하 '개인정보위'라 함)와 소관 부처가 공동으로 발간한 개인정보의 가명정보 처리에 관한 분야별 가이드라인*이 있는 경우에는 해당 분야의 가이드라인을 우선 적용함

* 보건의료 데이터 활용 가이드라인(보건복지부), 교육분야 가명·익명정보 처리 가이드라인(교육부), 공공분야 가명정보 제공 실무안내서(행정안전부) 등

☑ 또한, 본 가이드라인은 '가명정보 처리에 관한 특례'(보호법 제3장 제3절)에 근거하여 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 가명정보의 처리에 참고할 수 있도록 작성함

※ 보호법 제15조 제3항 및 제17조 제4항 등에 근거한 가명처리는 본 가이드라인의 적용대상이 아니지만, 가명처리에 관한 기술적 내용 등은 참고할 수 있음

개인정보 보호법

제15조(개인정보의 수집·이용) ③ 개인정보처리자는 당초 수집 목적과 합리적으로 관련된 범위에서 정보주체에게 불이익이 발생하는지 여부, 암호화 등 안전성 확보에 필요한 조치를 하였는지 여부 등을 고려하여 대통령령으로 정하는 바에 따라 정보주체의 동의 없이 개인정보를 이용할 수 있다.

제17조(개인정보의 제공) ④ 개인정보처리자는 당초 수집 목적과 합리적으로 관련된 범위에서 정보주체에게 불이익이 발생하는지 여부, 암호화 등 안전성 확보에 필요한 조치를 하였는지 여부 등을 고려하여 대통령령으로 정하는 바에 따라 정보주체의 동의 없이 개인정보를 제공할 수 있다.

개인정보 보호법 시행령

제14조의2(개인정보의 추가적인 이용·제공의 기준 등) 시행령 제14조의2(개인정보의 추가적인 이용·제공의 기준 등) ① 개인정보처리자는 법 제15조제3항 또는 제17조제4항에 따라 정보주체의 동의 없이 개인정보를 이용 또는 제공(이하 "개인정보의 추가적인 이용 또는 제공"이라 한다)하려는 경우에는 다음 각 호의 사항을 고려해야 한다.

1. 당초 수집 목적과 관련성이 있는지 여부
2. 개인정보를 수집한 정황 또는 처리 관행에 비추어 볼 때 개인정보의 추가적인 이용 또는 제공에 대한 예측 가능성이 있는지 여부
3. 정보주체의 이익을 부당하게 침해하는지 여부
4. 가명처리 또는 암호화 등 안전성 확보에 필요한 조치를 하였는지 여부

▷ 보호법 개정 및 시행('20.8.5.)으로,「개인정보 비식별조치 가이드라인」('16)은 더 이상 현행법에 따른 가이드라인이 아니므로 활용하지 않음

3 용어 정리

구분	용어설명
개인정보	살아있는 개인에 관한 정보로서 다음의 정보를 포함함 - 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보 - 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 함 - 가명처리를 거쳐 생성된 정보로서 그 자체로는 특정 개인을 알아볼 수 없도록 처리한 정보(이하 '가명정보'라 함) ※ 개인정보에 대한 판단기준은 개인정보처리자가 보유한 정보 또는 접근 가능한 권한 등 개인정보 처리 상황에 따라 다르게 판단되어야 함
가명처리	개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보(이하 '추가정보'라 함)가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것
개인정보파일	개인정보를 쉽게 검색할 수 있도록 일정한 규칙에 따라 체계적으로 배열하거나 구성한 개인정보의 집합물
개인정보처리자	업무를 목적으로 개인정보파일을 운용하기 위하여 스스로 또는 다른 사람을 통하여 개인정보를 처리하는 공공기관, 법인, 단체 및 개인 등
익명정보	시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보
추가정보	개인정보의 전부 또는 일부를 대체하는 가명처리 과정에서 생성 또는 사용된 정보로서 특정 개인을 알아보기 위하여 사용·결합될 수 있는 정보(알고리즘, 매핑테이블 정보, 가명처리에 사용된 개인정보 등) ※ 가명처리 과정에서 생성·사용된 정보에 한정된다는 점에서 다른 정보와 구분됨
재식별	특정 개인을 알아볼 수 없도록 처리한 가명정보에서 특정 개인을 알아보는 것
가명정보 처리시스템	개인정보를 가명처리하거나 가명정보를 처리할 수 있도록 체계적으로 구성한 시스템
결합키	결합 대상 가명정보의 일부로서 해당 정보만으로는 특정 개인을 알아볼 수 없으나 다른 결합대상정보와 구별할 수 있도록 조치한 정보로서, 서로 다른 가명정보를 결합할 때 매개체로 이용되는 값
결합키연계정보	결합키가 동일한 정보에 관한 가명정보를 결합할 수 있도록 서로 다른 결합신청자의 결합키를 연계한 정보
결합대상정보	결합신청자가 결합을 위해 결합전문기관에 제공하는 가명정보에서 결합키를 제외한 정보
결합정보	결합전문기관을 통해 결합대상정보를 결합하여 생성된 정보
반출정보	결합전문기관에서 결합된 결합정보 중 결합전문기관의 심사를 통해 반출된 정보

구분	용어설명
반복결합	시계열 분석 등을 위해 가명정보 결합을 반복하는 결합
반복결합 연결정보	반복결합을 통해 반출된 반출정보의 반복적인 분석을 위해 필요한 정보로, 반출시 해당 정보를 포함하여 반출
결합신청자	가명정보의 결합을 신청하는 개인정보처리자 등 * 가명정보를 제공하거나 이용하는 자(공공기관, 법인, 단체, 개인 등)
결합전문기관	보호법 제28조의3 제1항에 따라 서로 다른 개인정보처리자 간의 가명정보 결합을 수행하기 위해 개인정보위 또는 관계 중앙행정기관의 장이 지정하는 전문기관
결합키관리기관	보호법 시행령 제29조의3 제2항에 따라 특정 개인을 알아볼 수 없도록 가명정보의 안전한 결합을 지원(결합키연계정보를 생성하여 결합전문기관에 제공하는 등) 하는 업무를 하는 한국인터넷진흥원 또는 개인정보위가 지정하여 고시하는 기관
적정성 검토	개인정보처리자가 개인정보를 가명처리한 뒤, 적정성 평가 위원회 등을 구성하여 처리 목적의 적합성, 위험성 검토 결과의 적정성, 가명처리 결과의 적정성, 목적 달성 가능성 등을 검토하여 적절히 가명처리가 되었는지 확인하는 절차
반출 심사	결합된 가명정보에 대해 결합 목적, 반출 정보의 관련성, 특정 개인식별가능성, 반출정보에 대한 안전조치 계획 등을 검토하여 가명정보를 활용하고자 하는 자에게 반출하여도 문제가 없는지에 대하여 심사하는 절차
비정형데이터	영상, 이미지, 음성, 텍스트 등 일정한 규격이나 정해진 형태가 없이 구조화되지 않은 데이터

가명정보 관련 제도 현황에 대한 참고

연번	구분	내용	소관부처
1	법률	개인정보 보호법	개인정보위
2	법률	신용정보의 이용 및 보호에 관한 법률	금융위
3	시행령	개인정보 보호법 시행령	개인정보위
4	시행령	신용정보의 이용 및 보호에 관한 법률 시행령	금융위
5	고시	가명정보의 결합 및 반출 등에 관한 고시	개인정보위
6	고시	신용정보업감독규정	금융위
7	가이드라인	가명정보 처리 가이드라인	개인정보위
8	가이드라인	보건의료 데이터 활용 가이드라인	보건복지부
9	가이드라인	교육분야 가명·익명정보 처리 가이드라인	교육부
10	가이드라인	공공분야 가명정보 제공 실무안내서	행정안전부
11	가이드라인	금융분야 가명·익명처리 안내서	금융위

※ 법제처 국가법령정보센터(www.law.go.kr) 및 각 정부부처 홈페이지를 통해 확인

제 2 장

가명처리 및 가명정보의 처리

1 개요

개인정보

살아있는 개인에 관한 정보로 성명, 주민등록번호, 영상 등 개인을 알아볼 수 있는 정보

성명	홍길동
나이	32세
전화번호	010-1234-5678
주소	서울 종로구 한글길 12

가명정보

개인정보의 일부 또는 전부를 삭제·대체하는 등 가명처리를 통해 추가정보 없이는 특정 개인을 알아볼 수 없는 정보

성명	홍○○
나이	30대 초반
전화번호	010-★★★★-★★★★
주소	서울특별시

익명정보

시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보

성명	(삭제)
나이	30대
전화번호	(삭제)
주소	대한민국

☑ 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 이용, 제공, 결합 등 처리 할 수 있음(보호법 제28조의2 제1항, 제28조의3 제1항)

※ (주의) 「가명정보 처리에 관한 특례」에 따라 정보주체의 동의 없이 처리가 가능한 가명정보는 통계작성, 과학적 연구, 공익적 기록보존 등 목적에 한정되므로 처리 목적이 설정되지 않은 상황에서 보유하고 있는 개인정보를 가명처리하여 보관하는 것은 「가명정보 처리에 관한 특례」에 근거한 처리로 볼 수 없음

※ 불특정 제3자에게 제공하는 경우(공개 등) 익명정보로 처리하는 것을 원칙으로 함

제28조의2(가명정보의 처리 등) ① 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있다.

제28조의3(가명정보의 결합 제한) ① 제28조의2에도 불구하고 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 서로 다른 개인정보처리자 간의 가명정보의 결합은 보호위원회 또는 관계 중앙행정기관의 장이 지정하는 전문기관이 수행한다.

가명처리와 가명정보의 처리 차이점

"가명처리"는 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 과정,
"가명정보 처리"는 가명처리를 통해 생성된 가명정보를 이용·제공 등 활용하는 행위를 말함

☑ 가명정보는 가명처리를 수행한 당시의 목적과 처리 환경(활용 형태, 처리 장소, 처리 방법)에 따라 이용하는 것이 원칙임

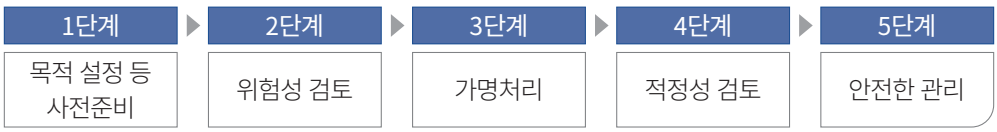
- 다만 보호법 제28조의2 제1항 및 제28조의3 제1항의 목적으로 사용하는 경우 가명정보를 당초 처리 목적과 다른 목적으로 이용하거나 제3자로부터 제공받은 가명정보를 다른 제3자에게 재제공하는 등을 금지하고 있지 않음

※ (예외) 제공 계약 시 재제공 제한이 있거나 반출 시 이용 범위의 제한이 있는 경우에는 가명정보의 재제공 또는 목적 외 이용이 불가할 수 있음

※ (주의) 가명정보를 다른 목적으로 사용하는 경우에는 목적 달성을 위해 꼭 필요한 항목만으로 구성되어야 하며 처리 환경이 달라지는 경우 추가 가명처리 과정을 거쳐야 함

☑ 개인정보의 가명처리는 ① 가명처리 목적 설정 등 사전준비, ② 위험성 검토 ③ 가명처리 수행, ④ 적정성 검토 및 추가 가명처리, ⑤ 가명정보의 안전한 관리 단계로 이루어짐

개인정보의 가명처리 단계별 절차도



기타 참고사항

▶ 개인정보처리자는 안전한 가명정보 처리를 위해 다음의 사항을 참고하여 업무에 반영할 수 있음

- 가명처리 관련 업무의 총괄·관리 및 의사결정을 위한 총괄부서(또는 담당자)를 지정할 수 있으며, 주요 업무는 다음과 같음

1) 가명정보 처리 목적의 적합성 검토
2) 가명처리
3) 가명처리 적정성 검토
4) 가명정보를 처리하는 자에 대한 관리·감독
5) 가명정보에 대한 안전성 확보조치 수행
6) 그 밖에 안전하고 효율적인 가명정보 처리를 위해 필요한 사항

※ 1), 3)의 경우 외부전문가를 포함한 심의위원회를 구성·운영할 수 있음

▶ 가명처리 관련 업무 담당자의 분리

- 가명정보의 재식별 위험을 방지하기 위해서는 가명처리, 적정성 검토, 가명처리가 완료된 가명정보 처리를 수행하는 업무담당자를 각각 분리하고, 해당 업무별로 접근권한을 분리하여 운영하는 것이 안전함

※ 추가정보의 내용을 알고 있는 자가 가명처리의 적정성 검토를 수행하거나 가명정보를 처리(활용)하는 경우 특정 개인을 알아볼 우려가 있음

2 1단계 목적 설정 등 사전 준비

– 가명정보 처리 목적을 명확히 설정하고 가명정보 처리 목적의 적합성 검토 및 계약서, 개인정보 처리방침(80p), 내부 관리계획(127p) 등 필요한 서류를 작성

☑ 목적 설정 : 개인정보처리자는 보호법에서 정한 목적 중에서 가명정보 처리 목적을 선정하고 명확히 설정하여야 함

“통계작성”을 위한 가명정보 처리

- “통계”란 특정 집단이나 대상 등에 관한 수량적인 정보를 의미함
- “통계작성을 위한 가명정보 처리”란 통계를 작성하기 위해 가명정보를 이용, 분석, 제공하는 등 가명정보를 처리하는 것을 말함
- 가명정보의 처리 목적이 시장조사를 위한 통계 등 상업적 성격을 가진 통계를 작성하기 위한 경우에도 가명정보를 처리하는 것이 가능함

예시

- 지방자치단체가 연령에 따른 편의시설 확대를 위해 편의시설(문화센터, 도서관, 체육시설 등)의 이용 통계(위치, 방문자수, 체류시간, 연령, 성별 등)를 작성하고자 하려는 경우
- 인터넷으로 상품을 판매하는 쇼핑몰 등에서 주간, 월간 단위로 판매상품의 재고를 관리하기 위해 판매상품에 대한 지역별 통계(품번, 품명, 재고, 판매수량, 금액)를 작성하고자 하려는 경우
- A공사가 도로구조 개선 및 휴게공간 추가설치 등 고객서비스 개선을 위하여 월별 시간대별 차량 평균속도, 상습 정체구간, 사고구간 및 원인 등에 대한 통계를 작성하고자 하려는 경우

“과학적 연구”를 위한 가명정보 처리

- “과학적 연구”란 과학적 방법을 적용하는 연구*로서 자연과학, 사회과학 등 다양한 분야에서 이루어질 수 있고, 기초연구, 응용연구뿐만 아니라 새로운 기술·제품·서비스 개발 및 실증을 위한 산업적 연구도 해당함

  * 과학적 방법을 적용하는 연구란 체계적이고 객관적인 방법으로 검증 가능한 질문에 대해 연구하는 것을 말함

- “과학적 연구를 위한 가명정보의 처리”란 과학적 연구를 위해 가명정보를 이용, 분석, 제공하는 등 가명정보를 처리하는 것을 말함

- 또한 과학적 연구와 관련하여 공적 자금으로 수행하는 연구뿐만 아니라 민간으로부터 투자를 받아 수행하는 연구에서도 가명정보 처리가 가능함

예시

- 코로나19 위험 경고를 위해 생활패턴과 코로나19 감염률의 상관성에 대한 가설을 세우고, 건강관리용 모바일앱을 통해 수집한 생활습관, 위치정보, 감염증상, 성별, 나이, 감염원 등을 가명처리하고 감염자의 데이터와 비교·분석하여 가설을 검증하려는 경우
- A지자체에서 특정 관광지의 활성화를 위해 국내의 유사 관광지 주변의 상권과 유동인구 분석을 통한 관광지 주변 상권에 대한 지원 및 전환 대책 수립을 위한 연구를 수행하려는 경우
- 공공기관이 보유하고 스팸정보와 민간 통신사에서 자체적으로 보유하고 있는 스팸정보를 가명정보 결합하여 보다 더 많은 스팸정보를 차단할 수 있다는 가설을 세우고, 스팸정보에 해당하는 전화번호, 유형, 날짜, 내용, 신고건수 등의 정보를 가명처리 및 결합을 통해 가설을 검증하고 결합에 참여한 스팸방지 시스템을 고도화하려는 경우

“공익적 기록보존”을 위한 가명정보 처리

- “공익적 기록보존”이란 공공의 이익을 위하여 지속적으로 열람할 가치가 있는 정보를 기록하여 보존하는 것을 의미함
- “공익적 기록보존을 위한 가명정보 처리”란 공익적 기록보존을 위해 가명정보를 이용, 분석, 제공하는 등 가명정보를 처리하는 것을 말함
- 공익적 기록보존은 공공기관이 처리하는 경우에만 공익적 목적이 인정되는 것은 아니며, 기업, 단체 등이 일반적인 공익을 위하여 기록을 보존하는 경우에도 공익적 기록보존 목적이 인정됨

예시

- 연구소가 현대사 연구 과정에서 수집한 정보 중 사료가치가 있는 생존 인물에 관한 정보를 가명처리하여 기록·보존하고자 하려는 경우
- 연구소가 코로나19 연구 과정에서 수집한 정보 중 공익적 연구가치가 있는 환자에 관한 정보를 가명처리하여 기록보존하고자 하려는 경우

☑ 가명처리 대상 선정(결합대상 속성정보 선정): 처리목적 달성에 필요한 정보의 종류, 범위를 명확히 하여 가명처리 대상을 선정함

※ 개인정보처리자는 정보주체가 자신의 개인정보에 대한 가명처리 정지를 요구하거나 개인정보 처리에 대한 동의를 철회한 경우, 가명처리 대상 정보에서 해당 정보주체의 정보를 제외하고 선정해야 함(보호법 제37조)

제37조(개인정보의 처리정지 등) ① 정보주체는 개인정보처리자에 대하여 자신의 개인정보 처리의 정지를 요구하거나 개인정보 처리에 대한 동의를 철회할 수 있다. 이 경우 공공기관에 대하여는 제32조에 따라 등록 대상이 되는 개인정보파일 중 자신의 개인정보에 대한 처리의 정지를 요구하거나 개인정보 처리에 대한 동의를 철회할 수 있다.

☑ 처리 목적 적합성 검토(개인정보 보유부서 또는 가명정보 활용 관련 전담부서 등) : 개인정보의 수집 목적 및 성격, 가명정보 활용 목적, 이용 목적에 대한 법률적 근거 등을 고려하여 가명처리 여부를 결정함

※ 필요시 적합성 검토위원회 심사 또는 외부전문가 평가 등을 통해 결정할 수 있음

☑ 가명정보 처리를 위한 안전조치 이행 : 개인정보 처리방침 수립·공개(보호법 제30조), 내부 관리계획 수립·시행(개인정보의 안전성 확보조치 기준 제4조) 등 가명정보 처리에 앞서 이행하여야 할 사항을 준비해야 함

- 가명정보 처리에 관한 내부 관리계획이 없는 경우 수립이 필요함 ([제4장 안전성 확보 조치] (77p) 참고)

☑ 필요서류 작성 : 가명정보의 처리 또는 가명처리를 위탁(보호법 제26조에 따라 수행)하거나 가명정보를 제3자에게 제공하는 경우 필요에 따라 재식별 금지에 관한 사항, 기타 처리에 있어 유의해야 할 사항* 등을 포함한 계약서를 작성할 수 있음

* (예시) 가명정보의 재제공 금지, 가명정보 재식별 금지, 가명정보의 안전성 확보조치, 가명정보의 처리기록 작성 및 보관, 가명정보의 파기, 재식별 시 책임 및 손해배상 등

- 또한, 가명정보 처리 또는 가명처리 위탁 시 위탁 관련 문서 작성, 위탁 업무 공개, 수탁사에 대한 관리·감독에 관한 사항 등 위탁 처리 시 준수하여야 할 사항들을 확인하여야 함

☑ 기타 : 개인정보 활용 및 가명처리 등에 대해 내부 승인 절차를 별도로 두고 있는 개인정보처리자는 이 단계에서 해당 절차를 진행하여야 함

- (적절하지 않은 예시) 신제품 개발을 위한 과학적 연구 수행
  ※ 목적이 구체적으로 명시되지 않아 적절하지 않음
- (적절한 예시) ○○제품의 성능 개선을 위해 개인별 ○○○특성에 대한 설문조사를 토대로 개인별 특성과 성능 요인의 연관성에 대한 과학적 연구 수행

기타 참고사항

▶ 민감정보와 고유식별정보의 처리

- 민감정보(보호법 제23조) 또는 고유식별정보(보호법 제24조)도 가명정보 처리 특례에 따라 가명처리하여 활용하는 것이 가능하지만, 개인정보 보호 원칙(보호법 제3조)을 준수하여 처리 목적에 필요하지 않은 민감정보 또는 고유식별정보는 삭제하여야 함
- 다만 주민등록번호는 법령에 주민등록번호를 처리할 수 있는 근거가 없는 경우 가명정보 처리 특례에 따른 가명처리는 허용되지 않음(보호법 제24조의2)

※ 가명정보 처리의 목적이 적합한지에 대한 입증 책임은 개인정보처리자에게 있으므로 개인정보처리자는 향후 처리 목적에 대한 증빙을 위해 연구계획서 등 목적설명서를 작성할 수 있음 ([참고자료] 참고5. 가명처리 및 결합 목적 증빙 자료 예시 (119p) 참고)

3 2단계 처리 대상의 위험성 검토

☑ 대상 선정 : [1단계. 사전 준비]에서 설정한 목적을 달성하기 위해 필요한 항목을 개인정보파일에서 선정함

※ 가명처리 대상 항목 선정 시 가명정보 처리 목적 달성에 필요한 최소 항목으로 해야 함

예시 가명처리 대상 항목 선정

▶ 가명처리 목적 : 성별과 지역에 따른 구매액 상관관계를 분석하고자 함

▶ 개인정보파일 내 항목 : 성명, 휴대폰번호, 성별, 이메일, 주소, 구매상품, 구매액, 장바구니 목록

▶ 가명처리 대상 항목(속성정보) : 성별, 주소(시군구), 구매액

* 분석 목적과 상관없는 속성정보는 제외하고 대상 선정

- 가명처리 시에는 가명정보 그 자체만으로 특정 개인을 알아볼 수 있는지와 가명정보를 처리할 자가 보유하거나 접근·입수가능한 정보*와의 사용·결합을 통해 식별할 수 있는지를 고려해야 함

* 다른 정보와의 사용·결합을 통해 개인을 식별할 수 있게 되는 경우 보호법 제2조제1호나목에 따른 개인정보에 해당할 수 있음

제2조(정의) 1. "개인정보"란 살아있는 개인에 관한 정보로서 다음 각 목의 어느 하나에 해당하는 정보를 말한다.

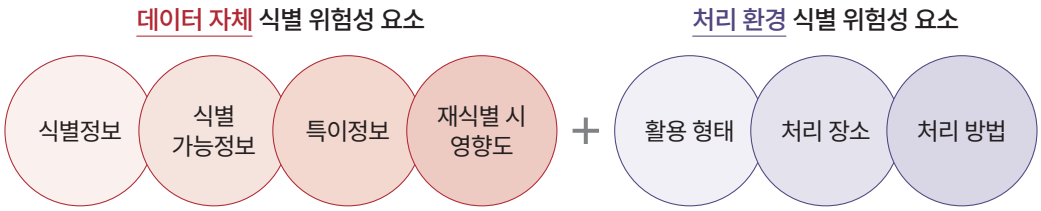
나. 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보. 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다.

예시 전화번호, 지명, 소속, 대화 상대방과의 관계 등을 추론할 수 있는 대량의 대화 문장 같은 경우 다른 정보와의 사용·결합을 통해 개인을 알아볼 가능성이 있음

- 가명처리 수준은 가명정보 처리 상황에 따라 달라지므로 당초 가명정보를 다른 목적으로 처리하거나 재제공하는 등 활용 형태, 처리 장소, 처리 방법 등 처리 상황에 변화가 있는 경우 해당 상황을 고려한 추가적인 가명처리가 필요함

☑ 위험성 검토 : 위험성 검토는 가명처리 대상 데이터의 식별 위험성을 분석·평가하여 가명처리 방법 및 수준에 반영하기 위한 절차이며,

- 식별 위험성은 **1) 데이터의 식별 위험성**과 **2) 처리 환경의 식별 위험성**으로 구분하여 검토해야 함



데이터 자체 식별 위험성 요소	식별정보	• 특정 개인과 직접적으로 연결되는 정보 • (예시) 성명, 고유식별정보, (개인)휴대전화번호, (개인)전자우편주소, 의료기록번호, 건강보험번호 등
	식별가능정보	• 다른 항목과 결합하는 경우 식별가능성이 높아지는 항목 • (예시) 성별, 연령, 거주 지역, 국적, 직업, 위치정보 등 • 해당정보를 처리하는 자를 기준으로 판단
	특이정보	• 전체데이터에 식별가능성을 가지는 고유(희소)값, 편중된 분포를 가지는 단일·다중이용항목 • (예시) 희귀성씨 등 특이한 값, 국내 최고령 등 극단값, 특정 데이터 분석집단에서 희소한 값 등
	재식별시 영향도	• 특정 정보주체에게 사회적 파장 등 영향도가 높은 항목 • 사회통념상 차별받을 수 있는 정보 또는 재식별로 인한 불이익이 큰 정보주체(대중적으로 유명한 사람 등)
처리환경 식별 위험성 요소	활용 형태	• 내부 이용 또는 외부 제공하는 경우 처리자(또는 취급자)가 보유하거나 접근·입수가능한 정보, 이용 범위 및 유형 등 • 보안서약서, 계약서 등을 통해 파악 가능한 범위 정보를 고려하여 식별 가능성 검토 가능
	처리 장소	• 해당 가명정보 외에 다른 정보의 접근·입수가 제한된 장소에서 처리되는지 여부 • 보안서약서, 계약서 등으로 내·외부 활용이 제한된 경우 폐쇄 환경에 준하여 검토가능
	처리 방법	• 가명정보를 다른 정보와 연계·분석·내부 결합하는 경우 결합 후 식별가능한 항목 • 가명정보 반복 제공시 식별 위험이 높아지는 항목

1) 데이터의 식별 위험성 검토

☑ 데이터 자체의 위험성 검토는 가명처리 대상이 되는 정보에 식별 가능한 요소가 있는지를 파악하는 것으로, ① 그 자체로 식별될 위험이 있는 항목, ② 다른 항목과 결합을 통해 식별될 가능성이 있는 항목, ③ 특이정보, ④ 그 밖에 데이터 특성만으로 재식별 시 사회적 파장 등 영향도가 높은 항목 등이 있는지 검토해야 함

- (식별정보) 다른 사람과 구분하기 위해 부여된 식별 정보는 특정 개인과 직접적으로 연결되는 정보로, 해당 정보가 포함되어 있는지 검토
- (식별가능정보) 단일 항목으로는 식별 가능성이 없으나, 가명처리 대상이 되는 다른 항목과 결합하는 경우 식별 가능성이 높아지는 항목이 있는지 검토

예시 개인 식별 가능성이 높은 정보

▶ **식별정보 :** 성명, 고유식별정보(주민등록번호, 여권번호, 외국인등록번호, 운전면허번호), (개인)휴대전화번호, (개인)전자우편주소, 의료기록번호, 건강보험번호 등 식별을 목적으로 생성된 정보

▶ **식별가능정보 :** 성별, 연령(나이), 거주 지역, 국적, 직업, 위치정보 등 개인정보처리자의 입장에서 개인을 알아볼 수 있는* 정보

* 개인을 '알아볼 수 있는지'는 해당 정보를 처리하는 자(정보의 제공 관계에 있어서는 제공받는 자를 포함)를 기준으로 판단하여야 함

- (특이정보 유무) 가명처리 대상 전체 데이터에 식별 가능성을 가지는 고유(희소)한 값이 있는지, 편중된 분포를 가지는 단일·다중 이용 항목이 있는지 검토

※ 가명처리 대상 정보의 항목별 분포와 특이정보의 포함 여부 등을 말하는 것으로 분포가 편중되어 있거나 특이정보가 다수 포함되어 있는 경우 식별 가능성이 높음

예시 특이정보

▶ 희귀 성씨, 희귀 혈액형, 희귀 눈동자 색깔, 희귀 병명, 희귀 직업 등 정보 자체로 특이한 값을 가지는 정보

▶ 국내 최고령, 최장신, 고액체납금액, 고액급여수급자 등 전체적인 패턴에서 벗어나는 극단값을 발생할 수 있는 정보

▶ 도서·산간 지역주민의 영유아에 대한 정보 등 특정 데이터 분석 집단에서 희소한 값을 가지는 정보

- (재식별시 영향도) 데이터가 지니는 특성만으로 재식별 시 특정 정보주체에게 사회적 파장 등 영향도가 높은 항목이 있는지 검토

※ 사회통념상 차별 정보 등으로 정보주체가 피해 또는 불이익을 받을 수 있는 정보 등

2) 처리 환경의 식별 위험성 검토

☑ 개인정보처리자는 가명정보 활용 형태(이용·제공), 처리 장소, 처리 방법(결합여부) 등 가명정보 처리 상황에 따라 발생할 수 있는 식별 위험성이 있는지 검토해야 함

- (활용 형태) 가명정보를 처리하는 처리자(또는 취급자)가 보유하고 있는 정보 또는 접근·입수 가능한 정보, 이용 범위 및 유형 등을 고려하여 식별가능한 항목이 있는지 검토

※ 처리자(또는 취급자)가 보유, 접근, 입수 가능한 모든 정보를 고려하여 식별가능성을 검토할 필요는 없으며, 보안서약서, 계약서 등을 통해 파악이 가능한 범위의 정보를 고려하여 식별 위험성을 검토하는 것이 가능함

예시 처리 주체가 보유하고 있는 정보

구분	처리자 (취급자)	가명처리전 정보	추가정보	다른 정보[1]	보유 경험 및 지식[2]
이용	동일부서	○	○	○	○
	타 부서	○	○	○	○
제공	제3자			○	○

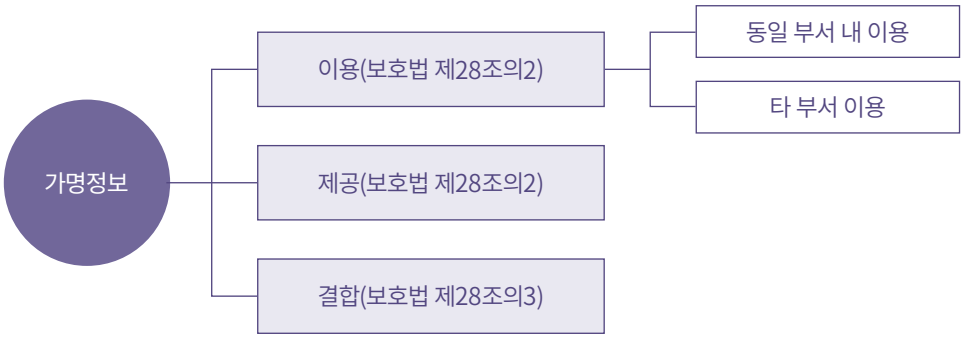
주1) 가명처리 전 정보와 추가정보를 제외한 개인정보처리자가 보유하고 있는 정보를 말함
주2) 내·외부 이용기관이 보유하고 있는 과거 유사 정보에 대한 수행 경험이나 지식 등을 말함

- (처리 장소) 가명정보가 해당 가명정보 외에 다른 정보의 접근·입수가 제한된 장소에서 처리되는지 검토

※ 다만 보안서약서, 계약서 등으로 내·외부 정보의 활용이 제한된 경우 폐쇄 환경에 준하여 검토 가능함

- (처리 방법)
  - 가명정보를 다른 정보와 연계 분석하는 경우 다른 정보와 결합 후 식별가능한 항목이 있는지 검토
  - 가명정보를 다른 정보와 내부 결합하는 경우 다른 정보와 결합 후 식별가능한 항목이 있는지 검토
  - 가명정보를 반복 제공하는 경우 반복 제공을 통해 식별 위험이 높아지는 항목이 있는지 검토

☑ 가명정보 이용 및 제공시 유의 사항



가. 동일 개인정보처리자 내 이용

□ 개인정보처리자가 보유한 개인정보*를 가명처리 또는 내부 결합하여 직접 활용 또는 다른 부서에 제공하는 경우를 의미함

* 정보주체로부터 직접 수집하거나 합법적으로 다른 개인정보처리자로부터 수집·제공받은 개인정보

□ 가명정보를 처리하는 소속 부서에서 이미 보유하고 있는(접근 가능한) 정보, 처리 시점을 기준으로 제공받는 다른 정보를 고려하여 식별 위험성을 검토함

잘못된 내부이용(동일 부서 내 이용) 사례1

동일 부서 내 이용으로 ◯◯화장품 회사의 A팀은 화장품 판매정보를 관리하는 팀으로서, 가명정보 또는 추가정보에 접근할 수 있는 권한을 분리하지 않고 해당 정보를 가명처리하여 신상품 수요조사 예측 모델 개발을 목적으로 활용

▶ (처리현황) A팀은 판매정보 내 개인식별 가능성이 있는 이름, 성별, 승인번호를 가명처리하고, 희귀 지역의 판매내역을 삭제하여 A팀 가명정보 분석담당자에게 제공

✓ 가명정보 분석담당자는 A팀의 판매정보 관리 업무를 병행하여 업무를 수행하고 있음

▶ (문제점) 가명정보 분석담당자는 가명정보 분석을 통해 최고가 화장품의 금액과 판매지역을 파악할 수 있으며, 판매정보가 관리되고 있는 개인정보처리시스템에 접근이 가능하여 금액과 지역을 통해 특정 개인을 식별할 가능성이 있음

▶ **해결방안** 가명정보 분석담당자가 가명정보 분석을 수행하는 경우를 제외하고는 특정 개인을 알아볼 수 있는 개인정보처리시스템에 접근할 수 없도록 제한해야 함

잘못된 내부이용(동일 부서 내 이용) 사례2

동일 부서 내 이용으로 OO유통사의 A팀은 매장의 판매정보시스템 고도화를 위해 매장고객(고객번호, 연령, 주소) 정보와 판매정보(제품번호, 제품명, 제품금액, 제품 제고 및 판매량)를 가명처리하여 내부적으로 분석하고자 함

▶ (처리현황) A팀은 가명정보 컨설팅 업체를 통해 가명처리에 관한 관리적·기술적 보호조치를 모두 준수하고 개인정보 가명처리 전용 솔루션을 통해 정형데이터에 대한 가명처리를 수행

✓ 가명처리된 정보는 판매정보시스템 고도화를 위해 활용되었으며, A팀은 보다 나은 서비스 개선을 위해 처리된 가명정보를 활용하여 신상품 개발을 위한 경진대회를 개최하였음

▶ (문제점) A팀은 개인정보 보호법에서 규정하고 있는 가명정보에 대한 관리적/기술적 보호조치를 모두 준수하였으며, 과학적 연구 목적 내에서 판매정보시스템을 고도화하였지만 신상품 개발을 위한 경진대회 개최의 목적은 개인정보 보호법 제28조의2제1항에서 규정한 목적에 해당하지 않음

▶ **해결방안** A팀은 원래 처리 목적 외로 가명정보를 활용하고자 하는 경우 개인정보 보호법 제28조의2제1항의 목적 범위 내에서 가명정보를 처리하여야 하며, 새로운 처리 환경에 맞추어 추가 가명처리를 수행하여야 함

잘못된 내부이용(타 부서 이용) 사례1

타 부서 이용으로 □□공사는 A부서의 고속도로 이용차량 빅데이터 분석 결과를 고속도로 통행요금을 관리하는 B부서에 교통서비스 개선을 위한 연구 목적으로 제공(이 때 B부서에서 처리하는 개인정보를 고려하지 않음)

▶ (처리현황) A부서는 개인식별 가능성이 있는 차량번호, 차종 등을 가명처리하고, 톨게이트 입출시간, 이동량, 사고정보 등의 정보를 B부서에 제공

✓B부서는 고속도로 통행요금 관리를 위해 고객번호와 차량번호, 톨게이트 입출시간 및 결제금액 정보를 보유하고 있음

▶ (문제점) B부서는 A부서에서 제공받은 정보의 이동시간 정보와 B부서가 보유한 톨게이트 입출시간을 활용하여 특정시간에 통과한 차량의 번호를 알 수 있으며, 해당 차량번호를 통해 특정 개인을 식별할 가능성이 있음

▶ **해결방안** A부서에서는 B부서가 보유하고 있는 정보를 고려하여 특정 시간에 대한 식별가능성이 없도록 이동시간 삭제 또는 가명처리 등을 수행하여야 함(필요시 가명처리를 위해 B부서가 보유한 톨게이트 입출시간 정보 제공 요청)

잘못된 내부이용(타 부서 이용) 사례2

타부서 이용으로 ○○사는 A부서의 고객 AS 및 민원처리 내역(비정형데이터)을 시스템 개발을 담당하는 B부서에 문의 유형별 민원처리 방안 연구를 위한 목적으로 제공(이때 A부서는 가명처리 후 적정성 검토를 수행하지 않음)

▶ (처리현황) A부서는 보유하고 있는 고객 AS 및 민원처리 내역(질문자, 질문분류, 질문내용, 답변내용, 처리만족도 등 비정형데이터)을 가명처리하여 B부서에 제공하였음

✓민원처리 내역 등의 비정형데이터는 정규 표현식 및 개인정보 검출 시스템을 통해 가명처리를 수행하였으며, B부서는 제공받은 정보를 그대로 알고리즘 고도화 문의 유형별 민원처리 시스템 개발에 이용

▶ (문제점) B부서는 제공받은 비정형데이터를 알고리즘 고도화 문의 유형별 민원처리 시스템 개발에 활용하려는 시점에 A부서로부터 제공받은 정보(민원처리 내역)의 가명처리가 미흡하였을 경우 개인을 식별할 수 있는 정보가 노출되어 특정 개인을 식별할 가능성이 있음

▶ **해결방안** 비정형데이터에 대한 가명처리는 현재 기술상 미흡한 부분이 많으므로 적정성 검토 시 신중하여야 하며, 또한 식별 위험이 높으므로 처리 시 폐쇄망 환경에서 이용하는 것을 권장하고, 식별이 된 경우 해당 정보의 처리를 즉시 중지하고 회수·파기하여야 하며 추가 가명처리 필요

✓현 시점에서 명확한 처리 방안이 없는 비정형데이터의 경우 정보주체의 동의를 받고 활용하는 것을 권고 함

나. 다른 개인정보처리자 제공(제3자 제공)

□ 개인정보처리자가 보유한 개인정보를 가명처리하여 특정 제3자에게 제공하는 경우를 의미함

- 제3자의 개인정보 보호수준 및 신뢰도를 고려하여 제공하는 가명정보로 발생할 수 있는 재식별 위험을 최소화하기 위하여 노력하여야 함*

* 보호수준이 낮은 기관에는 상대적으로 높은 수준의 가명처리 수준을 적용하는 방법 등

- 가명정보를 제3자에게 제공하는 경우 추가정보 등 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 제공하여서는 아니됨(보호법 제28조의2 제2항)

> **제28조의2(가명정보의 처리 등)** ② 개인정보처리자는 제1항에 따라 가명정보를 제3자에게 제공하는 경우에는 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함해서는 아니 된다.

- 또한 개인정보처리자는 제3자가 사전에 보유하고 있는(접근 가능한) 정보, 처리 시점을 기준으로 제공받는 다른(개인)정보 등을 고려하여야 하고, 이를 파악하기 위해 관련 정보*를 요청하는 것도 가능함

* 제3자가 관리하고 있는 개인정보 중 제공받는 가명정보와 연계 또는 조합 가능성이 있는 개인정보 목록 등

- 사전 준비 단계의 계약서에 데이터의 이용환경에 대한 제한 등에 대하여 명시한 사항*이 있다면 이를 고려할 수 있음

* 다른 정보의 접근이 제한된 폐쇄망 환경에서 이용하겠다는 사항 등

(참고) 가명정보 제공시 법적책임 범위

▶ 개인정보를 보호법에서 정한 처리 목적에 따라 가명처리하고 관련 안전조치 등 법률에서 정한 사항을 모두 준수하여 가명정보를 제공한 경우,

- 가명정보를 제공받은 자가 가명정보 이용 과정에서 의도치 않게 특정 개인을 알아볼 수 있는 정보가 생성되었다는 사실만으로는 가명정보를 제공한 자에 대해 개인정보 보호법상 행정처분을 하지 아니함

※ 단, 제공받은 자는 위 생성된 정보의 처리를 즉시 중지하고, 지체없이 회수·파기하여야 함

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

잘못된 외부제공 사례1

○○호텔에서는 최고급 객실을 이용한 VIP등의 특이정보를 삭제하지 않고 호텔 투숙 및 서비스 금액 등을 △△분석 회사에 제공하고, △△분석회사는 해당 정보를 분석하여 시간에 따른 객실이용현황 및 서비스이용에 대한 조사 연구를 수행

▶ (처리현황) △△분석회사는 온라인 SNS정보 및 다양한 기업의 정보를 수집하여 다양한 연구조사를 실시하는 회사로서 내부 관리계획을 수립하고, 관리적·기술적 보호조치를 준수하고 있음

✓ 호텔은 회원번호와 이름을 가명처리하고, 나이, 성별, 등급, 예약방법, 객실정보, 체크인, 체크아웃, 서비스 이용금액을 제공

▶ (문제점) △△분석회사의 분석담당자는 특정일에 최고급 객실을 이용한 내용을 분석과정에서 인지할 수 있으며, 기존 업무(온라인 SNS정보 수집)를 수행하며 공개된 정보(예 : 개인이 SNS에 올리는 정보, 여행후기 등)를 통해 특정 개인을 식별할 가능성이 있음

▶ **해결방안** ○○호텔은 제공하는 가명정보에 포함된 특이정보(최고급 객실)를 삭제 또는 가명처리 등을 수행하여야 함

잘못된 외부제공(위탁) 사례2

○○기관은 복지 서비스 정책 개선에 필요한 모델 개발 연구를 위해 △△대학에 해당 연구를 위탁함. ○○기관은 연구를 위해 ○○기관이 보유하고 있는 기초수급대상자 정보 및 정부예산 수급, 지원 내역을 가명처리하여 제공하였으나, 가명정보의 처리를 위탁하며 △△대학의 가명정보 처리에 관한 안전성 확보조치 이행 여부를 확인하지 않음

▶ (처리현황) ○○기관은 가명정보의 처리를 위해 사전에 기관 내 가명정보 분석을 위한 전담조직을 구성한 후 내부 관리계획 및 가명정보 처리에 대한 내부 지침 등을 마련하고 가명정보처리시스템에 대한 접근통제 및 권한관리 등 보호 조치를 준수하고 있음

✓ ○○기관은 보유하고 있는 기초수급대상자 정보 및 정부예산 수급, 지원 내역을 가명처리한 후 △△대학에 제공하고 △△대학은 ○○기관으로부터 연구비 지원을 받아 복지 서비스 개선에 관한 연구 용역을 실시

▶ (문제점) ○○기관은 △△대학의 요청으로 보유하고 있는 개인정보를 가명처리하여 △△대학에 제공하였지만, △△대학에 가명정보의 처리를 위탁하며 가명처리에 대한 안전성 확보조치(가명정보 처리에 관한 내부 관리계획 및 추가정보 분리 및 기록보관, 가명처리 시스템에 대한 안전성 확보조치 등)를 확인하지 않음

▶ **해결방안** ○○기관은 △△대학에 가명정보 처리에 대한 위탁 시 사전에 가명정보 처리에 관한 준수사항을 준수할 수 있도록 확인해야 하며(필요시 교육 또는 컨설팅 제공 지원), △△대학도 가명정보 처리에 대한 주기적인 교육 및 준수사항을 가명정보 처리 이전에 이행할 수 있도록 하여야 함

개인식별 위험성 검토 체크리스트

※ 해당 체크리스트는 가명처리 계획을 세우기 전에 개인식별 위험성을 검토하기 위한 것으로, 검토 결과가 "예"에 해당하더라도, 해당 위험을 낮추기 위한 적절한 가명처리 방안 적용 후 활용 가능

구분		개인식별 위험성 검토 사항	
데이터	식별성	개인 식별이 가능한 항목 여부	
		검토 항목	검토 결과
		① 식별이 가능한 단일항목의 정보가 있는가 * [항목설명] 직업에 개인의 식별성이 매우 높은 정보들이 포함되는 경우(장애인 여성 탁구 국가대표 감독, 지방자치단체장, 2급 이상의 공무원 등) 등	□ 예 □ 아니오
		② 두 개 이상의 컬럼(항목)을 조합하여 식별가능성이 높아지는 정보가 있는가 * [항목설명] △ 동일 데이터셋 내 여러 이용 항목을 동일 목적으로 함께 분석함에 따라 식별가능성이 높아지는 경우(질병, 투약, 약품 등 연관있는 이용 항목을 종합적으로 분석하는 경우) 등 △ 데이터셋 내 가족관계, 직책관계 등 계층적 특성을 가진 이용 항목이 포함되어 있어 개인 식별가능성이 높아질 수 있는 경우(회사 내 정보 분석 시 해당 데이터셋에 소수 직책이 포함되어 있는 경우) 등 △ 시간, 위치, 행위 등 이용 항목을 함께 분석하는 경우 등	□ 예 □ 아니오
		③ 공개된 데이터와 결합·대조하여 식별가능성이 높아질 수 있는 이용 항목이 있는가 * [항목설명] 통계청의 인구 센서스 데이터를 사용하여 식별가능한 이용 항목이 있는 경우 등	□ 예 □ 아니오
		④ 데이터셋의 크기가 적어 식별이 가능할 우려가 있는가 * [항목설명] 연구대상 질병이 극희귀질병(국내 유병자가 200명 미만인 질병)이라 이 질병을 가진 대상이 한정된 인원이라 식별 가능성이 높아지는 경우 등	□ 예 □ 아니오
		⑤ 원본데이터 전체가 아닌 일부의 데이터를 처리하는 샘플링을 적용하지 않았는가 * [항목설명] 상품 구매이력 분석에서 특정 고객층의 전체를 분석하는 것이 아니라 특정 고객층의 일부를 샘플링해서 분석하는 경우 등	□ 예 □ 아니오
		⑥ 시계열 성격을 가진 데이터가 포함되어 있는가 * [항목설명] 대학에서 학생들의 학점 데이터를 입학 때부터 졸업 때까지의 모든 학점에 대해 분석하는 경우 등	□ 예 □ 아니오

<table>
<tr><th colspan="2">구분</th><th colspan="2">개인식별 위험성 검토 사항</th></tr>
<tr><td rowspan="6">데이터</td><td rowspan="3">특이정보</td><td colspan="2">데이터 분포가 편중되어 있어 식별가능성이 있는 이용 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑦ 연속적인 숫자형 데이터에서 데이터 값의 분포가 양 끝단의 정보(분포 곡선에 따라 한쪽의 정보 포함)가 현저히 낮은 항목이 있는가</td><td>□ 예
□ 아니오</td></tr>
<tr><td>⑧ 일반적인 문자형 데이터(비 연속적인 숫자형 데이터 및 코드형 데이터 포함)에서 특정 값으로 현저히 낮은 항목이 있는가</td><td>□ 예
□ 아니오</td></tr>
<tr><td rowspan="3">재식별 시 영향도</td><td colspan="2">재식별 시 정보주체에게 심각한 피해 또는 불이익을 초래할 수 있는 이용 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑨ 사회통념상 차별 등으로 인해 정보주체가 피해 또는 불이익을 받을 수 있는 정보가 있는가</td><td>□ 예
□ 아니오</td></tr>
<tr><td rowspan="4">처리환경</td><td rowspan="4">이용 및 제공</td><td colspan="2">가명정보 활용 형태 및 이용 기관의 개인정보 보호 수준 등을 고려하여 식별가능성이 있는 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑪ 처리주체가 보유하고 있는 정보 또는 접근·입수 가능한 정보와 이용 범위 및 유형을 고려하여 식별가능한 항목이 있는가
* [항목설명] △ 시계열 분석 등을 위한 목적으로 가명정보를 반복 제공할 예정인 경우 반복 제공을 통해 식별 위험이 높아지는 항목이 있는 경우 등
△ 가명정보를 취급하는 자와 관련된 정보가 처리하는 데이터셋에 포함되어 있는 경우 등</td><td>□ 예
□ 아니오</td></tr>
<tr><td>⑫ 추가정보를 삭제하지 않고 보관하는가</td><td>□ 예
□ 아니오</td></tr>
<tr><td>⑬ 가명정보 제공 시 제공받는 자의 개인정보 보호 수준 및 신뢰할 수 있는 인증을 받았는가(ISMS, ISMS-P, ISO 27001 등)</td><td>□ 예
□ 아니오</td></tr>
</table>

<table>
<tr><th colspan="2">구분</th><th colspan="2">개인식별 위험성 검토 사항</th></tr>
<tr><td rowspan="5">처리환경</td><td rowspan="3">처리장소</td><td colspan="2">가명정보가 관리적·기술적·물리적으로 안전한 장소에서 처리되는지 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑭ 가명정보 처리 시 다른 정보를 접근·입수할 수 있는 장소인가
* [항목설명] △ 누구나 접근 가능한 개방형 형태의 장소 및 네트워크인지 △ 내부인원만 출입할 수 있는 장소 및 네트워크가 아닌지 △ 가명정보 처리 관련 담당자만 접근할 수 있는 장소 및 네트워크가 아닌지 등</td><td>□ 예
□ 아니오</td></tr>
<tr><td rowspan="3">다른 정보와의 결합</td><td colspan="2">가명정보를 다른 정보와 결합하여 활용 시 식별가능성이 있는 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑮ 다른 정보와의 연계 분석이 예정되어 있는가</td><td>□ 예
□ 아니오</td></tr>
<tr><td colspan="2">⑯ 처리주체가 보유하거나 접근·입수 가능한 정보 등 다른 정보와 연계 또는 결합하여 식별가능한 항목이 있는가</td><td>□ 예
□ 아니오</td></tr>
</table>

개인식별 위험성 검토 항목별 조치 가이드

구분		조치 가이드
데이터	식별성	개인을 직접적으로 알아볼 수 있는 식별정보는 원칙적으로 삭제하여야 하며, 결합 등 이용목적 상 필요한 경우 안전한 방식으로 대체할 수 있는 정보를 생성하여 대체
	특이정보	특이정보는 그 정보만으로 개인을 식별할 수 있는 정보는 아니더라도 고유(희소)한 특성 때문에 개인을 알아볼 수 있는 가능성이 높으므로, 이용목적 상 반드시 필요하지 않다면 삭제하고 필요 시 상하단 코딩, 범주화 등 처리
	재식별 시 영향도	사회통념상 차별, 기본권 침해 등 파급 영향이 클 수 있는 정보는 재식별 시 다른 일반정보와 다르게 개인의 피해와 더불어 사회적 파장이 있을 수 있으므로, 꼭 필요한 항목 이외에는 삭제 등 조치
처리환경	이용 및 제공	이용 및 제공의 위험성이 있는 경우 이용자와 제공자가 서로 위험성을 낮추기 위한 처리 환경에 대한 안전성 입증 관련 협의가 필요
	처리장소	물리적·관리적·기술적으로 처리 장소의 안전성 확보가 되지 않으면 가명처리의 수준을 높이거나 별도의 안전한 처리 장소를 물색
	다른 정보와의 결합	다른 정보와 연계·결합 예정에 있는 경우 연계·결합되는 정보와 결합하여 식별가능성이 높아지는 항목이 있는지 추가 검토 필요 처리주체가 보유하거나 접근·입수 가능한 정보를 통해 식별가능한 항목이 있는지 검토 - 가명정보를 제공받아 활용하게될 자가 가진 과거 유사 정보에 대한 수행 경험이나 지식 등은 가명정보를 제공하려는 자가 자체적으로 판단하기 어렵기 때문에 가명정보를 제공받아 활용하게 될 자에게 사전에 확인 및 검토 필요. 사전 검토가 어려운 경우, 가명처리의 수준을 높이는 방법 등으로 위험성을 낮춰야 함

□ 개인정보처리자는 데이터의 식별 위험성과 처리 환경의 식별 위험성 검토를 통해 가명처리에 대한 식별 위험성 평가 결과를 도출하여야 함

※ 최종 검토의견은 외부전문가에게 자문 및 작성을 요청할 수 있음

식별 위험성 검토 결과보고서 작성 예시(내부이용)

<table>
<tr><td colspan="2">가명정보
활용목적</td><td>▪ 본 기업은 전국적인 소매유통망을 가지고 있는 대형유통업체로 코로나 이전과 코로나 이후의 상품군별 판매 추이에 대한 통계학적 연구 분석을 통해 이후 코로나의 지속가능성이 높아짐에 따라 상품의 구매전략, 제품의 진열 위치 변경 등의 판매전략의 수립을 위해 데이터로 활용하기 위해 2019년 1월부터 12월까지의 주요 상품군별 판매액 정보와 2021년 1월부터 12월까지의 주요 상품군별 판매액 정보를 나이와 성별, 시군구 단위의 주소별로 비교하여 분석</td></tr>
<tr><td colspan="2">가명처리
대상 데이터
항목</td><td>▪ 고객ID, 나이, 주소, 성별, 2019년 1월~12월, 2021년 1월~12월까지의 여행용품, 식품류, 의류, 취미용품, 생활용품, 유아용품, 기타의 7개 범주의 구매금액의 월별 합계액, 월별 구매 총 금액, 월별 선호 제품군, 각 년도의 고객 등급 (전체 222개의 컬럼)
▪ 전체 고객 800만명 중 25%를 무작위 샘플링하여 구성한 200만명에 대한 데이터</td></tr>
<tr><td rowspan="3">데이터 위험성</td><td>식별성 유무</td><td>▪ ‘고객ID’는 개인식별정보임
▪ ‘나이’, ‘주소’, ‘성별’은 조합했을 때 개인의 식별이 가능한 개인식별 가능정보임</td></tr>
<tr><td>특이정보 유무</td><td>▪ 각 범주별 구매금액의 경우 특이정보로 인한 개인 식별성이 발생할 수 있음</td></tr>
<tr><td>재식별시
영향도</td><td>▪ 단순 고객의 구매데이터로 재식별 시 영향도는 크지 않을 것으로 판단됨</td></tr>
<tr><td rowspan="3">처리 환경 검토</td><td>이용 및 제공 형태</td><td>▪ 내부 이용</td></tr>
<tr><td>처리 장소</td><td>▪ 가명정보는 인터넷과 원본 DB에 접근할 수 없는 차단된 별도의 분석 PC에서 분석 예정
▪ 분석PC가 있는 환경은 별도의 분석실로 엄격한 출입통제가 되어 있으며 출입 시 출입 관리대장을 기재</td></tr>
<tr><td>다른 정보와의
결합 가능성</td><td>▪ 가명처리 전 개인정보와 구매정보를 보유하고 있음</td></tr>
</table>

최종 검토의견	▪ 해당 연구는 자사의 데이터를 자사의 내부에서 활용하는 것으로 데이터 자체 위험성과 처리 환경 위험성을 검토할 때 다음과 같은 조치가 필요함 - 결합 가능한 다른정보를 보유하고 있으나 처리 장소를 고려했을 때 결합 가능성은 매우 낮을 것으로 판단됨 - '고객ID'는 개인식별정보로 개인식별 가능성이 매우 높으며 이에 따라 연관관계가 없는 일련번호로 대체할 필요가 있음 - '나이', '주소', '성별'은 그대로 사용하는 경우 조합에 의한 개인식별 가능성이 있으며 이에 따라 다음과 같은 처리가 필요 • '나이' : 주 분석대상이 아닌 13세 미만의 경우 삭제처리가 필요하며 중학생과 고등학생은 하나로 묶어 처리하고 그 외의 나이에 대해서는 1살 단위로 제공하며 90세 이상의 나이에 대해서는 90세 이상으로 처리하는 것이 필요 • '주소' : 동단위와 상세 주소의 경우 통계목적에 필요하지 않기 때문에 삭제하며 시군구 단위의 주소까지만 사용하는 것이 필요 • '성별' : 성별은 분석목적에 필요하므로 그대로 사용 ▪ 구매액 관련 정보들은 구매금액별 특이정보를 검토하여 구매 금액에 대한 적절한 수준의 상단 코딩을 적용(19년 8월 취미용품 114,562,000원 → 1억원 이상)해야 함 ▪ 고객등급의 경우 식별성이 높은 VIP와 S를 하나로 묶어 식별성을 낮출 필요가 있음

식별 위험성 검토 결과보고서 작성 예시(외부 제공)

<table>
<tr><td>가명정보
활용목적</td><td colspan="2">▪ B통계업체는 코로나 이전과 코로나 이후의 상품군별 판매 추이에 대한 통계작성을 위해 당사의 구매정보 데이터 중 2019년 1월부터 12월까지의 주요 상품군별 판매액 정보와 2021년 1월부터 12월까지의 주요 상품군별 판매액 정보를 나이와 성별, 시군구 단위의 주소별 데이터를 분석</td></tr>
<tr><td>가명처리
대상 데이터
항목</td><td colspan="2">▪ 고객ID, 나이, 주소, 성별, 2019년 1월~12월, 2021년 1월~12월까지의 여행용품, 식품류, 의류, 취미용품, 생활용품, 유아용품, 기타의 7개 범주의 구매금액의 월별 합계액, 월별 구매 총 금액, 월별 선호 제품군, 각 년도의 고객 등급 (전체 222개의 컬럼)
▪ 전체 고객 800만명 중 25%를 무작위 샘플링하여 구성한 200만명에 대한 데이터</td></tr>
<tr><td rowspan="3">데이터
위험성</td><td>식별성 유무</td><td>▪ ‘고객 ID’는 개인식별정보임
▪ ‘나이’, ‘주소’, ‘성별’은 조합했을 때 개인의 식별이 가능한 개인식별 가능정보임</td></tr>
<tr><td>특이정보 유무</td><td>▪ 각 범주별 구매금액의 경우 특이정보로 인한 개인 식별성이 발생할 수 있음</td></tr>
<tr><td>재식별시
영향도</td><td>▪ 단순 고객의 구매데이터로 재식별 시 영향도는 크지 않을 것으로 판단됨</td></tr>
<tr><td rowspan="3">처리 환경
검토</td><td>이용 및 제공
형태</td><td>▪ 제3자 제공
- 데이터 제공 계약을 체결하여 데이터를 제공
- 데이터 제공 계약에는 재제공 금지, 목적 달성 후 삭제, 재식별 금지 및 재식별 시 조치에 관한 사항들이 포함되어 있음
▪ B통계업체는 개인정보(가명정보)처리시스템에 대한 ISMS-P 인증을 취득하고 있음</td></tr>
<tr><td>처리 장소</td><td>▪ B통계업체에서 가명정보는 인터넷에 접근할 수 없는 차단된 별도의 분석 PC에서 분석 예정
▪ 분석PC가 있는 환경은 별도의 분석실로 내부적인 출입통제를 적용하는 것으로 파악됨</td></tr>
<tr><td>다른 정보와의
결합 가능성</td><td>▪ B통계업체는 다양한 통계를 생성하는 업체로 유사 업종에 대한 통계정보 등 결합 가능성이 있는 정보를 보유하고 있음</td></tr>
</table>

최종 검토의견	▪ 해당 연구는 자사의 데이터를 B통계업체에 제공하는 것으로 데이터 자체 위험성과 처리 환경 위험성을 검토할 때 다음과 같은 조치가 필요함 - 통계전문업체의 특성 상 다른 정보의 결합가능성이 있으나 처리 장소와 개인정보 보호 수준을 검토할 때 결합에 대한 시도는 거의 없을 것으로 판단됨 - '고객ID'는 개인식별정보로 개인식별 가능성이 매우 높으며 이에 따라 다시 원래의 정보로 대체할 수 없는 Salt값이 포함된 해시처리 등의 기법의 적용이 필요 - '나이', '주소', '성별'은 그대로 사용하는 경우 조합에 의한 개인식별 가능성이 있으며 이에 따라 다음과 같은 처리가 필요 • '나이' : 물품의 주 구매대상이 아닌 20세 미만의 경우 삭제처리가 필요하며 그 외의 나이에 대해서는 일반적인 구매 분석 통계에 사용되는 10살 단위로 제공하며 80세 이상의 나이에 대해서는 80세 이상으로 처리하는 것이 필요 • '주소' : 동단위와 상세 주소의 경우 통계목적에 필요하지 않기 때문에 삭제하며 시군구 단위의 주소까지만 사용하는 것이 필요 • '성별' : 성별은 분석목적에 필요하므로 그대로 사용 ▪ 구매액 관련 정보들은 구매금액별 특이정보를 검토하여 구매 금액에 대한 적절한 수준의 상단 코딩을 적용(19년 8월 취미용품 114,562,000원 → 1억원 이상)하고 금액에 대해서 라운딩을 적용하는 것이 필요 ▪ 고객등급의 경우 식별성이 높은 VIP, S, A를 하나로 B, C를 하나로 D, E, F를 하나로 묶을 필요가 있음

4 3단계 가명처리

☑ 개인정보처리자는 식별 위험성 검토 결과를 기반으로 가명정보의 활용 목적 달성에 필요한 가명처리 방법 및 수준을 정하여 항목별 가명처리 계획을 설정함

- 식별 위험성 요소에 대한 주요 항목에 대하여 위험성을 낮출 수 있는 가명처리 방법 및 수준을 선택

※ 가명처리 기법 등은 [참고자료] 참고1. 정형데이터 가명처리 기술 및 예시 (85p) 참고

- 목적달성 가능성 검토를 위하여 가명처리 전 이용기관과 협의 가능하며, 가명처리 방법 및 수준 정의가 적정하지 않다고 판단되는 경우 다시 식별 위험성을 검토함

항목별 가명처리계획 작성 예시(내부 이용 또는 제3자 제공) 비교

순번	항목명	개인정보유형	내부 이용 또는 제3자 제공			
			처리 방법	처리 수준	처리 방법	처리 수준
1	고객ID	개인식별정보	대체	- 일련번호 대체	대체	- 일련번호 대체
2	나이	개인식별가능정보	범주화	- 14~19세 사이는 14~19세로 범주화	범주화	- 10살 단위 범주화
			상하단 코딩	- 13세 미만 삭제 - 90세 이상은 90세 이상 경계치 입력	상하단 코딩	- 20세 미만 삭제 - 80세 이상은 80세 이상 경계치 입력
3	주소	개인식별가능정보	부분삭제	- 동단위 이하 삭제	부분삭제	- 동단위 이하 삭제
4	성별	개인식별가능정보	처리 없음		처리 없음	
5	2019년 1월 여행용품구매액	개인식별가능정보	범주화	- 상단 99.9%를 초과하는 경우 경계치로 변경 - 모든 금액에 대해 1만단위 라운딩 처리	범주화	- 상단 99.9%를 초과하는 경우 경계치로 변경 - 금액은 다음과 같이 범주화 적용 - 0원 : 0원 - 10만단위 미만 : 1만단위로 라운드 업 - 1,000만 단위 미만 : 10만 단위로 라운딩 - 1,000만 단위 이상 100만 단위로 라운딩
6	2019년 1월 식품류 구매액	개인식별가능정보				
7	2019년 1월 의류 구매액	개인식별가능정보				
8	2019년 1월 취미용품구매액	개인식별가능정보				
9	2019년 1월 생활용품구매액	개인식별가능정보				
10	2019년 1월 구매 총금액	개인식별가능정보	범주화	- 각 구매액과 동일한 처리	범주화	- 각 구매액과 동일한 처리
~	~	~	~	~	~	
217	2021년 12월 식품류 구매액	개인식별가능정보	범주화	- 상단 99.9%를 초과하는 경우 경계치로 변경 - 모든 금액에 대해 1만단위 라운딩 처리	범주화	- 상단 99.9%를 초과하는 경우 경계치로 변경 - 금액은 다음과 같이 범주화 적용 - 0원 : 0원 - 10만단위 미만 : 1만단위로 라운드 업 - 1,000만 단위 미만 : 10 만 단위로 라운딩 - 1,000만 단위 이상 100만 단위로 라운딩
218	2021년 12월 의류 구매액	개인식별가능정보				
219	2021년 12월 취미용품구매액	개인식별가능정보				
220	2021년 12월 생활용품구매액	개인식별가능정보				
221	2021년 12월 구매 총금액	개인식별가능정보	범주화	- 각 구매액과 동일한 처리	범주화	- 각 구매액과 동일한 처리
222	2021년 고객 등급	개인식별가능정보	범주화	- 식별가능성이 높은 VIP와 S를 하나로 묶어 VIP로 처리	범주화	- 다음과 같이 범주화 처리 - VIP, S, A → 1등급 - B, C → 2등급 - D, E, F → 3등급

☑ 개인정보처리자는 '항목별 가명처리계획'을 기반으로 가명처리를 수행함

항목별 가명처리계획에 따른 가명처리 절차(예시)

(원본정보)

소유자명	연락처	주택구분	법정동코드	시도	시군구	읍면동	지번	건물명	전세(천원)	보증금(천원)	월세(천원)	전용면적	공급면적
김철수	090-1234-5678	아파트	2635010700	서울특별시	동작구	사당동	1388-4	한글아파트	-	25,000	750	104.00	84.00
이영희	090-2468-3579	오피스텔	3611011000	대전광역시	서구	둔산동	656	나주시티오	81,250	-	-	56.45	24.32
박민호	090-9876-5432	아파트	4311410100	부산광역시	해운대구	우동	111-13	세종아파트	125,000	-	-	100.00	84.00

선정 선정 선정

(대상선정)

- 목적 : 부동산 임대소득 계산 및 인근지역 시세자료 파악을 위한 연구

소유자명	연락처	주택구분	시도	시군구	읍면동	지번	전세(천원)	보증금(천원)	월세(천원)	전용면적	공급면적
김철수	090-1234-5678	아파트	서울특별시	동작구	사당동	1388-4	-	25,000	750	104.00	84.00
이영희	090-2468-3579	오피스텔	대전광역시	서구	둔산동	656	81,250	-	-	56.45	24.32
박민호	090-9876-5432	아파트	부산광역시	해운대구	우동	111-13	125,000	-	-	100.00	84.00

(위험성 검토)

- 데이터의 식별 위험성과 처리 환경의 식별 위험성 검토 결과를 반영하여 가명처리 방법 및 수준 정의

- 소유자명, 연락처는 개인정보로 분류하고 가명처리(암호화)
- 구체적인 지번은 분석목적에 관계없어 삭제조치 및 시세정보는 분석에 필요한 단위(만원)로 가명처리

분류 + 처리환경

- A사의 부동산 시세정보를 B기관에 제공(계약)
- 제공되는 항목의 '지번'의 경우 등기부열람을 통해 특정개인식별 가능성 존재

식별정보 / 식별가능정보

소유자 명	연락처	주택 구분	시도	시군구	읍면동	지번	전세 (천원)	보증금 (천원)	월세 (천원)	전용 면적	공급 면적
김철수	090-1234-5678	아파트	서울 특별시	동작구	사당동	1388-4	-	25,000	750	104.00	84.00
이영희	090-2468-3579	오피스텔	대전 광역시	서구	둔산동	656	81,250	-	-	56.45	24.32
박민호	090-9876-5432	아파트	부산 광역시	해운대구	우동	111-13	125,000	-	-	100.00	84.00

(소유자 명, 연락처)+Salt값 암호화

삭제 / 라운딩

(가명처리)

ID	주택 구분	시도	시군구	읍면동	전세 (천원)	보증금 (천원)	월세 (천원)	전용 면적	공급 면적
wd4e85D2C1qe89rwqe	아파트	서울 특별시	동작구	사당동	-	25,000	800	104.00	84.00
r5w1e2SXzi4wd64qwz	오피스텔	대전 광역시	서구	둔산동	81,300	-	-	56.45	24.32
ghe6W15Z5ax4Qe24jx	아파트	부산 광역시	해운대구	우동	125,000	-	-	100.00	84.00

☑ 가명처리 과정에서 생성되는 추가정보는 원칙적으로 파기하고 필요한 경우 가명정보와 분리하여 별도로 저장하여야함

- 추가정보의 분리보관은 [제4장 안전성 확보 조치] 기술적 보호조치 (81p) 참고

5 4단계 적정성 검토

☑ 1, 2, 3단계의 가명처리에 대해 결과 적정성을 최종 검토함

- 가명처리가 적정하게 수행되었는지 확인하고, 가명처리 한 결과가 가명정보의 처리 목적을 달성하기 위해 적절한지 등 검토
- 가명처리 적정성 검토는 내부 인원을 활용하여 자체적으로 검토하거나, 외부전문가를 통하여 검토할 수 있음

※ 최소 3명 이상으로 검토위원회를 구성하는 것을 권고하며, 외부전문가 섭외 시 가명정보 지원 플랫폼(dataprivacy.go.kr) ▸ 컨설팅·기술지원 ▸ 전문가 풀 지원 메뉴에서 분야별 가명정보 전문가 풀을 참고할 수 있음

- 재식별 가능성이 있는 경우 1, 2, 3단계의 개인정보 가명처리 절차를 다시 수행하거나 부분적으로 추가 가명처리를 수행함

※ 데이터의 분포, 내용 등을 검토하여 특이정보가 추가로 생성 또는 발견된 경우 재식별 가능성을 낮추기 위한 적절한 조치를 취하여야 함

☑ 적정성 검토는 ① 필요서류, ② 처리 목적 적합성, ③ 식별 위험성, ④ 가명처리 방법 및 수준의 적정성, ⑤ 가명처리의 적정성, ⑥ 처리 목적 달성 가능성 단계로 검토가 이루어 짐

☑ 적정성 검토 시 위원장을 선정하여 절차에 따라 검토를 진행할 수 있도록 하고, 종합적인 내용과 각 검토위원의 의견을 고려한 최종검토결과 및 종합검토의견을 개인정보처리자에게 제출함

적정성 검토 단계별 절차

단계	검토내용	부적정 시 조치사항
① 필요서류	사전준비 단계에서 필요서류가 법·제도 목적에 적합하게 작성되었는지 검토	해당 자료 보완 작성 등 재점검
	가명정보 이용 제공 신청서*, 가명정보 안전조치 의무이행 확약서 * 처리 위탁 및 제3자 제공에 대한 계약서, 이용환경에 대한 보호수준을 확인할 수 있는 서류 등	
▼		
② 목적 적합성 등	보호법에서 정한 가명정보 처리 목적에 해당하는지, 처리목적을 구체적으로 설정하였는지 검토	목적 명확화 및 재설정
	가명처리 및 결합 목적 증빙자료(가이드라인 참고5)	
▼		
③ 식별 위험성에 대한 결과 적정성	가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 식별 위험성 검토 점검표 및 결과보고서 기반으로 검토	식별위험성 재검토, 결과보고서 보완 등
	개인정보 유형 분류표, 활용데이터 요구 수준표, 식별 위험성 검토 결과보고서, 가명정보 안전조치 의무이행 확약서	
▼		
④ 항목별 가명처리 계획의 적정성	가명처리 단계에서 위험성 검토 결과를 반영하여 항목별 가명처리 방법 및 수준을 적정하게 계획하였는지 확인	항목별 가명처리계획 보완
	식별 위험성 검토 결과보고서, 항목별 가명처리계획	
▼		
⑤ 가명처리 결과에 대한 적정성	계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인 ※ 특히 대용량 정보의 경우 중간에 처리되지 않은 부분이 있을 수 있으므로 가능한 가명정보 항목 전체를 확인 필요	가명처리가 적정하지 않은 경우 가명처리를 다시 수행하거나 부분적으로 추가 가명처리를 수행
	가명정보 처리 기초자료 명세서 ※ 원본데이터 특징, 유형, 분포 등 가명정보의 생성 및 활용 등과 관련하여 설명할 수 있는 기초자료	
▼		
⑥ 처리 결과에 대한 목적 달성 가능성	가명처리된 정보가 당초 가명정보 처리 목적을 달성할 수 있는지 여부 검토	항목별 가명처리계획 보완, 추가 가명처리 등
	가명처리 및 결합 목적 증빙자료(가이드라인 참고5)	

6 5단계 안전한 관리

☑ 적정성 검토 이후 생성된 가명정보는 법에 따라 기술적·관리적·물리적 안전조치 등 사후관리를 이행하여야 함(보호법 제28조의4)

> **제28조의4(가명정보에 대한 안전조치의무 등)** ① 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 경우에는 원래의 상태로 복원하기 위한 추가 정보를 별도로 분리하여 보관·관리하는 등 해당 정보가 분실·도난·유출·위조·변조 또는 훼손되지 않도록 대통령령으로 정하는 바에 따라 안전성 확보에 필요한 기술적·관리적 및 물리적 조치를 하여야 한다.
> ② 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 경우 처리목적 등을 고려하여 가명정보의 처리 기간을 별도로 정할 수 있다.
> ③ 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하고자 하는 경우에는 가명정보의 처리목적, 제3자 제공 시 제공받는 자, 가명정보의 처리 기간(제2항에 따라 처리기간을 별도로 정한 경우에 한한다) 등 가명정보의 처리 내용을 관리하기 위하여 대통령령으로 정하는 사항에 대한 관련 기록을 작성하여 보관하여야 하며, 가명정보를 파기한 경우에는 파기한 날부터 3년 이상 보관하여야 한다.

※ 구체적 내용은 [제4장 안전성 확보 조치] (77p) 참고

1. 재식별 금지 및 재식별 가능성 모니터링

□ 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 자는 특정 개인을 알아보기 위한 목적으로 가명정보를 처리해서는 아니 되며(보호법 제28조의5 제1항), 가명정보 처리 과정에서 우연히 특정 개인이 식별되는 경우 처리중지, 회수, 파기 등과 같이 위험을 제거하기 위한 적절한 조치를 즉시 수행하여야 함(보호법 제28조의5 제2항)

> **제28조의5(가명정보 처리 시 금지의무 등)** ① 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 자는 특정 개인을 알아보기 위한 목적으로 가명정보를 처리해서는 아니 된다.
> ② 개인정보처리자는 제28조의2 또는 제28조의3에 따라 가명정보를 처리하는 과정에서 특정 개인을 알아볼 수 있는 정보가 생성된 경우에는 즉시 해당 정보의 처리를 중지하고, 지체 없이 회수·파기하여야 한다.

□ 또한, 개인정보처리자는 가명정보 처리 과정에서 특정 개인이 식별될 위험이 있는지 여부를 지속적으로 모니터링 하는 등 가명정보를 안전하게 처리하여야 함

※ 가명처리 기술의 취약점으로 인한 재식별 가능성 및 다른 정보와 결합 시 재식별가능성이 있는 새로운 공개데이터의 발생 여부

2. 안전조치 시행

□ 개인정보처리자는 사전준비 단계에서 수립한 내부 관리계획에 따라 가명정보를 안전하게 관리하여야 함

3. 가명정보 처리 관련 기록 작성 및 보관

□ 개인정보처리자는 가명정보의 처리 목적, 개인정보 항목, 이용내역, 제3자 제공 시 제공받는 자를 작성하여 보관하여야 함

4. 가명정보 처리(활용) 이후 데이터 파기

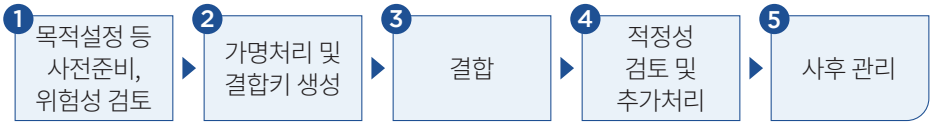
□ 개인정보처리자는 가명정보의 처리 목적을 달성하였거나 개인정보처리자가 설정한 별도의 처리 기간이 도래하였을 시 데이터를 파기하여야 함

기타 참고사항 : 내부 결합

▶ 개인정보처리자는 자신이 보유하고 있는 가명정보를 결합하여 활용할 수 있으며, 결합 절차가 정해져 있지는 않지만 결합 과정에서 특정 개인을 알아 볼 수 없도록 유의하여야 함

※ 안전한 결합을 위해 결합키를 이용한 결합방법을 선택할 수 있음

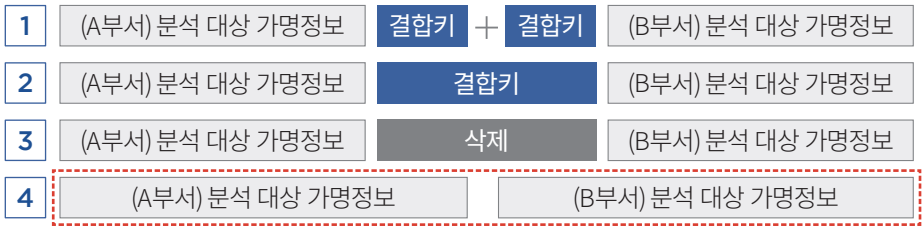
가명정보 내부 결합 절차도



- 개인정보처리자는 결합된 정보를 활용할 때 특별한 사유(시계열 분석 등)가 없는 한 결합키 등 결합을 위해 사용한 정보는 삭제하여야 함

※ (주의) 결합키 생성에 이용된 알고리즘, 매핑테이블 등은 추가정보에 해당하므로, 결합된 가명정보와 분리하여 보관하여야 하고, 접근권한을 분리하여야 함

가명정보 내부 결합 방식 예시



※ 결합키 생성 등의 구체적인 내용은 [제3장 가명정보 결합 및 반출] (57p) 참고

참고

비정형데이터 가명처리 기준

1 개요

☑ AI 기술 발전과 컴퓨팅 자원 발달로 데이터 활용수요가 전통적 정형데이터(수치)에서 비정형데이터(이미지, 영상, 음성, 텍스트)로 변화

* 전 세계 데이터 중 이미지, 영상, 음성, 텍스트 등 비정형데이터가 최대 90%를 차지 (IDC, '23)

참고 정형데이터와 비정형데이터의 차이점

▶ 정형데이터

(정의) 정해진 규칙에 맞게 구조화된 형식으로 존재하는 데이터

※ 예) DB에 열과 행으로 저장된 테이블형식의 자료 등

(특징) 데이터 연산, 분석 등 데이터 처리방식, 가명처리 기술·방법이 비교적 단순

▶ 비정형데이터

(정의) 일정한 규격이나 정해진 형태가 없이 구조화되지 않은 데이터

※ 예) 사진, 비디오, 통화음성, 대화기록, 보고서, 메일 본문 등

(특징) 연구목적·환경에 따라 데이터 처리방식 및 가명처리 기술·방법이 복잡·다양

☑ 비정형데이터도 가명정보 특례를 통해 과학적 연구 목적 등으로 정보주체 동의 없이 가명처리하여 AI 연구개발 등에 활용 가능

참고 비정형데이터의 가명처리·활용 예시

▶ **(이미지·영상)** 특정 질병을 진단(보조)하는 의료 AI 연구개발을 위해 병원이 보유한 MRI, CT, X-ray 사진·영상을 가명처리 후 학습데이터로 활용

▶ **(이미지·영상)** 불법현수막을 탐지하여 알려주는 지능형 CCTV 개발을 위해 지자체가 보유한 공공장소 CCTV 촬영영상을 가명처리하여 AI 연구개발에 활용

▶ **(음성·텍스트)** 민원인 상담·대응을 위한 음성생성 AI를 개발하기 위해 공공기관이 보유한 민원상담 음성정보와 상담기록 정보를 가명처리하여 학습데이터로 활용

2 비정형데이터 가명처리·활용의 특수성 및 고려사항

☑ (개인식별성 판단의 어려움) 개인식별 가능 정보와 그렇지 않은 정보의 구분이 상대적이며, 처리 목적·환경 등에 따라 다르게 판단될 수 있음

예시

- 얼굴 CT 사진 1장으로는 개인식별 위험성이 낮지만, 여러 위치·각도에서 촬영한 얼굴 CT 사진을 여러장 결합하면 얼굴형상 재건이 가능하여 개인식별 위험성 증가
- 눈·코·입을 알아볼 수 없는 거리에서 찍힌 CCTV 영상은 통상 개인식별 위험성이 낮지만, 흉터, 문신, 머리스타일 등 특이한 신체 특징이 있는 경우 개인식별 위험성이 높음

☑ (가명처리 기술의 불완전성) 비정형데이터 내 개인식별 위험성이 있는 모든 항목을 완벽하게 탐지·처리할 수 있는 기술이 부재

예시

- 이미지·영상 데이터의 경우, 해상도, 조명 각도, 객체 크기 등에 따라 얼굴 등을 탐지하지 못하는 경우가 존재 → 최근 AI 기술의 객체 탐지 정확도는 90~98% 수준[1)]
- '신뢰역 1번출구 앞 파란건물 1층 1호가 우리집' 텍스트를 주소로 인식하지 않아 처리하지 않거나, '김신뢰 김밥' 등 상호명을 개인정보(이름)로 인식해 불필요하게 삭제

☑ (재식별 공격 위험) AI 및 데이터 복원기술 발달로, 다른 정보와의 연계·결합 없이도 개인 재식별 공격 위험성 증가

예시

- 음성변조 규칙을 몰라도 화자의 원본 목소리를 복원할 수 있는 기술 존재[2)]
- 가명처리된 사진의 모자이크 패턴을 몰라도, AI를 통해 모자이크된 사진을 원본에 가깝게 복원해낼 수 있는 기술 연구 중[3)]

⇒ 비정형데이터 가명처리·활용 시 데이터 처리 맥락(context), 가명처리 기술의 한계, 재식별 공격 위험 등을 고려하여 개인식별 위험성을 낮춰야 함

1) Kaur, J., Singh, W., "Tools, Techniques, datasets and application areas for object detection in an image: a review", Multimed Tools Appl 81 (2022)
2) Deng, Jiangyi et al, "Catch You and I Can: Revealing Source Voiceprint Against Voice Conversion", ArXivabs/2302.12434 (2023)
3) R. Dahl, M. Norouzi and J. Shlens, "Pixel Recursive Super Resolution", 2017 IEEE International Conference on Computer Vision (ICCV) (2017)

3 비정형데이터 가명처리 기본원칙

1 데이터 처리 목적·환경, 민감도 등을 종합적으로 고려하여 개인식별 위험성이 있는 정보를 판단하고, 합리적인 처리 방법·수준 설정

※ 비정형데이터 식별 위험성 체크리스트(52p), 항목별 조치 가이드(55p) 참고 가능

□ 연구목적에 맞춰 데이터 자체 훼손을 최소화하면서 관리적·환경적 통제 등 다양한 안전성 확보 방안 적용 가능

- 연구목적 달성에 필수적인 정보항목을 남기는 대신 그 외 정보항목에 대한 가명처리 수준을 높이거나, 다른 정보 및 소프트웨어(SW) 반입제한 등 충분한 안전조치를 보완하여 활용

2 가명처리 기술의 한계 등을 보완하기 위해, 사전 준비단계(연구 및 기술개발 기획 시)부터 위험성을 충실히 검토하고 적절한 안전조치를 수행

□ 가명처리 기술의 한계 보완을 위해 다음 조치를 이행할 것을 권고

❶ 가명처리 기술의 적절성·신뢰성을 확인할 수 있는 근거 작성·보관

❷ 가명처리 기술 적용 이후, 처리 결과에 대해 자체적인 검수 수행

❸ 가명처리 적정성 검토 과정에서 ❶, ❷를 포함하여 점검(외부전문가 과반 이상 참여 바람직)

□ 사전에 식별된 개인정보 침해 위험을 예방하기 위해서 가명정보 활용에 참여하는 각 기관의 내부통제 강화 노력이 병행될 필요

□ 가명정보의 처리목적을 달성하면 신속히 가명정보를 파기하여 사후적 위험 최소화

3 데이터 복원기술 발달 등에 대응하여, 가명처리된 비정형데이터 활용 시 관련 시스템 · SW의 접근·사용 제한 등 통제방안 마련

* 원본 복원에 활용될 수 있는 추가정보 분리보관, 복원 SW에 대한 접근권한 제한 등

□ AI 개발·활용 상황에서 나타날 수 있는 다양한 위험을 사전에 완벽하게 제거하는 것은 불가능하므로, AI 서비스 제공과정에서도 개인식별 위험 등 정보주체 권익 침해 가능성을 지속 모니터링

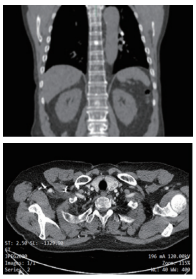
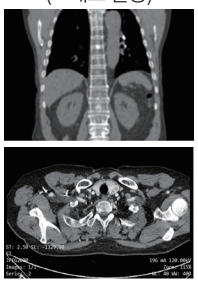
주요 비정형데이터 가명처리 시나리오 예시

※ 아래 가명처리 시나리오는 실제 비정형데이터를 가명처리하여 활용했던 사례를 관련 기업·기관 및 전문가 논의를 통해 재구성한 것으로 단순 참고용이며, 처리자 및 적정성 검토 위원회 등의 판단에 따라 데이터 활용 분야·상황에 맞게 가명처리 방법·수준 등을 자유롭게 적용할 수 있음

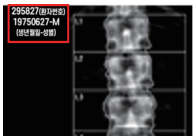
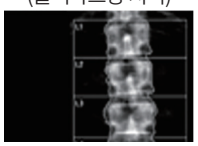
(사례1) 유방암·골밀도 감소 여부 진단 AI 개발

대학병원이 보유한 유방암 환자 CT사진(영상·이미지) 및 병리기록지(텍스트)를 가명처리하여 유방암 및 골밀도 감소 여부 진단 AI 개발을 위한 내부 연구에 활용한 사례

개인식별 위험이 발생하지 않도록 처리환경을 안전하게 통제하고 복원 SW 반입 제한 조치 등을 통해 별도의 가명처리 없이 CT 사진을 그대로 활용

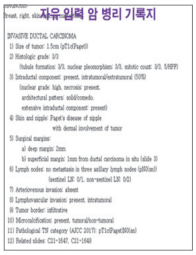
〈흉부 CT사진〉 	**개인식별 위험성 검토**	- 흉부 CT사진만으로는 개인식별 위험성 거의 없음 - 개인당 200장씩 촬영된 CT사진이 활용되는 연구로서 3차원 재건 기술 등을 통해 신체 형상의 입체적 복원이 가능하고, 복원 시 특이한 외형·흉터 등이 있는 극히 일부 환자의 경우 낮은 확률로 개인식별 위험성 존재 - 클라우드 기반 폐쇄연구분석환경*을 이용하고 인가되지 않은 데이터·프로그램 반입을 철저히 통제하고 있어 3차원 재건기술 적용 불가 * 클라우드 서버에 데이터를 저장하고 타 외부망에서는 클라우드 서버 접속이 제한되는 분석실에서 인가받은 인원만 데이터 접근 가능	(그대로 활용) 
	데이터 처리 방안	⇒ 3차원 재건으로 인한 개인식별 위험성이 존재하나, 환경적 통제로 인해 해당 위험의 발생 가능성이 없으므로 별도의 가명처리 없이 그대로 활용 가능	

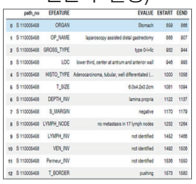
이미지 내 개인식별 위험성이 있는 메타데이터를 삭제하고 활용

〈CT사진 내 환자관련정보〉 	**개인식별 위험성 검토**	- 이미지 내 표시된 환자관련정보*는 타 정보와 결합되어 분석될 경우 개인식별 위험성이 있음 * DICOM 헤더정보(환자번호, 생년월일, 성별) 표시 - 해당 정보는 연구에 필요하지 않은 정보임	(블랙마스킹 처리) 
	데이터 처리 방안	⇒ 블랙마스킹 기법을 통해 환자관련정보 삭제	

비정형 텍스트데이터를 정형데이터 형태로 변환하여 활용

〈병리기록 텍스트〉


자유 입력 암 병리 기록지

개인식별 위험성 검토	- 암병리기록지에는 연구에 불필요한 다양한 개인 식별가능정보가 정제되지 않은 형태로 존재하여 개인식별 위험이 존재	(정형데이터로 변환 후 활용)
데이터 처리 방안	⇒ 자연어처리 기술을 통해 정형데이터로 변환한 후에 활용하고 개인식별 위험성이 있는 항목이 있으면 가명처리 수행 ⇒ 자연어처리 기술 및 텍스트데이터 가명처리 기술의 정확도가 100%가 아니므로, 정형데이터 변환 후 추가 전수검사 등을 통해 보완	

(사례2) 구강질환 진단 AI 개발

대학병원이 보유한 구강 건강검진 촬영 사진(이미지)을 가명처리한 뒤 기업에 제공하여, 충치·치주염 등 구강질환을 진단하는 AI 연구개발에 활용한 사례

연구 목적에 필요 없는 영역을 블러링 처리하고, 메타데이터를 삭제하여 활용

〈구강 촬영사진〉

개인식별 위험성 검토	- 구강사진 자체로는 개인식별 위험성 거의 없음 - 충치 영역 외 부분은 연구에 필요 없음 - 구강사진에 대한 메타데이터(이름, 나이 등)는 구강사진과 결합되어 개인식별 위험성 존재	(충치부분 : 그대로 활용) (그 외 : 블러링 처리)
데이터 처리 방안	⇒ 연구에 필요한 충치 영역은 그대로 활용하고, 연구에 필요 없는 그 외 영역은 블러링 처리 ※ 블러링 수준은 현재 복원기술 발전수준 및 데이터 처리 환경(타 정보·복원기술 접근성) 등을 고려하여 설정 ⇒ 메타데이터는 연구에 필요 없어 삭제	

(사례3) 안면골 골절 진단 AI 개발

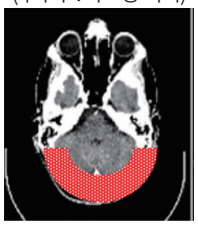
대학병원이 보유한 Facial CT 사진을 가명처리하여, 안면골(얼굴뼈) 골절여부를 진단하는 AI 개발을 위한 공동연구를 대학병원과 민간기업이 함께 수행한 사례

개인식별 위험성을 낮추기 위해 연구목적에 필요 없는 영역만 마스킹 처리하여 활용

〈Facial CT사진〉

개인식별 위험성 검토	- CT사진 자체로는 개인식별 위험성 거의 없음 - 3차원 재건 기술을 통해 입체적 복원이 가능하며, 복원 시 특이한 얼굴·외형, 알려진 얼굴 등 극히 일부 환자의 경우 낮은 확률로 개인식별 위험성 존재 - 대용량 영상·이미지를 활용하는 연구로 3차원 재건이 가능하나, 가장자리 마스킹 기법을 활용하여 3차원 재건 공격 위험을 낮출 수 있음 - 후두부(뇌 뒷부분) 영역은 연구에 필요 없음
데이터 처리 방안	**⇒연구에 필요한 안면부는 그대로 활용하고, 연구에 필요없는 후두부는 마스킹하여 3차원 재건위험을 낮추고 활용**

(안면부 : 그대로 활용)
(후두부 : 마스킹 처리)



(사례4) 자율주행차 주행 시 비정상 상황인지 AI 개발

연구기관이 보유한 도로 주행상황 촬영 영상을 가명처리한 뒤 기업에 제공하여, 자율주행자동차 운행 시 비정상 상황*을 인지하는 AI 연구개발에 활용한 사례

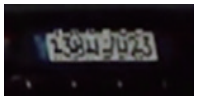
* 사람이 차도에 뛰어드는 상황, 다른 차가 갑자기 앞에 끼어드는 상황, 무단횡단 등

연구 목적에 필요 없는 영역만 마스킹 처리하여 활용

〈얼굴·차량번호판〉

개인식별 위험성 검토	- 사람의 얼굴이 선명히 보이는 경우, 차량 번호판이 그대로 노출되어 차량탑승자 유추가 가능한 경우 등에 개인식별 위험성 존재 - 연구목적상 사람·차량의 전체 형상과 움직임만 파악 가능하면 되므로 얼굴·차량번호판은 마스킹해도 무방
데이터 처리 방안	**⇒얼굴·차량번호판 영역을 사람과 컴퓨터가 식별 불가능한 수준으로 마스킹하여 활용**

(마스킹 처리)



(사례5) 고속도로 다인승전용차로 단속 AI 개발

지자체가 CCTV로 촬영한 고속도로 통행차량 이미지를 가명처리한 뒤 기업에 제공하여, 다인승전용차로 위반*을 단속하는 AI 개발에 활용한 사례

* 3명 이상 승차하지 않은 승용·승합자동차가 다인승전용차로를 이용한 경우

연구 목적에 필요 없는 영역만 블러링 처리하여 활용

〈도로 통행차량 사진〉

〈특이점 있는 차량〉

개인식별 위험성 검토	- 탑승자의 얼굴이 선명하게 촬영된 경우, 차량 외부에 특이점이 있는 경우 등은 개인식별 위험성 존재 - 연구 목적상 특정인 식별·구분이 필요없고 (1) 사람인지 아닌지 여부, (2) 차량 탑승인원이 몇 명인지만 확인할 수 있으면 됨 - AI가 사람인지 여부는 판단할 수 있도록 하되, 탑승자가 누구인지는 판별 불가능하도록 블러링 등 처리 필요
데이터 처리 방안	⇒ 특이점이 있는 차량 이미지는 삭제 ⇒ 블러링 수준(1~10단계)별로 데이터를 가명처리한 후, 식별위험이 없으면서 AI 정밀도를 어느 정도 확보할 수 있는 블러링 수준을 결정 ※ 적정성 검토 단계에서 목적 달성 가능성, 학습데이터의 개인식별 위험성 등을 검증

(① 탑승자 위치·범위 파악)

▼

(② 블러링 처리)

(사례6) 한국어 대화가 가능한 AI 챗봇 개발

인공지능 챗봇 전문기업이 채팅앱을 통해 수집한 사용자 간 일상대화 텍스트 데이터를 가명처리하여 한국어 대화가 가능한 AI 챗봇 연구개발에 활용한 사례

개인식별 위험 항목을 엄격히 필터링하여 제거하고 메타데이터는 삭제하여 활용

〈대화텍스트 파일〉

개인식별 위험성 검토	- 일상대화 데이터(텍스트)에는 사생활 관련 정보 등 개인식별 위험성이 높은 정보가 상당 수 포함
데이터 처리 방안	⇒ 메타데이터(대화 사용자 ID)를 삭제하고 랜덤ID로 대체하여 특정 개인과의 연결성을 제거 ⇒ 개인식별 위험이 있는 항목들을 필터링하여 가명처리(치환*, 삭제) * 이메일 주소를 '[MAIL]'로 대체하는 등

(메타데이터 삭제, 개인식별정보 필터링·제거)

학습에 활용된 가명정보가 AI 챗봇 답변에 그대로 발화되지 않도록 조치

〈챗봇 답변 시 위험성〉 	**개인식별 위험성 검토**	- 언어모델 학습에 활용된 가명정보가 AI 챗봇의 답변으로 발화될 시, 제대로 처리되지 않은 가명정보의 노출 등으로 인한 개인식별 위험성이 높음	(학습DB, 답변DB 분리) 
	데이터 처리 방안	⇒ 언어모델 학습을 위한 '학습데이터베이스'와 챗봇 답변을 위한 '답변데이터베이스'를 분리하여 학습에 활용된 문장이 그대로 노출되지 않도록 조치 * 답변데이터베이스에 개인 식별위험이 있는지 충분히 점검	

(사례7) 콜센터 직원 실습용 가상상담 시나리오 생성 AI 개발

기업에서 직원-고객간 음성 상담정보를 가명처리하여 자사 콜센터 직원들을 위한 상담 실습교육용 AI 개발에 활용한 사례

음성정보를 텍스트로 변환(STT, Speech To Text)한 뒤 가명처리하여 활용

〈음성 상담파일〉 	**개인식별 위험성 검토**	- 상담음성파일에는 고객 및 상담사의 실제 음성데이터(목소리, 음색, 억양, 발음 등)가 포함되어 있고, 대화내용에는 다양한 개인식별가능정보가 정제되지 않은 형태로 존재 - 가상상담 시나리오 생성 AI 개발에는 상담목적 및 고객 특성에 따른 질의-응답과 대화의 흐름 파악이 중요하며 실제 음성 자체는 필요하지 않음	(① 텍스트로 변환)  (② 개인식별정보 치환·삭제) 
	데이터 처리 방안	⇒ 음성변환(STT) 기술을 통해 텍스트로 변환한 뒤, 개인식별 위험성이 있는 항목들을 가명처리(치환·삭제)하여 활용 ⇒ 텍스트데이터에 대한 가명처리 기술의 정확도가 100%가 아니므로, 추가 전수 검사를 통해 식별 위험성이 있는 정보를 제거	

4 비정형데이터 가명처리 단계별 고려사항

- 비정형데이터의 가명처리와 관련하여서는「가명정보 처리 가이드라인」제2장의 가명처리 단계별 절차를 동일하게 따르되, 비정형데이터의 특수성을 반영한 개인식별 위험성 검토 및 안전조치 사항을 추가적으로 고려하여 시행하는 것을 권고

개인정보의 가명처리 단계별 절차

1단계	2단계	3단계	4단계	5단계
목적 설정 등 사전준비	위험성 검토	가명처리	적정성 검토	안전한 관리

1 사전준비 단계

※ 가명정보 처리 목적을 설정·검토하고 목적에 맞는 가명처리 대상을 선정하는 단계

□ 비정형데이터 내 개인식별 가능성이 있는 항목들을 도출하고, 목적 달성에 필요한 항목의 종류와 범위를 명확히 하여 가명처리 대상 선정

2 위험성 검토 단계

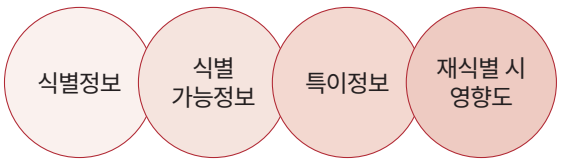
※ 가명처리 대상·처리 환경의 위험성을 검토하여 가명처리 방법·수준에 반영하기 위한 단계

□ 비정형데이터의 특성을 고려하여 '데이터 자체 식별 위험성'과 '처리 환경의 식별 위험성'을 종합적으로 검토하여 가명처리 방법·수준을 결정

(1) 데이터의 식별 위험성 검토

□ 개인식별 가능성이 높은 정보(❶식별정보, ❷식별가능정보), ❸특이정보, ❹데이터 특성만으로 재식별 시 영향도가 높은 항목이 있는지 검토

데이터 자체 개인식별 위험성 요소



- (식별성) 비정형데이터는 정형데이터와 달리 식별정보*와 식별가능정보**의 절대적인 구분이 어렵기 때문에, 처리 목적·방법 등 데이터 처리 맥락(context)을 고려하여 항목별 개인식별 가능성을 상대적으로 판단하여야 함

* (식별정보) 특정 개인과 직접적으로 연결되어 다른 사람과 구분되는 정보

** (식별가능정보) 단일 항목으로는 식별 가능성이 없으나, 가명처리 대상이 되는 다른 항목과 결합하는 경우 식별 가능성이 높아지는 정보

예시

- 사람 얼굴이 촬영된 CCTV 영상을 활용하고자 하는 경우
  - ▶ (상황에 따라 식별 가능성이 달라지는 경우)
    - 촬영된 장소가 명확하고 얼굴이 선명하게 촬영된 경우 식별 가능성이 높음
    - 촬영된 장소가 불분명하고 얼굴이 상당히 작게 찍히거나 해상도가 낮아, 특정 개인을 구분하게 어렵게 촬영된 경우 식별 가능성이 낮음
  - ▶ (처리 목적·방법에 따라 식별성이 달라지는 경우)
    - 얼굴 촬영영상을 신체특징, 걸음걸이, 이동동선 등 영상 내 다른 항목과 결합하여 촬영된 사람의 성별과 행동패턴을 분석하기 위한 목적으로 활용된다면, 식별 가능성이 높음
    - 얼굴 촬영영상이 단순히 사람인지 사람이 아닌지를 구분하기 위한 목적으로만 활용되거나, 다른 항목과 결합되지 않고 활용된다면 비교적 식별 가능성이 낮음
- 환자의 MRI 사진이 메타데이터를 포함하고 있고, 해당 메타데이터에 환자이름, 환자번호 등 개인식별번호가 포함되어 있는 경우 식별 가능성이 높음

- (특이정보) 개인의 특이한 신체적·외형적·행태적 특징 또는 개인과 연관된 객체·사물 등의 특이성으로 인한 식별 위험성에 대해 검토

예시

- 특이한 신체적·외형적 특징으로 인해 특정 개인을 식별할 가능성이 존재하는 경우
  - (이미지·영상) 신체적 특징, 체형, 머리스타일, 문신, 흉터 등이 특이한 경우
  - (음성) 발음(구개 파열음 등), 음색(목소리) 등이 특이한 경우
- 특이한 행태적 특징으로 인해 특정 개인을 식별할 가능성이 존재하는 경우
  - (이미지·영상) 걸음걸이, 몸짓이나 행위 등이 특이한 경우
  - (음성) 억양(사투리, 말투), 반복 어휘, 언어적 습관 등이 특이한 경우
  - (텍스트) 반복 어휘, 어법, 문체, 언어적 습관 등이 특이한 경우
- 개인과 연관된 특이한 객체·사물 등으로 인해, 개인 식별 가능성이 존재하는 경우
  - (이미지·영상) 거주하는 집, 차종(희귀 슈퍼카 등), 옷차림, 반려동물 등이 특이한 경우

- (재식별 시 영향도) 가명처리된 비정형데이터가 재식별될 경우 특정 정보주체에게 사회적 파장 등 영향도가 높은 항목이 있는지 검토

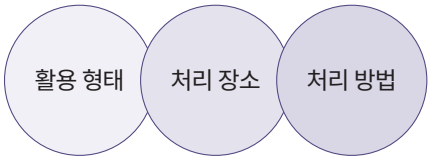
예시

- 범죄 피해현장이 촬영된 CCTV 영상, 민감한 사생활이 포함된 녹취파일 등

(2) 처리 환경의 식별 위험성 검토

□ ❶가명정보 활용 형태, ❷가명정보 처리 장소, ❸가명정보 처리방법 등 가명정보 처리 환경에 따라 발생할 수 있는 식별 위험성을 검토해야 함

처리 환경에 따른 개인식별 위험성 요소



- (활용 형태) 처리자(또는 취급자)가 보유하고 있는 정보 또는 접근·입수 가능한 정보와 이용 범위·유형을 고려하여 식별가능한 항목이 있는지 검토

예시

- 의사가 분석하고자 하는 환자 X-ray 사진 데이터셋에 본인이 직접 진료한 환자의 데이터가 포함된 경우, 해당 의사가 보유하고 있는 진료데이터와 진료경험을 토대로 특정 환자를 추정할 수 있는 위험성이 높아짐
- CT 영상에 포함된 메타데이터를 삭제할 경우, 데이터셋에 포함된 특정 환자에 대한 배경지식을 알고 있는 자가 해당 환자를 추정할 수 있는 위험성이 낮아짐
- AI 특성에 따라 텍스트정보를 학습하여 그대로 암기하고, 개인식별 위험이 있는 정보들을 그대로 생성·출력하는 형태로 구현될 경우, 추론 공격 등에 노출될 위험성이 높아짐
- AI가 CCTV에 촬영된 사람의 형상만 파악·집계하여 통계값만 추출하는 경우, 개개인에 대한 특징을 분석하지 않기 때문에 특정 개인이 식별될 위험성이 낮아짐

- (처리 장소) 가명정보가 해당 가명정보 외에 다른 정보의 접근·입수, 재식별 기술의 접근이 제한된 환경·장소에서 처리되는지 검토

예시

- 가명처리(모자이크)된 사진을 복원시킬 수 있는 기술이 존재하는 경우, 해당 사진에 대해 복원기술을 적용할 수 있는 환경에서 처리된다면 식별 위험성이 높아짐
- 개인식별이 어렵도록 음성변조된 음성정보를 처리할 시, 해당 음성정보와 비교·대조할 수 없는 다른 정보를 입수할 수 없는 환경에서 처리된다면 식별 위험성이 낮아짐

- (처리 방법) 가명정보가 다른 정보와 연계·결합되거나 반복 제공 등이 예정된 경우, 식별가능성이 높아지는 항목이 있는지 검토

예시

- 각 환자를 다양하게 촬영한 100장의 두경부 CT사진 활용 시, 영상 재건 기술 등을 활용하여 얼굴 형상을 3차원으로 복원해낼 수 있는 위험성이 존재
- 환자당 1장의 CT사진만 활용하여 연구를 수행하는 경우, 영상 재건 기술을 활용하여 얼굴 형상을 복원해낼 수 없으므로 개인식별 위험성이 낮음

3 가명처리 단계

※ 위험성 검토 결과 및 항목별 가명처리 계획을 기반으로 실제 가명처리를 수행하는 단계

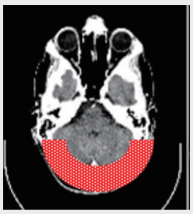
□ 가명처리가 필수적이지 않은 항목과 가명처리가 필요한 항목을 구분하고, 가명처리가 필요한 항목은 합리적인 가명처리 방법과 수준을 결정

⇒ 비정형데이터 항목 중 처리목적 달성을 위해 반드시 필요하지만, 개인식별 위험성이 낮은 정보는 가명처리하지 않고 그대로 사용 가능

⇒ 비정형데이터 항목 중 가명처리해도 처리목적 달성이 가능하고, 개인식별 위험성이 높은 정보는 가명처리하여 활용하여야 함

예시

- CT 사진을 활용하여 얼굴뼈 골절 진단을 위한 AI 연구개발 시
  - ▶ 얼굴 앞부분(안면부)
    : 얼굴뼈 골절 진단에 반드시 필요하므로, 그대로 활용
  - ▶ 뇌 뒷부분(후두부)
    : 얼굴뼈 골절 진단에 필요하지 않으므로, 마스킹처리하여 활용



예시

▪ 구강사진을 활용하여 충치를 분석·진단하는 AI 개발 시

▶ 충치 의심영역으로 라벨링된 치아부분
: 충치 분석·진단에 반드시 필요하므로, 그대로 활용

▶ 충치 의심영역 외 일반치아 및 잇몸 부분
: 충치 분석·진단에 필요하지 않고, 구강 내 특이구조나 치료흔적 등을 통한 개인식별 위험성이 존재하므로, 블러링 처리하여 활용

□ 비정형데이터 가명처리 기술 적용시, 해당 기술의 적절성·신뢰성을 평가하고 관련 근거*를 작성하여 보관할 것을 권고

예) CT사진의 가장자리를 마스킹 솔루션을 적용하여 가명처리한 경우, 해당 솔루션의 관련 가명처리 기능, 솔루션의 객체 인식률·처리 정확도(오류율)에 대한 증빙자료 등

□ 비정형데이터 가명처리의 기술적 한계를 보완하고 잔존 위험(Residual Risk)을 낮추기 위해, 처리결과에 대한 자체적인 추가검수 등이 필요

- 가명처리 목적, 데이터 성격, 적용한 기술의 특성, 처리환경 통제수준 등을 고려하여 위험도에 비례한 적절한 검수방법*을 적용하되, 검수 과정에서 발견된 위험을 낮추기 위한 조치사항을 기록·보관하고 이에 대한 적정성 검토를 받을 것을 권고

* 예) 사람에 의한 육안 전수검사, 통계적으로 신뢰할 수 있는 샘플링 검사 등을 통해 합리적으로 예측가능한 위험을 낮추기 위해 충분한 노력이 시행되었는지 검토

4 적정성 검토 단계

※ 외부전문가를 포함한 적정성 평가 위원회 등을 구성하여 처리 목적의 적합성, 위험성 검토 결과의 적정성, 가명처리 결과의 적정성, 목적 달성 가능성 등을 검토하는 단계

□ 비정형데이터의 특성과 처리 목적·환경 등을 고려하여 합리적인 방법·수준으로 가명처리를 수행하였는지 검토 권고

□ 비정형데이터 가명처리에 활용한 기술의 적절성·신뢰성을 검토하고, 해당 기술의 한계 등으로 인한 잔존 위험(Residual Risk)을 충분히 낮추기 위한 추가검수를 하였는지 검토

□ 비정형데이터는 가명처리 시 데이터의 특성과 관련 기술 발전 수준, 재식별 위험 등을 종합 고려하여야 하고 이를 위한 전문성이 필요하므로 적정성 검토 시 외부전문가를 과반수 이상으로 구성하여 객관성·전문성 있는 검증을 받을 것을 권고

5 안전한 관리 단계

※ 적정성 검토 이후 가명정보 활용 과정에서 재식별 가능성 등을 모니터링·관리하는 단계

□ 가명정보 특례를 활용하여 AI 학습·연구개발을 하고자 하는 자는 AI 기술·서비스의 특성을 고려하여 사전·사후적으로 발생할 수 있는 다양한 위험을 낮추기 위한 충분한 조치를 취해야 함

□ 다만, AI 개발·활용 상황에서 나타날 수 있는 다양한 위험을 사전에 완벽히 제거하는 것은 현재 기술상 불가능하므로,

- 잔존 위험(Residual Risk)을 최소화시키기 위한 노력 정도에 따라 사후관리의 이행 수준을 판단하여야 함

□ AI 서비스 운영 과정에서도 개인식별 위험 등 정보주체 권익 침해 가능성이 높아지지 않는지 지속적으로 모니터링하여야 함

- 또한, 개인식별 위험 증가 등 정보주체 권익 침해 관련 문제 발견 즉시, 해당 가명정보 처리를 중단하고 관련 위험을 제거하여야 함

비정형데이터 개인식별 위험성 검토 체크리스트

※ 해당 체크리스트는 가명처리 계획을 세우기 전에 개인식별 위험성을 검토하기 위한 것으로, 검토 결과가 "예"에 해당하더라도, 해당 위험을 낮추기 위한 적절한 가명처리 방안 적용 후 활용 가능

<table>
<tr><th colspan="2">구분</th><th colspan="2">개인식별 위험성 검토 사항</th></tr>
<tr><td rowspan="4">데이터</td><td rowspan="4">식별성</td><td colspan="2">개인 식별이 가능한 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>① 그 자체로 특정 개인을 식별할 가능성이 높은 정보가 있는가
* (예시) △ 얼굴 전체가 선명히 보이는 경우(안면, 옆면, 성형수술 전 · 후 얼굴, 거울·유리 등에 비치거나 반사된 사람의 얼굴 등) △ 명찰(이름표) △ 이미지 또는 영상에 포함된 메타데이터 상에 사람 이름이나 환자 번호 등 그 자체로 개인의 식별성이 높은 정보들이 그대로 포함된 경우 등</td><td>□ 예
□ 아니오</td></tr>
<tr><td>② 두 개 이상의 정보를 조합하여 식별가능성이 높아지는 정보가 있는가
* (예시) △ 휴대폰 사진에 일시, 장소, 촬영자 정보 등이 포함된 경우 △ 흉부 CT 영상 이미지에 환자정보를 나타내는 정보가 포함되어 있거나 환자의 특이사항에 대한 의사의 소견이 텍스트 등으로 적혀있는 경우</td><td>□ 예
□ 아니오</td></tr>
</table>

<table>
<tr><th colspan="2">구분</th><th colspan="2">개인식별 위험성 검토 사항</th></tr>
<tr><td rowspan="5">데이터</td><td rowspan="3">특이정보</td><td colspan="2">개인의 특이한 특징으로 인하여 식별 가능성이 있는 이용 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>③ 개인의 특이한 신체적·외형적 특징으로 인해 특정 개인을 식별할 가능성이 있는가
* (예시 : 이미지·영상정보) 신체적 특징, 체형, 머리스타일, (특정 위치의) 문신, 흉터 등이 특이한 경우
* (예시 : 음성정보) 발음(구개 파열음 등), 음색(목소리) 등이 특이한 경우</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td rowspan="2">특이정보</td><td>④ 개인의 특이한 행태적 특징으로 인해 특정 개인을 식별할 가능성이 있는가
* (예시 : 이미지·영상정보) 걸음걸이, 몸짓이나 행위 등이 특이한 경우
* (예시 : 음성정보) 억양(사투리, 말투), 반복 어휘, 언어적 습관 등이 특이한 경우
* (예시 : 텍스트정보) 반복 어휘, 어법, 문체, 언어적 습관 등이 특이한 경우</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td>⑤ 개인과 연관된 객체·사물 등의 특이성으로 인하여 식별가능성이 있는 이용 항목 여부
* (예시 : 이미지·영상정보) 거주하는 집, 차종(희귀 슈퍼카 등), 옷차림, 반려동물 등이 특이한 경우</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td rowspan="3">재식별 시 영향도</td><td colspan="2">재식별 시 정보주체에게 심각한 피해 또는 불이익을 초래할 수 있는 이용 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑥ 사회통념상 차별 등으로 인해 정보주체가 피해 또는 불이익을 받을 수 있는 정보가 있는가</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td>⑦ 재식별로 인하여 받는 피해 또는 불이익의 정도와 규모가 상당히 클 수 있는 정보주체에 관한 정보가 있는가
* (예시 : 음성·텍스트) 민감한 사생활 또는 질병관련 내용이 포함된 병원 진료·상담 녹취파일, 상담보고서</td><td>☐ 예
☐ 아니오</td></tr>
</table>

<table>
<tr><th colspan="2">구분</th><th colspan="2">개인식별 위험성 검토 사항</th></tr>
<tr><td rowspan="5">처리환경</td><td rowspan="5">이용 및 제공</td><td colspan="2">가명정보 활용 형태 및 이용 기관의 개인정보 보호 수준 등을 고려하여 식별가능성이 있는 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑧ 처리주체가 보유하고 있는 정보 또는 접근·입수 가능한 정보와 이용 범위 및 유형을 고려하여 식별가능한 항목이 있는가
* (예시: 이미지·영상정보) △ 의사가 분석하고자 하는 환자 X-ray 사진 데이터셋에 본인이 직접 진료했던 환자의 데이터가 포함된 경우, 해당 의사가 보유한 진료데이터·경험을 통해 특정 환자 추정 가능 △ CT 영상에 환자에 대한 메타데이터가 포함되어, 환자에 대한 배경지식을 알고 있는 자가 해당 환자 추정이 가능한 경우 등</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td>⑨ 추가정보를 삭제하지 않고 보관하는가</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td>⑩ 가명정보 제공 시 제공받는 자의 개인정보 보호 수준 및 신뢰할 수 있는 인증을 받았는가(ISMS, ISMS-P, ISO 27001 등)</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td rowspan="5">처리환경</td><td rowspan="3">처리장소</td><td colspan="2">가명정보가 관리적·기술적·물리적으로 안전한 장소에서 처리되는지 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑪ 가명정보 처리 시 다른 정보를 접근·입수할 수 있는 장소인가
* (예시) △ 누구나 접근 가능한 개방형 형태의 장소 및 네트워크인지 △ 허가된 인원만 출입할 수 있는 장소 및 네트워크인지 △ 다른 정보, 데이터 복원 기술 등에 대한 제한이 가능한 환경인지 등</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td rowspan="3">다른 정보와의 결합</td><td colspan="2">가명정보를 다른 정보와 결합하여 활용 시 식별가능성이 있는 항목 여부</td></tr>
<tr><td>검토 항목</td><td>검토 결과</td></tr>
<tr><td>⑫ 다른 정보와의 연계 분석이 예정되어 있는가</td><td>☐ 예
☐ 아니오</td></tr>
<tr><td>⑬ 처리주체가 보유하거나 접근·입수 가능한 정보 등 다른 정보와 연계 또는 결합하여 식별가능한 항목이 있는가</td><td>☐ 예
☐ 아니오</td></tr>
</table>

개인식별 위험성 검토 항목별 조치 가이드

구분		조치 가이드
데이터	식별성	▶ **(이미지·영상정보)** - 지문, 홍채 등 개인을 직접적으로 알아볼 수 있는 생체인식정보 등이 고해상도로 촬영된 경우 원칙적으로 삭제하여야 함 - 사람얼굴, 신체부위, 차량번호판 등의 개인식별가능정보가 처리 목적 상 반드시 필요한 경우 안전한 방식으로 가명처리하여 활용 가능 * 가명처리 방법으로는 이미지 필터링(블러링, 픽셀화(모자이크), 마스킹(블랙박스), 이미지 암호화, 얼굴 합성(k-same 모델 등), 인페인팅 기법 등 활용 가능 - 특히, 블러링이나 픽셀화 기술은 가명처리 수준에 따라 기술적으로 원본 복원이 가능하므로, 추가정보(이미지 필터링 알고리즘, 이미지 필터링 강도 등)는 삭제하는 것이 원칙 - 체형, 옷차림, 머리스타일, 문신 등이 특이한 경우 개인을 알아볼 수 있는 가능성이 높으므로, 이용목적 상 반드시 필요하지 않다면 삭제하고 필요 시 이미지 필터링 등 가명처리 필요 ▶ **(음성정보)** - 발화자나 혹은 대화 상대자 등의 음성 내용 중 개인을 식별할 수 있는 정보(이름, 주소, 전화번호 등)나 대화 맥락상 개인이 식별될 수 있는 정보 등은 삭제하거나 대체할 수 있는 정보를 생성하여 대체하여야 함 * 특히 음성 내용 중 민감한 담화주제(예. 성적취향, 종교, 질병)가 포함된 경우 개인이 식별되지 않도록 각별한 주의 필요 - 음성정보는 음성내용 외에도 발화자의 음성 자체를 통해서도 개인식별이 가능하므로, 규칙 기반 개인정보 단순 삭제나 혹은 대체, 음성 변형 원리에 기반한 음성 변형(transformation) 및 변환(conversion), STT(Speech To Text) 등을 통해 가명처리 필요 - 특히, 음성을 변조한 경우, 변조한 규칙을 알 경우 이를 역으로 실행하면 음성 복원이 가능하므로 추가정보(음성변조 알고리즘 등)는 삭제하는 것이 원칙 ▶ **(텍스트정보)** - 자유 형식의 텍스트, STT, 영상의 자막 등 텍스트 내용에 포함된 개인을 식별할 수 있는 정보(이름, 주소, 전화번호 등)는 삭제하거나 대체할 수 있는 정보를 생성하여 대체하여야 함 * 가명처리 방법으로는 규칙 기반 개인정보 단순 삭제나 혹은 대체, 스크러빙(scrubbing), 정규표현식, 주석달기(annotation) 등 활용 가능
	특이정보	▶ 특이정보는 그 정보만으로 개인을 식별할 수 있는 정보는 아니더라도 고유(희소)한 특성 때문에 개인을 알아볼 수 있는 가능성이 높으므로, 이용목적 상 반드시 필요하지 않다면 삭제하고 처리 목적 달성에 반드시 필요한 경우 안전하게 가명처리하여 활용
	재식별 시 영향도	▶ 사회통념상 차별, 기본권 침해 등 파급 영향이 클 수 있는 정보는 재식별 시 다른 일반정보와 다르게 개인의 피해와 더불어 사회적 파장이 있을 수 있으므로, 꼭 필요한 항목 이외에는 삭제 등 조치

구분		조치 가이드
처리환경	이용 및 제공	▶이용 및 제공의 위험성이 있는 경우 이용자와 제공자가 서로 위험성을 낮추기 위한 처리 환경에 대한 안전성 입증 관련 협의가 필요 - 예를 들어, 제공자 입장에서 내부 이용에 비하여 외부 제3자 제공의 경우가 처리환경으로 인한 식별 위험성이 높아 데이터의 가명처리 수준이 높게 책정될 수 있음. 이 경우, 가명정보 이용자가 해당 가명처리 수준으로 처리 목적 달성이 어렵다고 판단한다면, 제공자와의 협의를 통하여 처리 환경에 대한 통제 강화를 전제로 데이터의 가명처리 수준을 낮출 수 있는지 협의 가능 - 가명처리한 이후에도 계속 원본정보를 보유할 경우, 원본정보와의 대조 및 연결을 통하여 식별 위험이 높아질 수 있으며, 원본정보 내에 개인을 식별할 수 있는 메타데이터가 포함되어 있거나 그 밖에 원본정보에 대한 배경지식 등으로 인하여 식별 위험이 높아질 수 있어 각별한 주의 필요
	처리장소	▶데이터 자체의 가명처리 수준을 낮춰서 활용해야 할 경우, 물리적·관리적·기술적 조치 등 처리 장소의 안전성을 강화하여 종합적인 개인식별 위험성을 낮출 수 있음 - 가명처리 기술의 적절성·신뢰성을 담보하기 어렵거나 안전한 환경적 통제방안을 갖추기 어려운 경우, 데이터 자체에 대한 통제 강화를 통해 종합적인 식별위험성을 낮출 필요 ▶데이터 복원기술 등에 취약한 경우, 다른 정보 접근 및 복구 기술(SW)에 대한 접근·사용을 제한할 수 있는 환경을 갖출 필요가 있음
	다른 정보와의 결합	▶다른 정보와 연계·결합 예정에 있는 경우 연계·결합되는 정보와 결합하여 식별가능성이 높아지는 항목이 있는지 추가 검토 필요 ▶처리주체가 보유하거나 접근·입수 가능한 정보를 통해 식별가능한 항목 있는지 검토 - 가명정보를 제공받아 활용하게될 자가 가진 과거 유사 정보에 대한 수행 경험이나 지식 등은 가명정보를 제공하려는 자가 자체적으로 판단하기 어렵기 때문에 가명정보를 제공받아 활용하게 될 자에게 사전에 확인 및 검토 필요. 사전 검토가 어려운 경우, 가명처리의 수준을 높이는 방법 등으로 위험성을 낮춰야 함

제 3 장

가명정보 결합 및 반출

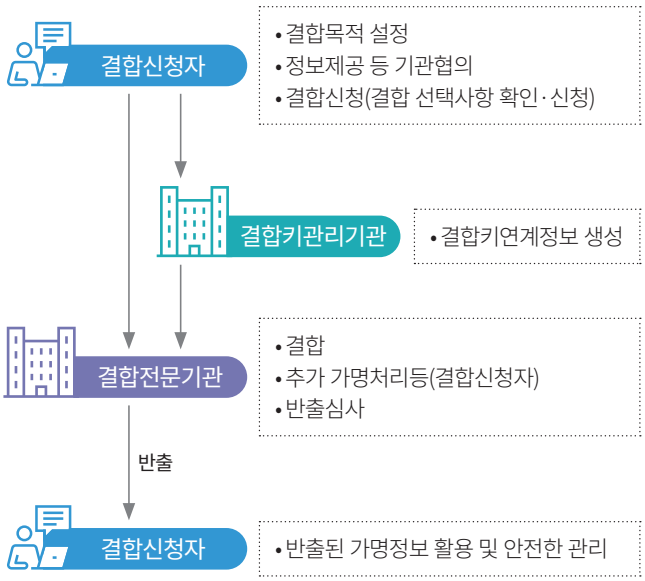
1 개요

1. 가명정보의 결합

□ 가명정보를 결합하여 활용하려는 개인정보처리자는 결합전문기관을 통해 통계작성, 과학적 연구, 공익적 기록보존 등의 목적으로 가명정보 결합이 가능함

- 서로 다른 개인정보처리자가 보유한 가명정보의 결합은 개인정보위 또는 관계 중앙행정기관의 장이 지정한 결합전문기관이 수행함(보호법 제28조의3 제1항)

□ 가명정보 결합은 ① 결합신청자의 결합신청, ② 결합키관리기관의 결합키연계정보 생성, ③ 결합전문기관의 가명정보 결합 및 반출, ④ 결합신청자의 반출정보 활용 및 관리 등으로 진행함



가명정보 결합과정에서의 결합신청자, 결합전문기관, 결합키관리기관 역할·업무

▷ 결합신청자 : 가명정보를 제공하거나 결합된 가명정보를 활용하는 개인정보처리자 등
- 결합 목적의 설정, 정보보유기관 등 결합신청자 간 사전 협의, 결합전문기관 선정 및 결합신청 등 수행
- 결합신청 시 선택사항(모의결합, 결합률 확인, 가명정보 추출) 확인 및 필요한 사항 선택·신청
- 결합전문기관의 결합 완료 후 추가 가명처리 수행 및 반출신청, 반출된 가명정보의 활용 및 안전한 관리 등 수행

* 구체적인 사항은 결합신청자의 유형(가명정보 제공 또는 결합정보 이용)에 따라 다름

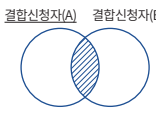
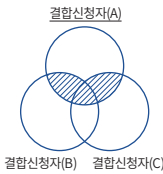
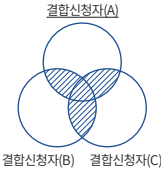
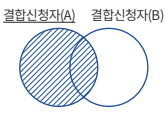
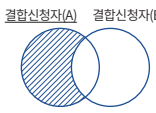
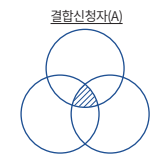
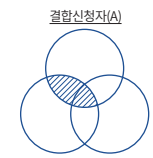
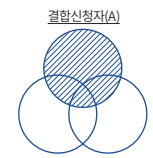
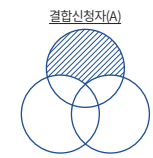
▷ 결합전문기관 : 가명정보의 결합, 반출심사 등 수행
- 결합신청자가 신청 시 선택한 모의결합, 기술지원, 결합신청자의 요청업무 등 추가 수행
- 결합 전 가명처리, 결합, 추가 가명처리 및 분석, 반출된 정보의 분석 등 지원

▷ 결합키관리기관 : 안전한 가명정보 결합지원을 위해 결합키 생성 협의 및 결합키연계정보 생성 등 수행
- 결합신청자가 신청 시 선택한 결합률 확인, 가명정보 추출 및 반복결합연결정보 생성·관리 등을 추가 수행

2. 가명정보의 결합 유형

□ 가명정보 결합은 결합신청자 간의 공통되는 결합키에 의해 이루어지며, 결합신청자가 결합 후 활용할 수 있는 정보(반출정보 등)는
- 공통된 결합키로만 결합(공통결합)된 정보,
- 각 결합신청자 기준, 공통된 결합키로 결합된 정보와 그 외 결합키의 정보로 구성(확대결합, 잔여결합)된 정보임

※ 결합신청자(A)는 결합결과로서 자신의 결합되지 않은 정보(A) 이용 가능

<table>
<tr><th colspan="3">공통결합
(INNER JOIN)</th><th>확대결합
(OUTER JOIN)</th><th>잔여결합
(ANTI-INNER JOIN)</th></tr>
<tr><th>단일
(INNER SINGLE)</th><th>다중
(INNER MULTI)</th><th>완전
(INNER FULL)</th><th>단일
(OUTER SINGLE)</th><th>단일
(ANTI-INNER SINGLE)</th></tr>
<tr>
<td> 
</td>
<td rowspan="2"> 
</td>
<td rowspan="2"> 
</td>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td> 
</td>
<td> 
</td>
<td> 
</td>
<td> 
</td>
</tr>
</table>

※ 결합 유형의 세부사항은 [참고자료] 참고3. 결합의 다양한 유형 (115p) 참고

3. 가명정보의 반복결합

□ 시계열 분석 등을 목적으로 가명정보를 결합할 때에는 동일한 서로 다른 개인정보처리자 간의 가명정보를 지속적·반복적으로 반복하며 결합 할 수 있음

- 반복결합이 필요한 경우 결합신청 시 반복결합을 선택하여 신청함

※ 반복결합의 경우 반출정보에 반복적인 분석을 위해 필요한 정보(반복결합연결정보)가 추가 포함됨

※ 시계열 분석을 위한 반복결합 절차의 구체적인 내용은 [참고자료] 참고4. 시계열 분석을 위한 반복결합 절차 (117p) 참고

2 가명정보 결합·반출 절차

☑ 가명정보 결합·반출은 ① 결합신청, ② 결합 및 추가처리, ③ 반출 및 활용, ④ 안전한 관리의 총 4단계를 거쳐 진행함

- 결합신청자는 결합신청 시 모의결합, 결합률 확인, 가명정보 추출을 선택하여 신청할 수 있음

1단계 결합신청

☑ 결합신청자는 신청자 간 결합신청에 필요한 사항*의 협의, 결합신청서 작성 등 가명정보 결합에 필요한 사전 준비사항을 확인하고 결합전문기관에 결합을 신청함

* 개인정보파일에서 가명정보 결합 목적 달성에 필요한 항목을 선정, 반복결합 여부, 모의결합/결합률 확인/가명정보 추출 신청여부, 결합키 생성항목 등

- 결합신청자는 결합전문기관과 결합일정, 전송방법 등을 협의함

2단계 결합 및 추가처리

☑ 가명정보를 제공하는 결합신청자는 결합키관리기관으로부터 결합키 생성에 이용되는 정보(Salt값)를 수신하여 결합키를 생성하고 결합신청 시 선택한 모의결합, 결합률 확인, 가명정보 추출 등이 완료되면 결합에 필요한 정보를 각 기관에 전송함

- 결합정보를 이용하는 결합신청자는 결합전문기관 공간에서 추가 가명·익명처리를 하거나, 결합전문기관이 지원하는 분석기능을 신청·이용하여 분석할 수 있음

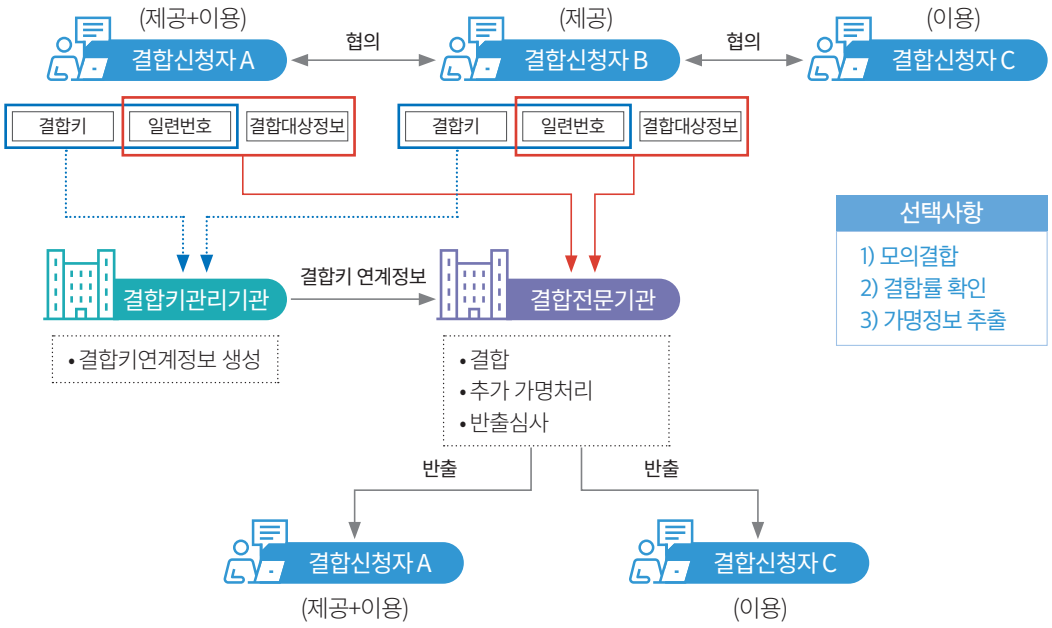
3단계 반출 및 활용

☑ 결합정보 또는 분석결과 등을 반출하려는 경우 결합전문기관에 반출을 신청함

4단계 안전한 관리

☑ 결합정보를 이용하는 결합신청자는 반출한 결합정보(이하 반출정보)를 당초 결합신청서 및 반출신청서에 기재한 목적에 따라 처리하고 안전조치 의무 등을 준수하여야 함

가명정보 결합·반출 업무 흐름도



가명정보 결합·반출 세부 절차

절차		결합신청자	결합전문기관	결합키관리기관
1 결합신청		①결합 신청	②결합신청서 검토 및 접수 ③결합 일정·절차 등 협의(결합신청자)	-
2 결합 및 추가처리	1. 결합키 생성	①결합키 생성 협의 ③결합키 및 일련번호 생성	-	②결합키 생성 협의 (Salt값 전송)
	2. 모의결합 (선택)	①결합키 전송 (→결합전문기관)	②모의결합 가능성 검토 및 통지 ③모의결합 대상 결합키 선정 및 전송	-
		④모의결합 대상 가명처리 ⑤모의결합대상정보, 가명처리 내역 및 결합키 전송	⑥가명처리 수준 검토 (필요시 추가처리 요청) ⑦모의결합 수행	-
		⑧모의결합된 정보 분석 (결합전문기관 내) * 반출 불가	⑨모의결합 관련 정보 파기	-
	3. 결합률 확인(선택)	①결합키 및 일련번호 전송 (→결합키관리기관) ④결합률 확인	-	②결합키연계정보 생성 ③결합률 측정 및 통보
	4. 가명정보 추출(선택)	①결합키 및 일련번호 전송 (→결합키관리기관) ③추출 요청	-	②추출 가능 여부 검토 및 통지 ④추출에 필요한 일련번호 선정 및 전송
	5. 가명처리 및 검토	①가명처리 대상 정보 확정 ②가명처리 ③결합대상정보, 가명처리 내역 및 일련번호 전송(→결합전문기관)	★가명처리 지원(가능한 경우) ④가명처리 수준 검토 (필요시 추가처리 요청)	-
	6. 결합	①결합키 및 일련번호 전송	③결합키연계정보 수신 및 가명정보 결합 * 반복결합의 경우 반복결합연결정보 포함	②결합키연계정보 생성 및 전송 * 반복결합의 경우 반복결합연결정보 포함
	7. 추가처리 (필요시)	①결합정보의 추가처리 및 분석(결합전문기관 내) * 결합전문기관에 지원 요청 가능	★추가처리 및 분석 지원 (가능한 경우)	-
3 반출 및 활용	1. 반출	①반출신청	②반출심사위원회 구성·운영 ③반출 승인 및 결합정보 반출 * 반복결합의 경우 반복결합연결정보 포함 ④결합키연계정보 파기	④결합키 및 결합키연계정보 파기 (필요시 유지*) * 반복결합의 경우 결합키 생성방법 (Salt값)·반복결합연결정보 생성방법(Salt값) 보관
	2. 활용	• 반출정보의 활용 원칙 준수 * 반복결합의 경우 반복결합연결정보를 통해 내부에서 연계하여 분석	-	-
4 안전한 관리		• 안전성 확보 조치 이행 • 가명정보 처리 내역 기록·보관	★반출한 정보 분석 지원 (가능한 경우) ★개인정보 보호 교육 제공 (가능한 경우)	-

★ : 가명정보의 결합 및 반출 등에 관한 고시 제11조의2에 따른 결합전문기관의 업무지원 사항으로, 결합신청자는 결합전문기관이 해당 업무에 대해 지원 가능한 경우 요청할 수 있음

※ 본 가이드라인에서는 결합신청자 기준으로 결합절차 안내

3 사전준비

1. 목적 설정 및 결합 가능정보의 탐색

□ 결합 목적 설정

- 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등 결합의 목적*을 명확히 설정하여야 함

* 목적 설정에 관한 설명은 [제2장 가명처리 및 가명정보의 처리] (9p) 참고

- 통계작성을 위한 결합이란 특정 집단이나 대상 등에 대하여 수량적인 정보를 처리하여 통계 작성을 목적으로 가명정보를 결합하는 것을 말하며 상업적 성격의 통계 작성도 가능함
- 과학적 연구를 위한 결합이란 과학적 방법을 적용한 연구로서 자연과학, 사회과학, 기초연구, 응용연구뿐만 아니라 새로운 기술·제품·서비스 개발 및 실증을 위한 산업적 연구를 포함한 과학적 연구를 목적으로 가명정보를 결합하는 것을 말함
- 공익적 기록보존을 위한 결합이란 공공의 이익을 위하여 지속적으로 열람할 가치가 있는 정보를 기록하여 보존하는 것을 의미하며 공공기관뿐 아니라 기업, 단체 등이 일반적인 공익을 위하여 기록을 보존하는 경우도 포함한 공익적 기록보존을 목적으로 가명정보를 결합하는 것을 말함

□ 결합 가능정보의 탐색

- 결합을 추진하려는 결합신청자는 결합 목적을 달성하기 위한 결합 가능정보를 탐색·확인하고 해당 정보의 보유기관과 협의*함

* 결합목적을 달성하기 위한 항목의 보유 여부, 가명정보의 제공 가능 여부 등

참고사항

▶ 개인정보위가 지원하는 '매칭지원 서비스'를 통해 가명정보 결합 시 필요한 역량(데이터, 아이디어 등)과 함께 결합을 할 수 있는 기관(기업)을 찾을 수 있음

※ 가명정보 지원 플랫폼(dataprivacy.go.kr) → 매칭지원

- 결합 아이디어(연구)는 존재하나 데이터 보유기관을 찾고자 하는 경우
- 보유한 데이터를 기반으로 가명정보의 결합을 추진 시 아이디어 및 결합대상 탐색하는 경우

※ 정보보유기관은 결합을 위한 가명정보를 제공할 의무가 없으므로, 해당 정보가 필요한 개인정보처리자 등은 정보보유기관과 협의를 진행하여야 함

2. 기관협의

□ 결합신청자는 결합 목적을 설정하고 다른 결합신청자와 가명정보 결합에 대해 협의하는 등 필요한 사항을 사전준비 함

※ 협의사항 : 개인정보파일에서 가명정보 결합 목적 달성에 필요한 항목 선정, 반복결합 여부, 모의결합/결합률 확인/가명정보 추출 신청여부, 결합키 생성항목 등

- 결합신청자 간 개인정보의 공통항목 중에서 결합키 생성에 활용할 항목을 결정함

▶ 결합키 생성 항목 정의(예시)

- 결합신청자(A) : 성명, 전화번호, 생년월일, 주소, 차량 정보, 배기량, 주유금액 등
- 결합신청자(B) : 성명, 전화번호, 생년월일, 주소, 주거형태, 보증금 유무, 월세 유무 등

⇨ 결합신청자가 동일하게 가지고 있는 **성명, 전화번호, 생년월일**을 결합키 생성 항목으로 선정

□ 결합신청자는 가명정보 결합에 관한 별도의 내부승인절차 등을 진행할 수 있으며, 결합에 대한 계약 체결 등 필요한 조치를 할 수 있음

4 결합신청

1. 결합 신청

□ 결합을 위해 가명정보를 제공하는 개인정보처리자 또는 결합정보를 이용하려는 자*(이하 결합신청자) 모두 결합신청서를 작성하여 신청하여야 함

* 현재 가명정보를 보유하고 있지 않으나, 결합된 가명정보를 처리할 예정인 자 포함

※ 결합전문기관은 결합 및 반출 등에 필요한 비용을 결합신청자(결합정보를 이용하려는 자)에게 요청할 수 있음

□ 결합신청자는 가명정보 지원 플랫폼(dataprivacy.go.kr)을 이용하여 결합전문기관*에 결합을 신청함

* 결합전문기관 현황은 가명정보 지원 플랫폼(dataprivacy.go.kr)에서 확인 가능함

▶ 결합신청자는 결합대상정보에 대한 전문성, 분석 및 가명처리에 필요한 시스템 성능, 소요일정, 가명처리 또는 분석 지원 여부, 모의결합 지원 등 결합전문기관의 지원사항 등을 고려하여 결합전문기관을 선택 할 수 있음

▶ 결합전문기관(보호법)과 데이터전문기관(신용정보법)

- 신용정보법은 신용정보회사등의 정보와 결합하고자 하는 경우 데이터전문기관을 통해 결합하도록 규정(신용정보법 제17조의2 제1항)하므로, 신용정보회사등과 결합하는 경우에는 데이터전문기관에 결합을 신청하여야 함

※ 결합대상정보의 성격이 아닌 해당 정보를 보유한 기관에 따라 결합전문기관(보호법) 또는 데이터전문기관(신용정보법)을 구분하여 결합신청 필요

- 신용정보회사등이 아닌 기관이 보유한 금융·신용정보는 보호법에 따라 결합전문기관을 통해 결합을 수행하여야 함

□ 결합신청자는 '가명정보의 결합 및 반출 등에 관한 고시(이하 '결합 고시'라 함)'의 제8조에 따른 [별지 제3호] 결합신청서와 첨부 서류*를 결합 신청 시 제출함

* 단, 결합신청자가 결합의 선택사항 진행 등의 사유로 결합대상정보에 대한 검토 및 확정을 완료하지 않은 경우 결합전문기관과 협의하여 결합대상정보의 가명처리 내역에 관한 서류는 가명정보 전송 시 제출할 수 있음

※ 가명정보 지원 플랫폼(dataprivacy.go.kr)을 통해 제출할 결합신청서 및 첨부 서류의 구체적인 작성 방법은 [참고자료] 참고6. 결합신청서 작성 방법 (122p) 참고

□ 결합신청자는 결합전문기관이 신청서 작성내용(결합 목적 적합성 등) 및 첨부 서류에 대한 보완을 요청한 경우 결합신청자는 해당사항을 보완하여 다시 제출하여야 함

- 결합전문기관은 서류 누락 등 신청서류에 더 이상 보완사항이 없는 경우 결합신청서를 접수하고 결합신청자에게 신청접수 사실을 통지함

- 결합신청을 접수한 이후에도 결합전문기관은 결합 목적, 결합대상 항목 등이 적절한지 여부를 추가로 확인하여 보완이 필요한 경우에는 결합 목적 증빙 자료 제출이나 결합대상 변경 등을 요청할 수 있음

□ 결합신청자는 결합 신청 내역에 따라 결합 절차 및 필요한 정보 등을 결합전문기관 및 결합키관리기관과 협의할 수 있음

2. 모의결합, 결합률 확인, 가명정보 추출 등의 선택 신청

□ 모의결합 : 결합 목적 달성을 위한 정보의 일부를 결합·분석하여 결합의 유용성을 확인하는 절차를 말함

□ 결합률 확인 : 결합대상정보의 가명처리에 앞서 결합률을 확인하는 절차로써 결합키와 일련번호로 결합률을 확인함

* 결합률 확인 후 결합 절차의 진행여부를 결정할 수 있으며, 결합 절차 진행 결정 이후 가명처리를 수행할 수 있음

□ 가명정보 추출 : 결합대상정보 중 결합되는 정보를 가명처리하여 전송할 수 있도록 해당 정보를 추출*하는 절차를 말함. 결합대상정보 전부를 가명처리하거나 전송하는 방식에 비하여 효율적이며 안전할 수 있음

* 개인정보 침해 우려가 없도록 결합되지 않는 정보를 일부 포함하여 추출(결합키관리기관)

신청항목	처리기관	비고
모의결합	결합전문기관	결합신청자(결합키) → 처리기관(모의결합 대상 결합키) → 결합신청자(모의결합대상정보) → 처리기관(모의결합) →결합신청자(분석)
결합률 확인	결합키관리기관	결합신청자(결합키, 일련번호) → 처리기관(결합률) → 결합신청자(결합률 확인)
가명정보 추출	결합키관리기관	결합신청자(결합키, 일련번호) → 처리기관(추출된 일련번호) → 결합신청자(가명정보 추출)

5 결합 및 추가 가명처리

1. 결합키 생성

□ 가명정보를 제공하는 결합신청자는 결합키관리기관과 결합키 생성에 관한 사항을 협의*하고 결합키관리기관으로부터 결합키 생성에 필요한 정보(Salt값)를 수신함

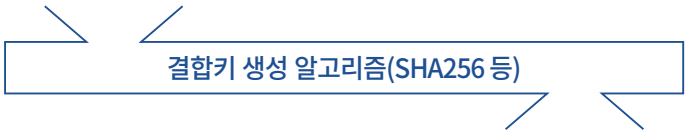
* 결합키 생성 항목, 인코딩 방식, 결합키 생성 알고리즘

- 결합키 생성시에는 결합신청자 간 결합키 생성 항목, 인코딩* 방식, 알고리즘을 동일하게 사용하여야 함

* 한글 인코딩 방식(EUC-KR, UTF-8)이 다를 경우 동일한 일방향 암호화 알고리즘으로 데이터를 암호화하여도 서로 다른 값으로 결합키가 생성되어 결합이 되지 않음(UTF-8 인코딩을 권고)

결합키 생성 예시

'홍길동'+'01012345678'+'생년월일'+'abc123'
(성명/전화번호/생년월일/Salt값)



'a9fe0r · · · sr13'

▶ 일반적으로 결합키의 대상은 성명, 전화번호, 생년월일 등 특정 개인을 식별할 수 있는 정보이며, 비정형데이터의 경우 이미지나 혹은 영상 등에 포함된 메타 데이터 내 항목(예 : 의료 DICOM (Digital Imaging and Communications in Medicine) 표준 헤더 내 환자 번호 등)도 가능

▶ 결합키 생성 알고리즘은 결합키 생성 항목으로 특정 개인을 식별할 수 없도록 일방향 암호화 알고리즘을 사용함

※ 일방향 암호화 알고리즘은 가명정보의 보호에 큰 영향을 미치게 되어 일방향 암호화 기법 중 SHA256 이상의 알고리즘(Salt값 포함)을 이용할 것을 권고함

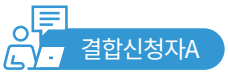
※ Salt 값의 길이는 Hash처리 결과값의 크기와 동일한 크기를 사용하는 것이 안전함

* 참고 : '개인정보의 암호화 조치 안내서(2020.12.)', 개인정보위

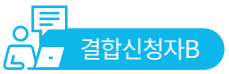
- 결합신청자는 가명처리 대상 정보에 정보주체별로 중복되지 않는 일련의 값(일련번호*)을 생성함

* 일련번호는 모의결합 시에는 활용되지 않으므로, 모의결합 절차가 종료된 이후 생성할 수 있음

일련번호 생성 예시



일련번호	성명	전화번호	생년월일	...
A1	강감찬	090-4562-7895	1947	...
A2	권율	090-7854-5689	1975	...
A3	유관순	090-4567-9876	1982	...
...	...	...	...	...



일련번호	성명	전화번호	생년월일	...
B1	유관순	090-4567-9876	1982	...
B2	권율	090-7854-5689	1975	...
B3	강감찬	090-4562-7895	1947	...
...	...	...	...	...

□ 반복결합의 경우 결합신청자는 추후 반출되는 정보와의 연계·분석을 위하여 결합키에 사용된 결합키 생성 항목, 인코딩 방식, 알고리즘(Salt값 제외*)을 보관함

* 반복결합에 사용된 Salt값은 결합키관리기관이 보관하였다가, 추후 반복결합 진행시 재안내 예정

2. 모의결합

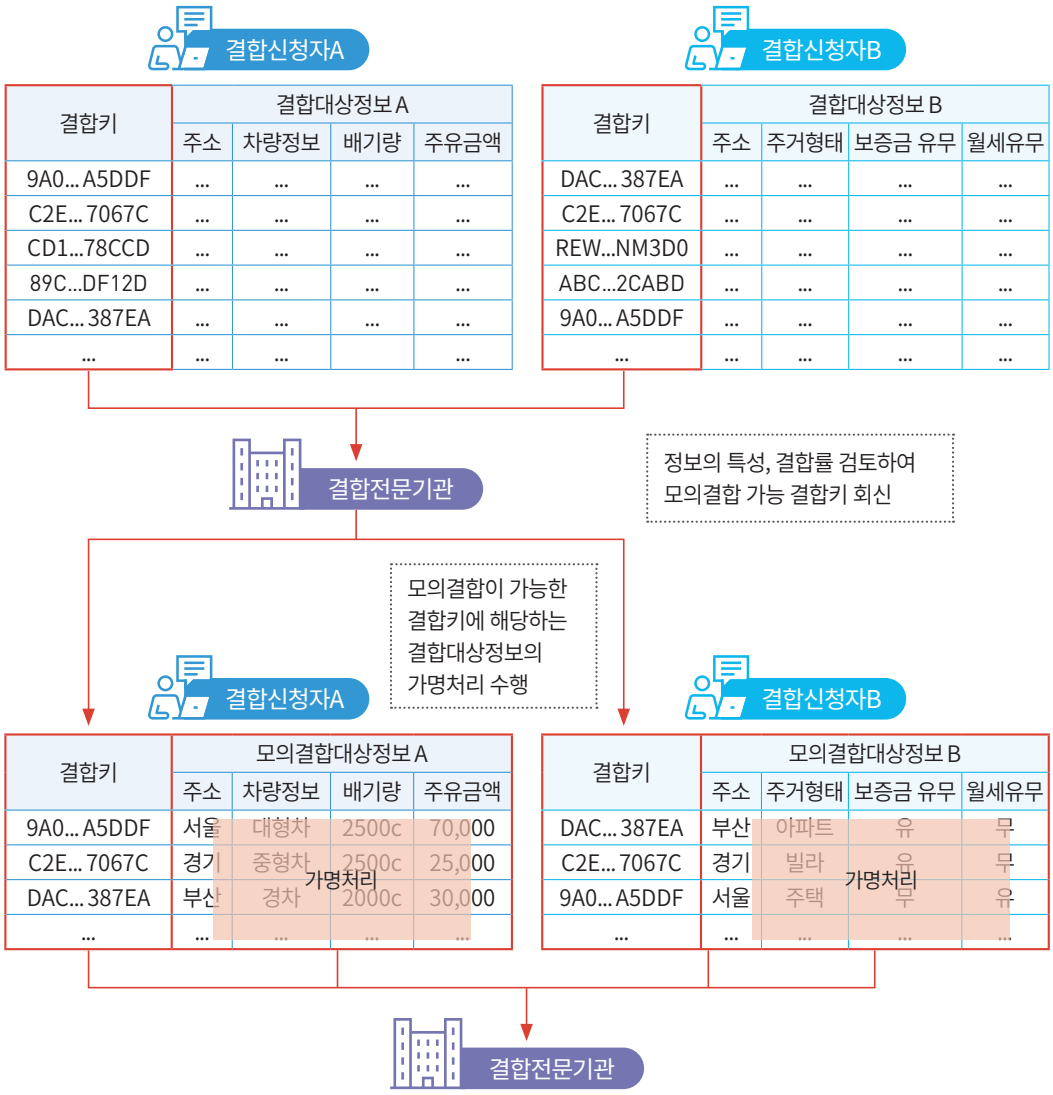
□ 가명정보를 제공하는 결합신청자는 결합키 관리기관 또는 모의 결합을 수행 할 결합전문기관과의 협의에 따라 생성한 모의결합키를 해당 결합전문기관에 전송함

※ 결합신청자가 모의결합을 위한 결합키 생성을 위해 결합키 관리기관과 협의하는 것은 선택사항

□ 결합전문기관은 모의결합대상정보의 특성, 결합률 등을 고려하여 모의결합 가능여부를 판단하고, 모의결합이 가능한 경우 모의결합대상정보(결합키)를 선정하여 해당 결합키를 결합신청자에게 전송함

※ 결합전문기관은 개인정보 침해의 우려가 없는 범위에서 결합신청자가 결합의 유용성을 확인할 수 있도록 모의결합 대상 결합키를 선정

□ 결합신청자는 결합전문기관으로부터 결합키를 제공받아 해당 모의결합대상정보를 가명처리하고 가명처리 내역과 함께 결합전문기관에 전송함



- 결합전문기관은 결합신청자의 가명처리 내역을 확인하고(보완 필요시 보완 요청) 결합키를 사용하여 모의결합대상정보의 결합을 수행함

- 결합신청자는 결합전문기관(추가 가명처리 공간 등)에서 모의결합된 정보를 분석할 수 있음

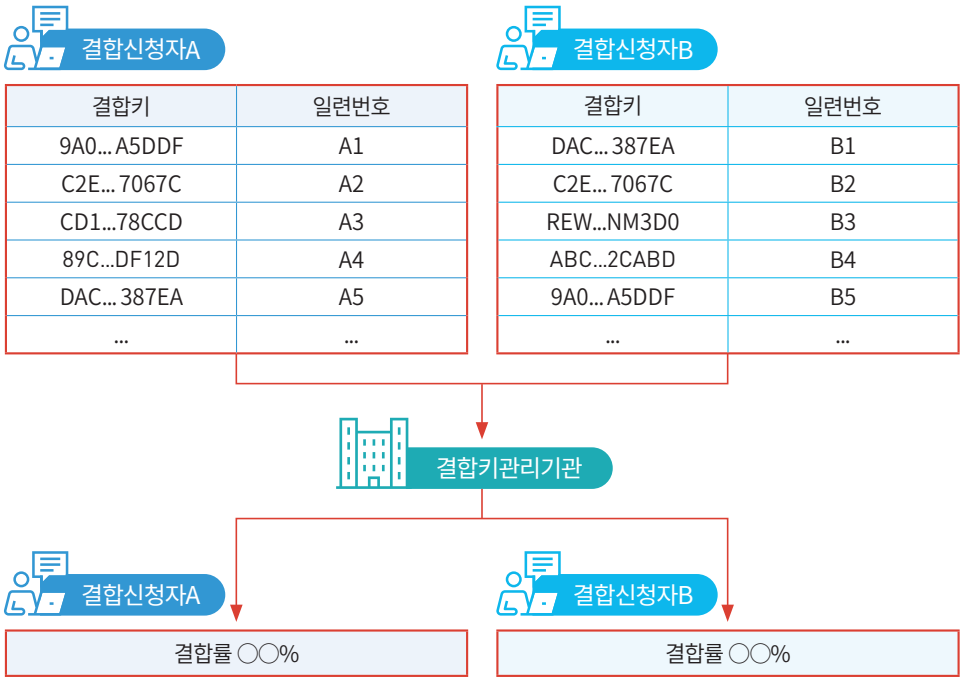
  - 결합신청자는 모의결합 분석 결과에 따라 본결합의 진행 또는 종료를 결정할 수 있음

  ※ 단, 결합신청자는 분석한 결과물 및 모의결합 정보를 반출할 수 없음

- 결합전문기관은 결합신청자의 모의결합정보 분석이 완료되면 모의결합에 사용된 정보를 파기하여야 함

3. 결합률 확인

□ 결합률 확인을 신청한 결합신청자는 결합키와 일련번호를 결합키관리기관에 전송함



□ 결합키관리기관은 결합률을 측정하며 해당 결합신청자에게 해당 정보의 결합률을 통지함

- 결합신청자는 결합률 확인 후 결합의 진행 또는 종료를 결정할 수 있음

4. 가명정보 추출

□ 가명정보 추출을 신청한 결합신청자는 결합키와 일련번호를 결합키관리기관에 전송함

□ 결합키관리기관은 추출 여부를 판단하는데 필요한 정보(결합 목적 등)를 결합신청자로부터 제공받아 추출 가능 여부를 검토하고, 추출이 가능한 경우 추출에 필요한 일련번호를 결합신청자에게 전송함

▶ 가명정보 추출의 가능여부 검토 및 대상 선정 기준

- (정보의 특성) 결합의 목적 및 데이터의 특성을 검토하여 결합되지 않는 정보의 수가 너무 적어 재식별 위험이 있는 등과 같은 정보주체에 대한 불이익 발생 가능성 등을 고려. 필요한 경우 가명정보 추출 대상 조정
- (결합률) 전국민 데이터 등 대규모 데이터의 경우 결합되는 정보에 비해 많은 양의 정보를 가명처리 및 전송해야하는 부담 등을 고려. 필요한 경우 가명정보 추출 대상 조정

▶ 가명정보 추출 원칙

1. 결합되는 정보(B)에 결합되지 않는 정보 일부(비결합대상정보, C)의 일련번호를 추가 (결합되는 정보(B)의 수와 동일)하여 추출(추출 대상 정보, D)

〈 결합신청자별 가명정보 추출 〉

전체 정보(A)	결합되는 정보(B)	비결합대상정보(C) (결합되지 않는 정보 일부)	추출 대상 정보(D=B+C) (일련번호 수)
1,000,000	100,000	100,000	200,000

2. 추출되지 않는 정보(E)의 수가 1,000개 이하인 경우 정보의 특성을 검토하여 추출되는 비결합대상정보(C)의 수를 조정(축소)

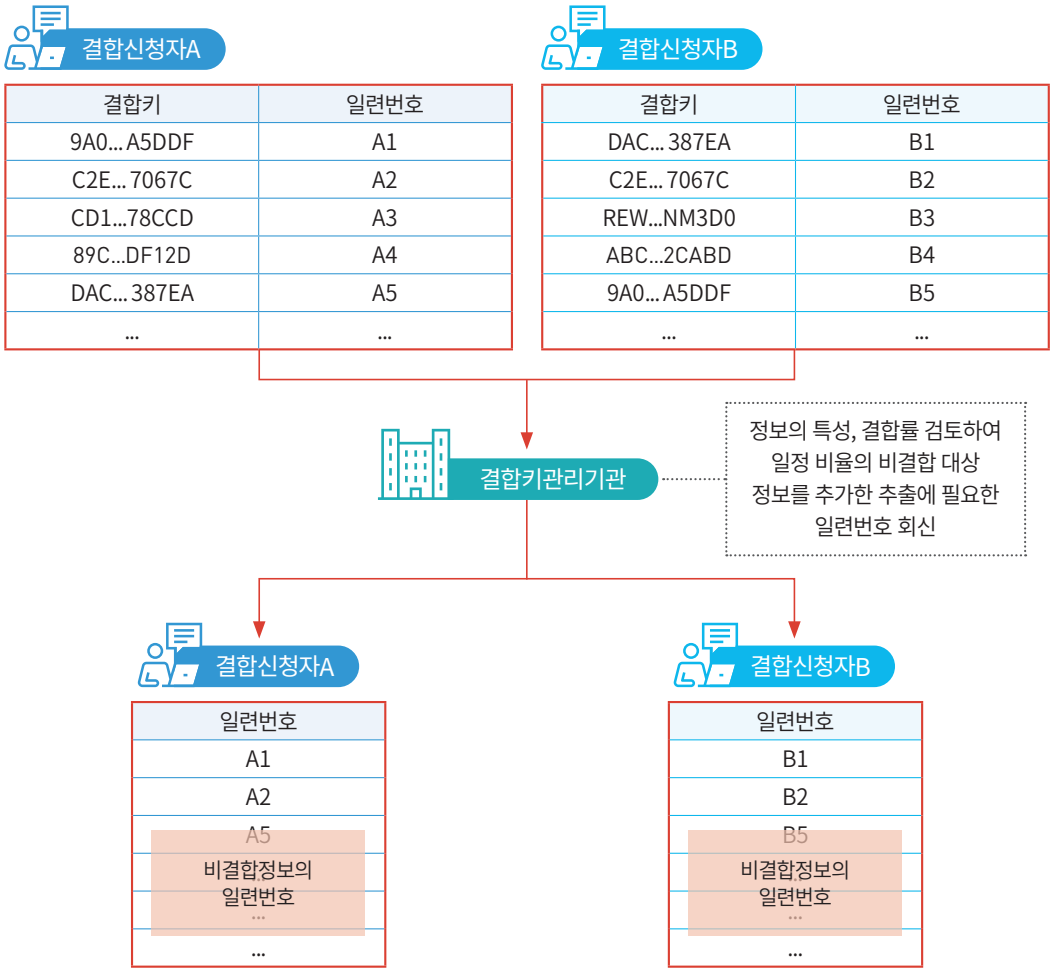
전체 정보(A)	결합되는 정보(B)	비결합대상정보(C)	추출 대상 정보 (D=B+C)	비추출 대상 정보 (E=A-D)
5,000	2,000	2,000	4,000	1,000

▼

전체 정보(A)	결합되는 정보(B)	비결합대상정보(C)	추출 대상 정보 (D=B+C)	비추출 대상 정보 (E=A-D)
5,000	2,000	1,600	3,600	1,400

3. 결합률(B/A)이 50%를 넘는 경우 정보의 특성을 고려하여 가명정보 추출 불가

전체 정보(A)	결합되는 정보(B)	비결합대상정보(C)	추출 대상 정보 (D=B+C)	비추출 대상 정보 (E=A-D)
5,000	2,700	2,300	(추출 불가)	-


결합신청자A
결합키
일련번호
9A0... A5DDF
A1
C2E... 7067C
A2
CD1...78CCD
A3
89C...DF12D
A4
DAC... 387EA
A5
...
...
결합신청자B
결합키
일련번호
DAC... 387EA
B1
C2E... 7067C
B2
REW...NM3D0
B3
ABC...2CABD
B4
9A0... A5DDF
B5
...
...
결합키관리기관
정보의 특성, 결합률 검토하여
일정 비율의 비결합 대상
정보를 추가한 추출에 필요한
일련번호 회신
결합신청자A
일련번호
A1
A2
A5
비결합정보의
일련번호
...
...
결합신청자B
일련번호
B1
B2
B5
비결합정보의
일련번호
...
...

5. 가명처리 및 검토

□ 결합을 진행하기로 결정한 결합신청자는 가명처리 대상 정보를 가명처리하여 결합대상정보를 결합전문기관에 전송함

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 가명처리 지원을 요청할 수 있음

- 가명정보 추출을 신청한 결합신청자는 결합키관리기관이 제공한 추출에 필요한 일련번호를 확인하고, 해당 일련번호의 결합대상정보의 가명처리 내역(결합대상정보 등)을 결합전문기관에 전송함

▶ 결합신청자가 보유한 개인정보 항목(예시)

- 결합신청자(A) : (성명, 전화번호, 생년월일), 주소, 차량 정보, 배기량, 주유금액 등
- 결합신청자(B) : (성명, 전화번호, 생년월일), 주소, 주거형태, 보증금 유무, 월세 유무 등

결합키 생성 항목

▶ 결합신청자별 가명처리 대상 항목

- 결합신청자(A) : 주소, 차량 정보, 배기량, 주유금액 등
- 결합신청자(B) : 주소, 주거형태, 보증금 유무, 월세 유무 등

* 가명처리 대상 중 분석목적에 필요하며, 식별 가능성이 현저히 낮은 항목인 경우 처리대상에서 제외 가능함

※ 결합키 생성 항목을 결합대상정보로 활용하고자 하는 경우 식별 가능성이 존재하지 않는 것을 확인한 후 활용하여야 함

□ 결합전문기관은 결합신청자가 제출한 결합대상정보 및 가명처리 내역을 검토함

- 보완이 필요한 경우 결합전문기관은 결합신청자에게 보완 사항을 적시하여 보완 요청함

6. 결합

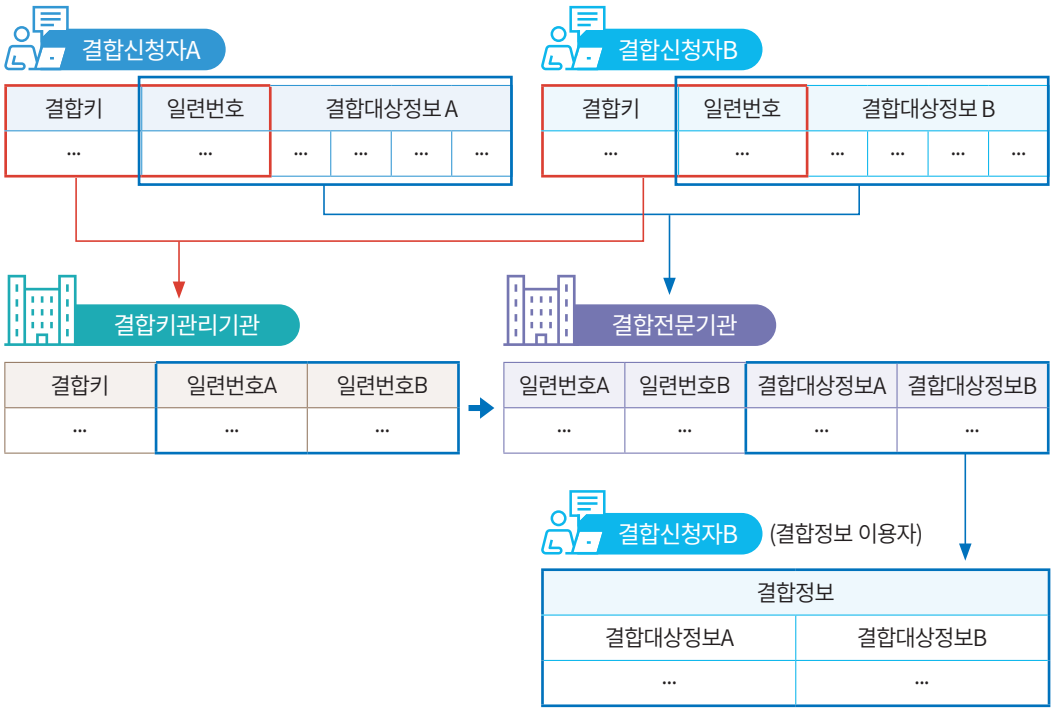
□ 결합신청자는 결합키와 일련번호를 결합키관리기관에 전송함

※ 결합률 확인 또는 가명정보 추출을 위해 결합키와 일련번호를 기 전송하고 결합대상정보의 변경이 없는 경우 전송을 생략할 수 있음

□ 결합키관리기관은 결합키와 일련번호를 사용하여 결합키연계정보를 생성하고 결합전문기관에 결합키연계정보를 전송함

※ 반복결합의 경우 반복결합연결정보를 포함하여 결합키연계정보를 생성함

□ 결합전문기관은 결합키연계정보와 일련번호, 결합대상정보를 사용하여 결합함



※ 필요시 추가처리 및 분석 수행, 반출심사 후 반출 가능

7. 추가처리 및 분석

□ 결합정보를 이용하는 결합신청자는 결합전문기관(추가 가명처리 등 공간)에서 결합정보의 식별 위험성(가능성)을 확인하고, 보완이 필요한 경우 해당 부분에 대한 추가 가명처리를 수행함

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)으로부터 추가 가명처리에 대한 자문 및 지원을 받을 수 있음

- 결합신청자는 결합정보의 재식별 가능성이 없거나 추가 가명처리가 필요하지 않다고 판단하는 경우 추가 가명처리 없이 반출신청이 가능함

□ 결합신청자는 결합전문기관에 마련된 분석에 필요한 시설, 장비를 갖춘 공간(추가 가명처리 등 공간)에서 결합정보를 분석할 수 있음

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 결합정보의 분석지원을 요청할 수 있음

6 반출 및 활용

1. 반출신청 및 반출심사

□ 결합정보를 반출하려는 결합신청자는 결합 고시 [별지 제4호] 반출신청서와 첨부 서류*를 제출하며 반출을 신청함

* 추가적인 서류 제출이 필요한 경우에 한하여 추가 처리 내역, 반출정보를 증명할 수 있는 서류, 반출정보에 대한 안전조치 계획을 제출

※ 반출신청서 및 첨부 서류의 구체적인 작성 방법은 [참고자료] 참고7. 반출신청서 작성 방법 (125p) 참고

□ 반출신청을 받은 결합전문기관은 반출신청서 및 첨부 서류를 확인하고 접수함

- 보완이 필요한 경우 결합전문기관은 해당 사유를 적시하여 결합신청자에게 보완 요청하고 보완사항을 확인한 이후 접수함

□ 결합신청자가 반출을 요청하면 결합전문기관은 접수일로부터 영업일 기준 5일 이내 반출심사위원회 구성 등에 관한 사항을 결합신청자에게 통지함

- 결합신청자는 결합전문기관으로부터 회의개최 일정 및 장소, 반출가능 예정 시기 등이 포함된 계획서를 받을 수 있음
- 시계열 분석 등 반복결합의 반출심사는 2회부터 최초(첫회) 반출과 결합대상, 가명처리 방법 등이 거의 동일한 경우 서면회의 등으로 간소화할 수 있음
- 반출심사위원회는 3명의 위원으로 구성함. 단 반출심사를 위해 필요한 경우 다른 결합전문기관에 소속된 전문가를 추가로 포함하여 구성할 수 있음
- 반출심사위원은 개인정보 보호와 관련한 업무 경력이 있거나 관련 단체로부터 추천을 받은 사람, 개인정보처리자로 구성된 단체에서 활동한 경력이 있거나 관련 단체로부터 추천을 받은 사람, 그 밖에 개인정보 보호와 관련한 경력과 전문성이 있는 사람이어야 함(결합 고시 제11조제2항)
- 반출심사위원회는 결합 목적과 반출정보의 관련성, 특정 개인의 식별가능성, 반출정보에 대한 안전조치 계획 등을 심사하여야 함(보호법 시행령 제29조의3제4항, 결합 고시 제11조제3항)

□ 결합신청자는 반출심사위원회의 요청에 따라 추가 서류를 제출하거나 직접 출석하여 설명할 수 있음

2. 반출

□ 결합전문기관이 반출을 승인하면 결합신청자는 결합정보를 분석한 결과물을 반출하거나, 결합정보(데이터셋)를 반출할 수 있음

3. 활용

□ 반출정보는 결합신청자가 반출심사 시 제출한 환경(가명정보 활용 형태, 처리 장소, 방법)과 목적범위에서 활용하는 것이 원칙임

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 반출정보에 대한 분석 지원을 요청할 수 있음

- 결합신청자가 반출정보를 반출심사 시와 다른 목적으로 활용하거나 제3자에게 제공하는 것이 금지되어 있지는 않으나(보호법 제28조의2 제1항), 반출심사 시 제출한 처리 상황의 변경이 있는 경우 해당 처리 상황에 맞게 가명처리하여 활용하여야 함

□ 반복결합의 반출정보에는 반복결합연결정보가 포함되어 내부에서 연계하여 분석할 수 있음

7 안전한 관리

1. 안전한 관리

□ 결합신청자는 반출정보를 특정 개인을 알아보기 위한 목적으로 처리하여서는 아니 되며(보호법 제28조의5 제1항), 재식별되지 않도록 지속적으로 모니터링하여야 함

□ 반출정보를 활용하는 결합신청자는 안전성 확보에 필요한 기술적·관리적·물리적 조치를 수행하여야 함

※ 결합신청자는 결합전문기관(지원 가능 기관에 한함)에 개인정보 보호 교육에 관한 지원을 요청할 수 있음

※ 안전조치에 관한 세부사항은 [제4장 안전성 확보 조치] (77p) 참고

2. 결합전문기관 업무지원 사항

▶ 결합전문기관은 보호법에서 규정한 가명정보 결합·반출 업무를 수행하여야 하며, 기관의 상황에 따라 가명처리 컨설팅, 분석 지원 등 가명정보 처리에 대한 전문성 있는 기관으로의 역할을 수행할 수 있음

□ (모의결합) 결합 전 모의결합 절차를 수행*할 수 있음

* 모의결합 가능성 검토 및 통지, 모의결합대상정보 가명처리 수준 검토(필요시 추가처리 요청), 모의결합 수행

□ (결합 전 처리) 결합 전 결합대상정보의 가명처리를 지원할 수 있음

□ (반출 전 처리) 반출 전 결합정보의 추가 가명처리를 지원할 수 있음

□ (분석) 반출 전 결합정보의 분석 및 반출 후 반출정보의 분석을 지원할 수 있음

□ (교육) 가명정보를 반출하려는 결합신청자에 대한 개인정보 보호 교육*을 지원할 수 있음

* 반출정보의 안전조치에 관한 교육, 가명처리 지원 제도 안내 등

제 4 장

안전성 확보 조치

1 관리적 보호조치

◉ 개인정보처리자는 가명정보 또는 추가정보의 안전한 관리를 위하여 내부 관리계획의 수립, 수탁자 관리·감독 등의 관리적 보호조치를 하여야 함

1. 개인정보처리자는 가명정보 및 추가정보를 안전하게 관리하기 위한 내부 관리계획을 수립·시행하여야 함(보호법 시행령 제29조의5 제1항 제1호)

※ 다만 개인정보 개념에 가명정보 개념이 포함되므로, 개인정보의 안전한 관리를 위하여 수립·시행된 내부 관리계획이 있을 경우 가명정보의 처리에 관한 내용만 추가하여 수립·시행하는 것도 가능

제29조의5(가명정보에 대한 안전성 확보 조치) ① 개인정보처리자는 법 제28조의4 제1항에 따라 가명정보를 원래의 상태로 복원하기 위한 추가 정보(이하 이 조에서 "추가정보"라 한다)에 대하여 다음 각 호의 안전성 확보 조치를 해야 한다.

1. 제30조에 따른 안전성 확보조치
2. 가명정보와 추가정보의 분리 보관. 다만, 추가정보가 불필요한 경우에는 추가정보를 파기해야 한다.
3. 가명정보와 추가정보에 대한 접근 권한의 분리. 다만, 「소상공인기본법」 제2조에 따른 소상공인으로서 가명정보를 취급할 자를 추가로 둘 여력이 없는 경우 등 접근 권한의 분리가 어려운 정당한 사유가 있는 경우에는 업무 수행에 필요한 최소한의 접근 권한만 부여하고 접근 권한의 보유 현황을 기록으로 보관하는 등 접근 권한을 관리·통제해야 한다.

② 법 제28조의4제3항에서 "대통령령으로 정하는 사항"이란 다음 각 호의 사항을 말한다.

1. 가명정보 처리의 목적
2. 가명처리한 개인정보의 항목
3. 가명정보의 이용내역

4. 제3자 제공 시 제공받는 자
5. 가명정보의 처리 기간(법 제28조의4제2항)에 따라 가명정보의 처리 기간을 별도로 정한 경우로 한정한다)
6. 그 밖에 가명정보의 처리 내용을 관리하기 위하여 보호위원회가 필요하다고 인정하여 고시하는 사항

☑ 내부 관리계획에는 추가정보의 별도 분리 보관 및 이에 대한 접근권한 분리에 대한 사항 등을 포함하여야 함

가명정보 처리 내부 관리계획에 포함될 사항(예시)

가. 가명정보 및 추가정보의 분리 보관에 관한 사항
나. 가명정보 및 추가정보에 대한 접근권한 분리에 관한 사항
다. 가명정보 또는 추가정보의 안전성 확보조치에 관한 사항
라. 가명정보를 처리하는 자의 교육에 관한 사항
마. 가명정보 처리 기록 작성 및 보관에 관한 사항
바. 개인정보 처리방침 공개에 관한 사항
사. 가명정보의 재식별 금지에 관한 사항
아. 가명정보의 처리기간을 별도로 정한 경우에 관한 사항
※ 상기 내용에 포함되지 않은 항목은 '개인정보의 안전성 확보조치 기준 해설서' 참조

※ 가명정보 처리 내부 관리계획 작성 예시는 [참고자료] 참고8. 내부 관리계획 작성 예시 (127p) 참고

☑ 개인정보처리자는 내부 관리계획에서 정한 사항에 중요한 변경이 있는 경우 이를 즉시 반영하여 내부 관리계획을 수정·시행하고, 관리책임자는 연 1회 이상 내부 관리계획의 이행 실태를 점검·관리하여야 함

2. 수탁자 관리·감독의 의무(보호법 제26조)

☑ 개인정보처리자는 가명정보 처리업무를 외부에 위탁하는 경우 가명정보도 개인정보에 해당하므로 보호법 제26조에 따라 위탁업무 수행 목적 외 가명정보의 처리 금지에 관한 사항 등을 포함한 문서를 작성하여야 함

☑ 또한 위탁자는 위탁하는 업무의 내용과 가명정보 처리업무를 위탁받아 처리하는 자를 공개하여야하며, 업무 위탁으로 인하여 가명정보가 분실·도난·유출·위조·변조·훼손 또는 재식별되지 아니하도록 수탁자를 교육하고, 처리현황 점검 등 수탁자가 가명정보를 안전하게 처리하는지를 감독하여야 함

가명정보 처리업무 위탁계약서에 포함되어야 할 사항(예시)

구분	위탁계약서에 포함되어야 할 사항
위탁업무 수행 목적 외 처리금지	가명정보를 위탁받은 범위 외로 처리하는 것을 금지하는 사항
가명정보의 안전조치 사항	가명정보와 추가정보의 분리 보관, 가명정보와 추가정보에 대한 접근권한 분리, 가명정보에 대한 안전조치 등에 대한 사항
위탁업무의 목적 및 범위	가명정보를 위탁하는 목적과 범위에 대한 사항
재위탁 제한	재위탁 가능한 범위에 대한 사항
관리·감독에 관한 사항	위탁업무와 관련하여 보유하고 있는 개인정보, 가명정보, 추가정보 등에 대한 안전성 확보조치에 관한 관리·감독사항
재식별 금지	가명정보를 제공받거나 처리를 위탁 받은 사업자 등은 다른 정보와 결합을 통해 재식별 시도가 금지됨을 명시
재식별 위험 발생시 통지	가명정보가 재식별 되었거나, 재식별 가능성이 높아지는 상황이 발생한 경우에는 가명정보 처리 중지 및 위탁자에게 통지 의무 명시

가명정보 처리업무 위탁계약서 특수조건 반영 사례(예시)

제○○조(재식별 금지)

① □은 △으로부터 제공받은 가명정보를 ××한 목적으로 안전하게 이용하고, 이를 이용해서 개인을 재식별하기 위한 어떠한 행위도 하여서는 아니 된다.

② □은 △으로부터 제공받은 정보가 재식별 되거나 재식별 가능성이 현저하게 높아지는 상황이 발생하면 즉시 해당 정보의 처리를 중단하고 관련 사항을 △에게 알리며, 필요한 협조를 하여야 한다.

③ □은 제1항에서 제2항까지의 사항을 이행하지 않아 발생하는 모든 결과에 대해 형사 및 민사상 책임을 진다.

※ 가명정보를 제공받은 기업은 "□", 제공한 기업은 "△"로 표시

3. 개인정보 처리방침 수립 및 공개(보호법 제30조)

☑ 개인정보처리자는 가명정보 처리와 관련하여 아래와 같은 내용을 개인정보 처리방침에 포함하여 공개하여야 함

※ 다만 개인정보의 처리에 대하여 기 작성한 개인정보 처리방침이 있을 경우 가명정보 처리에 관한 내용만 추가 가능

가명정보 활용 관련 개인정보 처리방침에 포함될 사항(예시)

1. 가명정보 처리 목적
2. 가명정보 처리 기간(선택)
3. 가명정보 제3자 제공에 관한 사항(해당되는 경우)
4. 가명정보 처리의 위탁에 관한 사항(해당되는 경우)
5. 처리하는 개인정보의 항목
6. 보호법 제28조의4(가명정보에 대한 안전조치의무 등)에 따른 가명정보의 안전성 확보 조치에 관한 사항

가명정보 활용 관련 개인정보 처리방침 반영 사례(예시)

제○○조(가명정보의 처리)

① □□□(개인정보처리자명)는 수집한 개인정보를 특정 개인을 알아볼 수 없도록 가명 처리하여 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 처리할 수 있습니다. 가명정보 처리의 위탁 및 제3자 제공은 하지 않으며, 가명정보는 재식별 되지 않도록 분리하여 별도 저장·관리하고 가명정보의 처리 내용에 대해 기록을 작성하여 보관하는 등 필요한 기술적·관리적 보호조치를 취합니다.

구분	수집·이용 목적	처리항목	보유 및 이용기간
△△△ 연구	연령대별 △△ 등 분석	휴대전화번호, △△일시, △△유형	결합데이터 분석 완료시까지

2 기술적 보호조치

◉ 개인정보처리자는 가명정보 및 추가정보의 분리 보관, 접근권한 관리, 접근통제 및 접속기록의 보관 및 점검 등의 기술적 보호조치를 하여야 함

1. 추가정보의 분리 보관(보호법 시행령 제29조의5 제1항 제2호)

☑ 개인정보처리자는 추가정보를 가명정보와 분리하여 별도로 저장·관리하고, 추가정보가 가명정보와 불법적으로 결합되어 재식별에 악용되지 않도록 접근권한을 최소화하고 접근통제를 강화하는 등 필요한 조치를 적용하여야 함

- 추가정보와 가명정보는 분리하여 보관하는 것을 원칙으로 하고, 불가피한 사유로 물리적인 분리가 어려운 경우 DB 테이블 분리 등 논리적으로 분리*하는 것도 가능 함

* 논리적으로 분리할 경우 엄격한 접근통제를 적용하여야 함

※ 추가정보의 활용 목적 달성 및 불필요한 경우에는 추가정보를 파기할 수 있으며, 이 경우 파기에 대한 기록을 작성하고 보관할 필요가 있음

2. 접근권한의 분리(보호법 시행령 제29조의5 제1항 제3호)

☑ 개인정보처리자는 가명정보 또는 추가정보에 접근할 수 있는 담당자를 가명정보 처리 업무 목적달성에 필요한 최소한의 인원으로 엄격하게 통제하여야 하며, 접근권한도 업무에 따라 차등부여 하여야 함

- 가명정보를 취급할 자를 추가로 둘 여력이 없는 경우 등 접근권한의 분리가 어려운 정당한 사유가 있는 경우*에는 업무 수행에 필요한 최소한 접근권한 부여 및 접근권한의 보유 현황을 기록으로 보관하는 등 접근권한을 관리·통제하여야 함

* 「소상공인 보호 및 지원에 관한 법률」 제2조에 따른 소상공인 등

- 가명정보를 처리하는 자가 가명처리를 수행하는 경우를 제외하고는 특정 개인을 알아볼 수 있는 개인정보처리시스템(가명정보처리시스템 제외)에 접근할 수 없도록 제한할 필요가 있음

☑ 전보 또는 퇴직 등 인사이동이 발생하여 가명정보를 처리하는 자가 변경되었을 경우 지체 없이 가명정보처리시스템의 접근권한을 변경 또는 말소하여야 함

☑ 가명정보처리시스템의 접근권한 부여, 변경 또는 말소에 대한 내역을 기록하고, 그 기록을 최소 3년간 보관하여야 함

☑ 가명정보처리시스템에 접속할 수 있는 사용자 계정을 발급하는 경우 가명정보를 처리하는 자 별로 사용자 계정을 발급하여야 하며, 다른 가명정보를 처리하는 자, 추가정보를 처리하는 자, 해당 가명정보 이외의 다른 개인정보취급자와 공유되지 않도록 하여야 함

☑ 가명정보를 처리하는 자가 안전한 비밀번호를 설정하여 이행할 수 있도록 비밀번호 작성규칙을 수립하여 적용하여야 함

☑ 가명정보에 대한 처리 권한이 있는 자만이 가명정보처리시스템에 접근할 수 있도록 계정정보 또는 비밀번호를 일정 횟수 이상 잘못 입력한 경우 접근을 제한하는 등 필요한 기술적 조치를 하여야 함

3. 가명정보 처리 관련 기록 작성·보관(보호법 시행령 제29조의5 제2항)

☑ 개인정보처리자는 가명정보의 처리목적, 가명처리한 개인정보 항목, 가명정보의 이용내역, 제3자 제공 시 제공받는 자를 작성하여 보관하여야 함

작성방법 예시

가명정보 처리 관리 대장

구분	내용
이용신청 접수번호	
가명정보의 처리 목적	가명정보의 처리 목적을 기재 (통계작성, 과학적 연구, 공익적 기록보존 등)
가명처리한 개인정보의 항목	가명처리의 대상이 된 이용 항목을 말함 (예: 성별, 나이, 주소 등)
가명정보의 이용내역	① **책임자** : 가명정보 처리 관련 책임자 ② **가명정보 및 추가정보를 처리하는자** : 가명정보를 처리하는 자 또는 추가정보를 처리하는 자(필요시 처리자 명단) ③ **가명처리 일시** : 가명처리한 일시 ④ **이용방법** : 목적외 이용, 내부이용, 외부제공, 내부 결합, 결합전문기관을 통한 결합 등
제공받는 자 (제3자 제공시)	(제3자에게 제공하는 경우) 가명정보를 제공받는 자의 명칭
관련 파일명	
가명정보 이용기간	년 월 일~ 년 월 일
가명정보 파기일자	년 월 일
대장 기록자	(인)
기록 확인자	(인)

3 물리적 보호조치

⊙ 개인정보처리자는 가명정보 또는 추가정보의 안전한 관리를 위하여 물리적 안전조치를 취하여야 함

☑ 개인정보처리자는 가명정보 또는 추가정보를 전산실이나 자료보관실에 보관하는 경우 비인가자의 접근으로부터 보호하기 위하여 출입 통제 등의 절차를 수립하여야 함

☑ 또한 가명정보 또는 추가정보가 보조저장매체 등에 저장되어 있는 경우 잠금장치가 있는 안전한 장소에 보관하여야 하며, 이러한 보조저장매체 등의 반·출입 통제를 위한 보안대책을 마련하여야 함

4 정보주체의 권리보장

⊙ 개인정보처리자는 보호법 제37조에 따라 정보주체가 자신의 개인정보에 대한 가명처리 정지를 요구하는 경우 이를 보장하여야 함

☑ 개인정보처리자는 정보주체의 가명처리 정지를 요구 받았을 때에는 지체 없이 해당 정보주체의 개인정보 처리의 전부 또는 일부를 정지하여야함

- 다만 이미 해당주체의 개인정보가 가명처리된 경우에는 가명처리 정지 요구가 적용되지 않으며, 해당 정보주체의 개인정보에 대해서는 향후 통계작성, 과학적 연구, 공익적 기록보존 등 목적으로 가명처리가 이루어지지 않도록 처리하여야 함

※ 가명정보는 특정 개인을 알아볼 수 없는 정보로 현행법상 재식별이 불가하며, 이에 따라 해당 정보주체의 개인정보가 가명처리 되었는지 여부를 확인할 수 없음(보호법 제28조의5 제1항)

제28조의5 (가명정보 처리 시 금지의무 등) ① 누구든지 특정 개인을 알아보기 위한 목적으로 가명정보를 처리해서는 아니 된다.

부록 1

참고자료

참고1 정형데이터 가명처리 기술 및 예시

1 개인정보의 가명처리 기술 종류

※ 아래 분류는 이해를 돕기 위해 ISO/IEC 20889, 그리고 EU ENISA에서 발간한 보고서 등 국내·외 자료들을 참고하여 작성했으며 표준이 아님

분류	기술	세부기술	설명
개인정보 삭제	삭제기술	삭제 (Suppression)	• 원본정보에서 개인정보를 단순 삭제
		부분삭제 (Partial suppression)	• 개인정보 전체를 삭제하는 방식이 아니라 일부를 삭제
		행 항목 삭제 (Record suppression)	• 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제
		로컬 삭제 (Local suppression)	• 특이정보를 해당 행 항목에서 삭제
개인정보 일부 또는 전부 대체	삭제기술	마스킹 (Masking)	• 특정 항목의 일부 또는 전부를 공백 또는 문자(‘*’, ‘_’ 등이나 전각 기호)로 대체
	통계도구	총계처리 (Aggregation)	• 평균값, 최댓값, 최소값, 최빈값, 중간값 등으로 처리
		부분총계 (Micro aggregation)	• 정보집합물 내 하나 또는 그 이상의 행 항목에 해당하는 특정 열 항목을 총계처리. 즉, 다른 정보에 비하여 오차 범위가 큰 항목을 평균값 등으로 대체
	일반화 (범주화) 기술	일반 라운딩 (Rounding)	• 올림, 내림, 반올림 등의 기준을 적용하여 집계 처리하는 방법으로, 일반적으로 세세한 정보보다는 전체 통계정보가 필요한 경우 많이 사용
		랜덤 라운딩 (Random rounding)	• 수치 데이터를 임의의 수인 자리 수, 실제 수 기준으로 올림(round up) 또는 내림(round down)하는 기법
		제어 라운딩 (Controlled rounding)	• 라운딩 적용 시 값의 변경에 따라 행이나 열의 합이 원본의 행이나 열의 합과 일치하지 않는 단점을 해결하기 위해 원본과 결과가 동일하도록 라운딩을 적용하는 기법
		상하단코딩 (Top and bottom coding)	• 정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 식별성을 가질 수 있음 • 이를 해결하기 위해 적은 수의 분포를 가진 양 끝단의 정보를 범주화 등의 기법을 적용하여 식별성을 낮추는 기법

1) EU ENISA(European Union Agency for Network and Information Security), Recommendations on shaping technology according to GDPR provisions, An overview on data pseudonymisatin, November 2018

EU ENISA(European Union Agency for Network and Information Security), Pseudonymisation and best practices, November 2019

분류	기술	세부기술	설명
개인정보 일부 또는 전부 대체	일반화 (범주화) 기술	로컬 일반화 (Local generalization)	• 전체 정보집합물 중 특정 열 항목(들)에서 특이한 값을 가지거나 분포상 의 특이성으로 인해 식별성이 높아지는 경우 해당 부분만 일반화를 적용하여 식별성을 낮추는 기법
		범위 방법 (Data range)	• 수치 데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 또는 구간(interval)으로 표현
		문자데이터 범주화 (Categorization of character data)	• 문자로 저장된 정보에 대해 보다 상위의 개념으로 범주화하는 기법
	암호화	양방향 암호화 (Two-way encryption)	• 특정 정보에 대해 암호화와 암호화된 정보에 대한 복호화가 가능한 암호화 기법 • 암호화 및 복호화에 동일 비밀키로 암호화하는 대칭키(Symmetric key) 방식과 공개키와 개인키를 이용하는 비대칭키(Asymmetric key) 방식으로 구분
		일방향 암호화-암호학적 해시함수 (One-way encryption-Cryptographic hash function)	• 원문에 대한 암호화의 적용만 가능하고 암호문에 대한 복호화 적용이 불가능한 암호화 기법 • 키가 없는 해시함수(MDC, Message Digest Code), 솔트(Salt)가 있는 해시함수, 키가 있는 해시함수(MAC, Message Authentication Code)로 구분 • 암호화(해시처리)된 값에 대한 복호화가 불가능하고, 동일한 해시 값과 매핑(mapping)되는 2개의 고유한 서로 다른 입력 값을 찾는 것이 계산상 불가능하여 충돌 가능성이 매우 적음
		순서보존 암호화 (Order-preserving encryption)	• 원본정보의 순서와 암호값의 순서가 동일하게 유지되는 암호화 방식 • 암호화된 상태에서도 원본정보의 순서가 유지되어 값들 간의 크기에 대한 비교 분석이 필요한 경우 안전한 분석이 가능
		형태보존 암호화 (Format-preserving encryption)	• 원본 정보의 형태와 암호화된 값의 형태가 동일하게 유지되는 암호화 방식 • 원본 정보와 동일한 크기와 구성 형태를 가지기 때문에 일반적인 암호화가 가지고 있는 저장 공간의 스키마 변경 이슈가 없어 저장 공간의 비용 증가를 해결할 수 있음 • 암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용
		동형 암호화 (Homomorphic encryption)	• 암호화된 상태에서의 연산이 가능한 암호화 방식으로 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용가능 • 암호화된 상태의 연산값을 복호화 하면 원래의 값을 연산한 것과 동일한 결과를 얻을 수 있는 4세대 암호화 기법
		다형성 암호화 (Polymorphic encryption)	• 가명정보의 부정한 결합을 차단하기 위해 각 도메인별로 서로 다른 가명처리 방법을 사용하여 정보를 제공하는 방법 • 정보 제공 시 서로 다른 방식의 암호화된 가명처리를 적용함에 따라 도메인별로 다른 가명정보를 가지게 됨
	무작위화 기술	잡음 추가 (Noise addition)	• 개인정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법

분류	기술	세부기술	설명
개인정보 일부 또는 전부 대체	무작위화 기술	순열(치환) (Permutation)	• 분석 시 가치가 적고 식별성이 높은 열 항목에 대해 대상 열 항목의 모든 값을 열 항목 내에서 무작위로 순서를 변경하여 식별성을 낮추는 기법 • 개인정보를 다른 행 항목의 정보와 무작위로 순서를 변경하여 전체정보에 대한 변경 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법
		토큰화 (Tokenisation)	• 개인을 식별할 수 있는 정보를 토큰으로 변환 후 대체함으로써 개인정보를 직접 사용하여 발생하는 식별 위험을 제거하여 개인정보를 보호하는 기술 • 토큰 생성 시 적용하는 기술은 의사난수생성 기법이나 양방향 암호화, 형태보존 암호화 기법을 주로 사용
		(의사)난수생성기 ((P)RNG, (Pseudo) Random Number Generator)	• 주어진 입력값에 대해 예측이 불가능하고 패턴이 없는 값을 생성하는 메커니즘으로 임의의 숫자를 개인정보와 대체
가명·익명처리를 위한 다양한 기술 (기타 기술)		표본추출 (Sampling)	• 데이터 주체별로 전체 모집단이 아닌 표본에 대해 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법
		해부화 (Anatomization)	• 기존 하나의 데이터셋(테이블)을 식별성이 있는 정보집합물과 식별성이 없는 정보집합물로 구성된 2개의 데이터셋으로 분리하는 기술
		재현데이터 (Synthetic data)	• 원본과 최대한 유사한 통계적 성질을 보이는 가상의 데이터를 생성하기 위해 개인정보의 특성을 분석하여 새로운 데이터를 생성하는 기법
		동형비밀분산 (Homomorphic secret sharing)	• 식별정보 또는 기타 식별가능정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉐어(share)*로 대체 *기밀사항을 재구성하는데 사용할 수 있는 하위 집합
		차분 프라이버시 (Differential privacy)	• 특정 개인에 대한 사전지식이 있는 상태에서 데이터베이스 질의(Query)에 대한 응답 값으로 개인을 알 수 없도록 응답 값에 임의의 숫자 잡음(Noise)을 추가하여 특정 개인의 존재 여부를 알 수 없도록 하는 기법 • 1개 항목이 차이나는 두 데이터베이스간의 차이(확률분포)를 기준으로 하는 프라이버시 보호 모델

2 개인정보의 가명처리 예시

※ 아래 모든 예시는 각 기법의 적용에 대한 예시이며 전체 데이터에 대한 가명처리에 대한 예시가 아닙니다.

1. 개인정보 삭제

▶ 삭제기술 : 선택된 항목을 제거하는 기술

❶ 삭제(Suppression) 수치형데이터 문자형데이터

- 원본정보에서 개인정보를 단순 삭제

※ 이때 남아 있는 정보 그 자체로도 분석의 유효성을 가져야 함과 동시에 개인을 식별할 수 없어야 하며, 인터넷 등에 공개되어 있는 정보 등과 결합하였을 경우에도 개인을 식별할 수 없어야 함

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적 포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성별	나이	통신료	단말기금액	누적포인트
남	41세	98,700	1,198,700	356,800
여	61세	69,400	505,400	203,000
남	30세	104,400	1,604,400	198,000
여	57세	954,800	3,954,800	20,532,000
남	28세	83,600	883,600	400,900

❷ 부분삭제(Partial suppression) 수치형데이터 문자형데이터

- 개인정보 전체를 삭제하는 방식이 아니라 일부를 삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김	남	41세	8888	서울특별시 중구	98,700	1,198,700	356,800
이	여	61세	2222	부산광역시 북구	69,400	505,400	203,000
박	남	30세	7777	광주광역시 서구	104,400	1,604,400	198,000
이	여	57세	4444	전라남도 나주시	954,800	3,954,800	20,532,000
최	남	28세	6666	세종특별자치시	83,600	883,600	400,900

❸ 행 항목 삭제(Record suppression) 수치형데이터 문자형데이터

- 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제
- 통계분석에 있어서 전체 평균에 비하여 오차범위를 벗어나는 자료를 제거할 때 사용

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

❹ 로컬 삭제(Local suppression) 수치형데이터 문자형데이터

- 특이정보를 해당 행 항목에서 삭제

(설명) 다른 누적포인트에 비하여 뚜렷이 구별되는 누적포인트를 항목에서 삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적 포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적 포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

❺ 마스킹(Masking) 수치형데이터 문자형데이터

- 특정 항목의 일부 또는 전부를 공백 또는 문자(‘*’, ‘_’ 등이나 전각 기호)로 대체

※ 분류는 개인정보 일부 또는 전부 대체로 분류되지만, 기술적으로 마스킹된 부분은 데이터로써의 가치가 없어져 일부 문건에서는 삭제로 분류되기도 함

성명	성별	나이	핸드폰번호
김철수	남	41세	010-6666-8888
이영희	여	61세	010-9999-2222
박민호	남	30세	010-2222-7777
이윤정	여	57세	010-3333-4444
최동욱	남	28세	010-5555-6666

마스킹

성명	성별	나이	핸드폰번호
김**	남	4*세	***-****-****
이**	여	6*세	***-****-****
박**	남	3*세	***-****-****
이**	여	5*세	***-****-****
최**	남	2*세	***-****-****

2. 개인정보 일부 또는 전부 대체

▶ 통계도구 : 데이터의 전체 구조를 변경하는 통계적 성질을 가진 기법

❶ 총계처리(Aggregation) 수치형데이터

- 평균값, 최댓값, 최소값, 최빈값, 중간값 등으로 처리

※ 단, 데이터 전체가 유사한 특징을 가진 개인으로 구성되어 있을 경우 그 데이터의 대푯값이 특정 개인의 정보를 그대로 노출시킬 수도 있으므로 주의 필요

❶-1. 부분총계(Micro Aggregation) 수치형데이터

- 정보집합물 내 하나 또는 그 이상의 행 항목에 해당하는 특정 열 항목을 총계처리즉, 다른 정보에 비하여 오차 범위가 큰 항목을 평균값 등으로 대체
- 동질 집합 내의 특정 항목을 총계처리 하거나 특정 조건에 너무 특이한 값이 있어 개인의 식별 가능성이 높지만 분석에 꼭 필요한 값인 경우 처리

(설명) 지역, 나이 기준으로 동질집합을 형성하고,
오차 범위가 큰 소득금액을 동질집합 내 평균값으로 대체

지역	나이	소득금액
서울	30대	5,987,900
서울	30대	28,169,700
서울	30대	3,009,600
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

지역	나이	소득금액
서울	30대	12,389,067
서울	30대	12,389,067
서울	30대	12,389,067
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

▶ 일반화기술 : 범주화로도 불리며, 특정한 값을 상위의 속성으로 대체

❶ 라운딩(Rounding) 수치형데이터

❶-1. 일반 라운딩

- 올림, 내림, 반올림 등의 기준을 적용하여 집계 처리하는 방법

나이
33세
61세
47세
66세
40세

올림	내림	반올림
40세	30세	30세
70세	60세	60세
50세	40세	50세
70세	60세	70세
40세	40세	40세

※ 적절하지 않은 라운딩의 경우 라운딩 후에도 남은 값의 유일성이 남게 될 수 있으며, 적용하는 단위에 대한 판단이 중요

금액	백 단위 라운딩
983,116,785	983,117,000
984,715,591	984,716,000
984,932,383	984.932.000
985,660,262	985,660,000
986,047,778	986,048,000

적절하지 않은 라운딩

금액	백만 단위 라운딩
983,116,785	980,000,000
984,715,591	980,000,000
984,932,383	980,000,000
985,660,262	990,000,000
986,047,778	990,000,000

적절한 라운딩

❶-2. 랜덤 라운딩(Random Rounding) 수치형데이터

- 수치 데이터를 임의의 수인 자리 수, 실제 수 기준으로 올림(round up) 또는 내림(round down)하는 기법

금액		금액
869,250	만 단위 라운딩	900,000
4,559,120	십만 단위 라운딩	4,000,000
13,601,564	십만 단위 라운딩	14,000,000
979,118	만 단위 라운딩	900,000
122,848,878	백만 단위 라운딩	120,000,000

❶-3. 제어 라운딩(Controlled rounding) 수치형데이터

- 라운딩 적용 시 값의 변경에 따라 행이나 열의 합이 원본의 행이나 열의 합과 일치하지 않는 단점을 해결하기 위해 원본과 결과가 동일하도록 라운딩을 적용하는 기법

※ 컴퓨터 프로그램으로 구현하기 어렵고 복잡한 통계표에는 적용하기 어려우며, 해결할 수 있는 방법이 존재하지 않을 수 있어 아직 실무에서는 잘 사용하지 않음

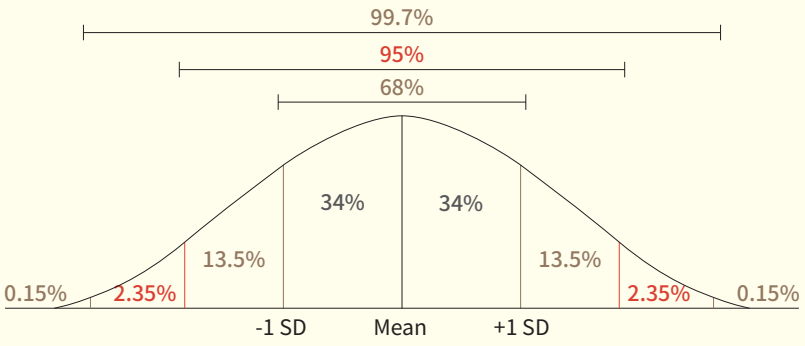
(설명) 나이에 대한 평균 분석 시 원본의 경우 평균이 51세가 되나 일반 라운딩을 적용한 경우 평균이 50세가 되어 결과가 다르게 되고, 이에 일부 값을 다르게 라운딩(제어)하여 평균 나이가 원본과 일치되도록 함

원본(나이)
33세
61세
50세
72세
43세
44세
23세
67세
68세
49세
평균 : 51세
합계 : 510

일반 라운딩	제어 라운딩
30세	30세
60세	60세
50세	50세
70세	70세
40세	40세
40세	50세
20세	20세
70세	70세
70세	70세
50세	50세
평균 : 50세	평균 : 51세
합계 : 500	합계 : 510

❷ 상하단코딩(Top and bottom coding) 수치형데이터

- 정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 식별성을 가질 수 있으며, 이를 해결하기 위해 적은 수의 분포를 가진 양 끝단의 정보를 범주화 등의 기법을 적용하여 식별성을 낮추는 기법



❸ 로컬 일반화(Local generalization) 수치형데이터

- 전체 정보집합물 중 특정 열 항목(들)에서 특이한 값을 가지거나 분포상의 특이성으로 인해 식별성이 높아지는 경우 해당 부분만 일반화를 적용하여 식별성을 낮추는 기법

(설명) 서울 지역의 30대 중 분포 상 다른 금액에 비해 특이한 값을 동질집합 내 범주화
※ 특이한 로컬(28,169,700)에만 3,009,600 ~ 28,169,700으로 범주화 할 수 있음

지역	나이	소득금액
서울	30대	5,987,900
서울	30대	28,169,700
서울	30대	3,009,600
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

지역	나이	소득금액
서울	30대	3,009,600~28,169,700
서울	30대	3,009,600~28,169,700
서울	30대	3,009,600~28,169,700
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

❹ 범위 방법(Data range) 수치형데이터

- 수치 데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 또는 구간(interval)으로 표현

(예시) 소득 3,300만원을 소득 3,000만원~4,000만원으로 대체 표기

❺ 문자데이터 범주화(Categorization of character data) 문자형데이터

- 문자로 저장된 정보에 대해 상위의 개념으로 범주화하는 기법

품목	품목
분유	육아용품
기저귀	육아용품
젖병	육아용품
샤워타올	육아용품
욕실화	육아용품

▶ 암호화 : 정보 가공 시 일정한 규칙의 알고리즘을 적용하여 대체

❶ 암호화(Encryption) 수치형데이터 문자형데이터

※ 암호화에 따른 세부적인 내용은 한국인터넷진흥원 암호이용활성화 관련 안내서 참조

❶-1. 양방향 암호화(Two-way encryption)

- 특정 정보에 대해 암호화와 암호화된 정보에 대한 복호화가 가능한 암호화 기법
- 암호화 및 복호화에 동일한 비밀키로 암호화하는 AES, ARIA 등 대칭키(Symmetric key) 방식과 공개키와 개인키를 이용하는 RSA 등 비대칭키(Asymmetric key) 방식으로 구분되며, 키(key) 관리에 주의 필요

❶-2. 일방향 암호화 – 암호학적 해시함수(One-way encryption-Cryptographic hash function)

- 원문에 대한 암호화의 적용만 가능하고 암호문에 대한 복호화 적용이 불가능한 암호화 기법
- 키가 없는 해시함수(MDC, Message Digest Code), 키가 있는 해시함수(MAC, Message Authentication Code), 솔트(Salt)가 있는 해시함수로 구분
- 암호화(해시처리)된 값에 대한 복호화가 불가능하고, 동일한 해시 값과 매핑(mapping)되는 2개의 고유한 서로 다른 입력값을 찾는 것이 계산상 불가능하여 충돌 가능성이 매우 적음

❶-3. 순서보존 암호화(Order-preserving encryption)

- 원본정보의 순서와 암호값의 순서가 동일하게 유지되는 암호화 방식
- 암호화된 상태에서도 원본정보의 순서가 유지되어 값들 간의 크기에 대한 비교 분석이 필요한 경우 안전한 분석이 가능

❶-4. 형태보존 암호화(Format-preserving encryption)

- 원본 정보의 형태와 암호화된 암호값의 형태가 동일하게 유지되는 암호화 방식
- 원본 정보와 동일한 크기와 구성 형태를 가지기 때문에 일반적인 암호화가 가지고 있는 저장 공간의 스키마 변경 이슈가 없어 저장 공간의 비용 증가를 해결할 수 있음
- 암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용

❶-5. 동형 암호화(Homomorphic encryption)

- 암호화된 상태에서의 연산이 가능한 암호화 방식
- 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용가능
- 암호화된 상태의 연산한 값을 복호화 하면 원래의 값을 연산한 것과 동일한 결과를 얻을 수 있는 4세대 암호화 기법

❶-6. 다형성 암호화(Polymorphic encryption)

- 가명정보의 부정한 결합을 차단하기 위해 각 도메인별로 서로 다른 가명처리 방법을 사용하여 정보를 제공하는 방법
- 정보 제공 시 서로 다른 방식의 암호화된 가명처리를 적용함에 따라 도메인별로 다른 가명정보를 가지게 됨

▶ 무작위화기술 : 속성의 값을 원래의 값과 다르게 변경

❶ 잡음 추가(Noise addition) 수치형데이터 문자형데이터

- 개인정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법
- 지정된 평균과 분산의 범위 내에서 잡음이 추가되므로 원 자료의 유용성을 해치지 않으나, 잡음값은 데이터 값과는 무관하기 때문에 유효한 데이터로 활용하기 곤란하여, 중요한 종적정보는 동일한 잡음을 사용해야함(예시로 입원일자에 +3이라는 노이즈를 추가하는 경우 퇴원일자에도 +3이라는 노이즈를 부여해야 전체 입원일수에 변화가 없음)

생년월일	잡음추가	잡음추가생년월일
2011-12-05	+3	2011-12-08
2016-08-09	-2	2016-08-07
2009-02-11	-5	2009-02-06
1998-05-27	-6	1998-05-21
1991-06-18	+9	1991-06-27

❷ 순열(치환)(Permutation) 수치형데이터 문자형데이터

- 기존 값은 유지하면서 개인이 식별되지 않도록 데이터를 재배열하는 방법
- 개인정보를 다른 행 항목의 정보와 무작위로 순서를 변경하여 전체정보에 대한 변경 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법

※ 데이터의 훼손 정도가 매우 큰 기법으로 무작위로 순서를 변경하는 조건 선정에 주의 필요

(설명) 원본과 비교하여 평균 분석 시 전체 재배열은 결과가 다르며
동질집합 내 재배열 결과는 동일

지역	나이	소득금액(원본)	소득금액(전체 재배열)	소득금액(동질집합 내 재배열)
서울	30대	5,987,900	2,789,200	3,009,600
서울	30대	8,169,700	4,607,300	5,987,900
서울	30대	3,009,600	5,987,900	8,169,700
나주	30대	4,607,300	2,940,100	2,940,100
나주	30대	3,560,800	8,169,700	4,607,300
나주	30대	2,940,100	5,048,300	3,560,800
세종	30대	6,088,400	3,009,600	2,789,200
세종	30대	2,789,200	3,560,800	5,048,300
세종	30대	5,048,300	6,088,400	6,088,400

원본 분석결과	지역	서울	나주	세종
	평균소득	5,722,400	3,702,733	4,641,967

전체 재배열 분석결과	지역	서울	나주	세종
	평균소득	4,461,467	5,048,300	4,219,600

동질집합 내 재배열 분석결과	지역	서울	나주	세종
	평균소득	5,722,400	3,702,733	4,641,967

❸ 토큰화(Tokenisation) 수치형데이터 문자형데이터

- 개인을 식별할 수 있는 정보를 토큰으로 변환 후 대체함으로써 개인정보를 직접 사용하여 발생하는 개인에 대한 식별 위험을 제거하여 개인정보를 보호하는 기술
- 토큰 생성 시 적용하는 기술은 의사난수생성 기법이나 일방향 암호화, 순서보존 암호화 기법을 주로 사용

고객번호	이름	성별	핸드폰번호	나이	회원등급	연간 이용액
D1304365	이공재	남	010-1234-5678	30세	2등급	3,782,459

의사난수 생성기 / 암호화 기법 / 형태보존 암호화

고객번호	이름	성별	핸드폰번호	나이	회원등급	연간 이용액
AD921648	Wzcd88qdp ekfhandkcosekrn	남	159-6857-6384	30세	2등급	3,782,459

❹ (의사)난수생성기((P)RNG, (Pseudo) Random Number Generator)

수치형데이터 문자형데이터

- 주어진 입력 값에 대해 예측이 불가능하고 패턴이 없는 값을 생성하는 메커니즘으로 임의의 숫자를 개인정보에 할당

※ 난수는 원칙적으로 규칙적인 배열순서가 없는 임의의 수를 의미하며 컴퓨터는 원천적으로 입력에 의한 처리 결과를 반환하는 것으로 처리의 방법과 입력이 동일하면 항상 동일한 출력이 발생하기 때문에 완전한 난수의 생성은 불가능

3. 가명·익명처리를 위한 다양한 기술(기타 기술)

❶ 표본추출(Sampling) 수치형데이터 문자형데이터

- 데이터 주체별로 전체 모집단이 아닌 표본에 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법
- 확률적 표본추출 방법과 비확률적 표본추출 방법으로 나누어지며, 확률적 표본추출이 통계적 분석에 많이 사용
- 확률적 표본추출 : 무작위 표본추출(복원 표본추출, 비 복원 표본추출), 계통적 표본추출, 층화 표본추출, 집락 표본추출 등
- 비확률적 표본부출 : 임의 표본추출, 판단 표본추출, 할당 표본추출, 누적 표본추출 등

❷ 해부화(Anatomization) 수치형데이터 문자형데이터

- 기존 하나의 데이터셋(테이블)을 식별성이 있는 정보집합물과 식별성이 없는 정보집합물로 구성된 2개의 데이터셋으로 분리하는 기술

Record ID	이름	성별	나이	월 납입금액	총 납부금액
1	조미선	F	33	817,250	66,300,000
2	홍길병	M	61	4,559,120	327,700,000
3	김영심	F	50	13,601,564	41,300,000
4	이미정	F	70	979,118	64,600,000
5	김경태	M	40	5,501,809	23,549,000
6	유영근	M	43	609,622	13,900,000

Record ID	이름	성별	나이
1	조미선	F	33
2	홍길병	M	61
3	김영심	F	50
4	이미정	F	70
5	김경태	M	40
6	유영근	M	43

Record ID	월 납입금액	총 납부금액
1	817,250	66,300,000
2	4,559,120	327,700,000
3	13,601,564	41,300,000
4	979,118	64,600,000
5	5,501,809	23,549,000
6	609,622	13,900,000

❸ 재현데이터(Synthetic data) 수치형데이터 문자형데이터

- 원본과 최대한 유사한 통계적 성질을 보이는 가상의 데이터를 생성하기 위해 개인정보의 특성을 분석하여 새로운 데이터를 생성하는 기법

※ 원본 데이터 포함 여부에 따라 완전 재현 데이터(Fully Synthetic Data), 부분 재현 데이터(Partially Synthetic Data), 하이브리드 재현 데이터(Hybrid Synthetic Data)로 구분

❹ 동형비밀분산(Homomorphic secret sharing) 수치형데이터 문자형데이터

- 식별정보 또는 기타 식별가능정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉐어(share)*로 대체

* 기밀사항을 재구성 하는 데 사용할 수 있는 하위 집합

※ 재식별은 가명·익명처리된 데이터의 쉐어를 소유한 모두가 동의하는 경우만 가능

❺ 차분 프라이버시(Differential privacy) 수치형데이터 문자형데이터

- 특정 개인에 대한 사전지식이 있는 상태에서 해당정보가 포함된 데이터베이스와 포함되지 않은 데이터베이스 질의(Query)에 대한 응답 값으로 개인을 알 수 없도록 응답 값에 임의의 숫자 잡음(Noise)을 추가하여 특정 개인의 존재 여부를 알 수 없도록 하는 기법
- 1개 항목이 차이나는 두 데이터베이스간의 차이(확률분포)를 기준으로 하는 프라이버시 보호 모델

※ 질의응답 값을 확률적으로 일정 크기 이하의 차이를 갖도록 함으로써 차이에 따른 차분 공격 방지

3 특이정보 처리 사례

1. 필요성

▶ 개인정보를 가명처리를 통해 특정 개인을 알아볼 수 없게 처리했더라도 '특이정보'를 통해 다른 정보와 쉽게 결합하여 개인을 알아 볼 수 있음

- 따라서 특이정보의 유형 등을 살펴보고 가명정보 내 해당 유형의 정보가 존재하고 있는지 검토할 필요가 있음

※ 특이정보는 관측된 데이터의 범위에서 많이 벗어난 아주 작은 값이나 아주 큰 값을 의미

2. 특이정보 사례

▶ 특정 기관의 급여가 2천만원에서 6천만원까지 고루 분포되어 있는데, 일부 고액 급여 수령자가 발생하는 경우

▶ 특정 직업의 소속인원이 전국에서 약 300명 정도로 추정되는데, 지역에 극소수(1~2인)만 존재하고 있는 경우

▶ 정보공개 규정에 따라 공개되는 정보에서 특정 나이대가 현저하게 적게 나타나는 경우

3. 특이정보 관찰 방법

▶ 정보의 특이정보는 3시그마규칙 또는 도수분포표 등을 이용하여 검토할 수 있음

- 3시그마 규칙 : 68-95-99.7규칙이라고도 하며, 정보의 분포의 3시그마(표준편차) 범위에 거의 모든 값들(99.7%)가 들어가는 것을 의미
- 도수분포표 : 항목에 대한 값을 적당한 범위로 분류하고, 각 범위에 해당하는 수량을 조사하여 표로 나타내는 것을 의미

〈급여〉

직원	급여(만원)
직원1	2,200
직원2	3,400
직원3	4,600
직원4	5,300
직원5	10,000
직원6	6,700

※ 3시그마 규칙을 이용 하여 표준편차에 벗어난 특이정보 검토

〈지역, 직업〉

주소	직업	빈도
경기	국회의원	5
경기	국회의원	5
강원	국회의원	1
경기	국회의원	5
경기	국회의원	5
경기	국회의원	5

※ 지역에 대한 도수분포(빈도)를 이용하여 특이정보 검토

〈나이〉

나이(세)	빈도
10~20	4
20~30	11
30~40	21
40~50	18
50~60	5
60~70	1

※ 특정 나이에 도수분포(빈도)를 측정 하여 특이정보 검토

4. 특이정보 처리 사례

▶ 삭제 기법을 활용한 목적별 사용 예시

- 분석 목적에 해당 정보가 없어도 분석에 크게 영향이 없는 경우에만 가능한 기법, 해당 특이 정보를 삭제하여 개인 식별성을 제거

가. 로컬 삭제(Local suppression)

일반적으로 특이정보 처리에 많이 사용되는 기법으로 도수분포표를 활용하여 빈도가 적은 항목을 삭제하여 처리하는 방법

〈로컬삭제 기법 예시〉

나이	주소	직업	월소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	1,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	울릉도	변호사	200만원

나이	주소	직업	소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	1,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	Null	변호사	200만원

나. 행 삭제(Record suppression)

특이정보로 인해 개인의 식별가능성이 있는 경우 사용되는 기법으로 특이정보를 가지고 있는 행 전체를 삭제하여 처리하는 방법

※ 통계 분석에서 특이정보는 분석 목적을 달성하기보다 분석의 목적을 저해하는 요소로 작용하는 경우가 있으며, 이 경우 행 삭제 기법이 가장 적절한 기법이 될 수 있음

〈레코드 삭제 기법 예시〉

나이	주소	직업	월소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	7,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

나이	주소	직업	소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

▶ 통계적 기법의 종류와 목적별 사용 예시

- 분석 목적에 특이정보를 가지고 있는 해당 정보가 필요한 경우 활용하는 기법으로, 해당 특이정보를 통계적인 방법을 통해 통계값으로 변경하여 사용

가. 단일 속성으로 대체(Combining a set of attributes into a single attribute)

숫자형 정보가 아닌 경우(문자형 등) 주로 사용되는 방법으로 분류군의 상위로 묶어 처리하는 방법

※ 특정한 직업이 희귀하여 개인의 식별이 가능한 경우 상위의 분류로 변경하여 사용함으로 희귀성을 제거

〈단일속성 대체 예시〉

나이	주소	직업	월소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	판사	1,200만원
35	경기	검사	800만원
35	경기	변호사	600만원
35	경기	변호사	1,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

나이	주소	직업	소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	법조인	1,200만원
35	경기	법조인	800만원
35	경기	변호사	600만원
35	경기	변호사	1,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

나. 로컬 일반화(Local generalization)

선택한 행에서 일부 특정 값을 일반화하여 활용하는 기법으로, 다른 행의 속성값은 수정하지 않고 희귀 값을 가진 속성값만 처리하여 사용

〈로컬 일반화(상단 코딩) 기법 예시〉

나이	주소	직업	월소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
36	서울	변호사	1,200만원
36	경기	변호사	800만원
36	경기	변호사	600만원
36	경기	변호사	1,300만원
37	경기	변호사	300만원
…	…	…	…
84	경기	변호사	800만원
88	경기	변호사	200만원

나이	주소	직업	소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
36	서울	변호사	1,200만원
36	경기	변호사	800만원
36	경기	변호사	600만원
36	경기	변호사	1,300만원
37	경기	변호사	300만원
…	…	…	…
80초과	경기	변호사	800만원
80초과	경기	변호사	200만원

다. 부분 총계(Micro Aggregation)

부분 총계는 일부(특정그룹 값의 합)속성에서 정확한 통계적 값을 확인하는 기법으로, 로컬일반화보다 일부 속성에서 정확한 값을 알 수 있음

〈부분 총계 기법 예시〉

나이	주소	직업	월소득
35	경기	변호사	600만원
35	경기	변호사	700만원
35	경기	변호사	500만원
35	경기	변호사	700만원
35	경기	변호사	6,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	7,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

나이	주소	직업	소득
35	경기	변호사	600만원
35	경기	변호사	700만원
35	경기	변호사	500만원
35	경기	변호사	700만원
35	경기	변호사	6,750만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	6,750만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

참고2 비정형데이터 가명처리 기술 및 예시

※ 아래 가명·익명처리 기술은 비정형데이터의 가명처리 시 처리자의 단순참고를 위해 국내·외 자료와 전문가 논의를 통해 구성된 예시자료로, 해당 가이드라인에서 제시되지 않은 기술도 처리자의 판단에 따라 데이터 활용 분야·상황에 맞게 자유롭게 활용할 수 있음

1. 영상정보

① 이미지 필터링 기술

▶ 필터링 기술(❶ 블러링, ❷ 픽셀화, ❸ 마스킹)로 처리된 영상 이미지

* 얼굴 이외에도 사람 형상, 신체, 옷차림 등, 사람과 관련된 사물/동물 등도 필터링 대상이 될 수 있음

원본 이미지 ❶ 블러링 ❷ 픽셀화 ❸ 마스킹

※ 출처 : Tao Li and Lei Lin, Natural Face De-identification with Measurable Privacy(CV-COPS, 2019)

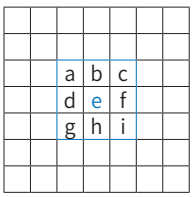
▶ 필터링 기술 : 블러링(Blurring), 픽셀화(Pixelization), 마스킹(Masking)

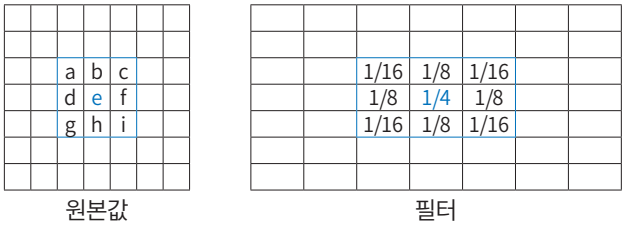
❶ 이미지 블러링(Blurring)

- 필터의 격자 사이즈 : 3*3 또는 5*5 등
- 필터링의 종류 : 평균, 가우시안, 중간값, 바이레터럴(Bilateral)등

〈 필터링의 종류 〉

종류	설명
평균	• 입력 이미지의 현재 위치에서 예를 들어 3×3 격자 범위의 주변 픽셀값의 평균을 구하여 원 픽셀값을 결과 이미지의 픽셀값으로 대체 ⇒ e = 1/9(a+b+c+d+e+f+g+h+i)



종류	설명
가우시안	• 입력 이미지의 현재 위치에서 예를 들어 3×3 격자 범위의 주변 픽셀값에 가중치를 부여(중심으로 갈수록 높은 가중치를 부여)하여 원 픽셀값을 결과 이미지의 픽셀값으로 대체 ※ 입력값으로 커널크기, 가우시안 표준편차, 필터의 데이터 타입이 사용됨 ⇒ e = 1/16*a+1/8*b+1/16*c+1/8*d+1/4*e+1/8*f+1/16*g+1/8*h+1/16*i  원본값 필터
중앙값	• 입력 이미지의 현재 위치에서 예를 들어 3×3 범위의 주변 픽셀값의 중앙값을 구하여 원 픽셀값을 결과 이미지의 픽셀값으로 대체
바이레터럴 (bilateral)	• 원본 이미지로부터 최대한 노이즈는 제거하고 에지는 보존하기 위한 것으로 공간과 밀도를 함께 고려하여 원 픽셀값을 결과 이미지의 픽셀값으로 대체

- 추가정보 : 원본 정보, 필터 사이즈, 가우시안 표준편차, 필터링의 종류 등

※ 블러링 처리된 이미지를 원본 이미지로 직접 복원은 불가하나 최근에는 자연 영상의 통계적 특성(natural image statistics)에 기반한 모션 블러 추정 및 제거 등 디블러링 방법들이 제안되고 있음

❷ 이미지 픽셀화(Pixelization) (=모자이크(Mosaic))

- 이미지 블러링의 평균 필터와 유사하나 계산한 평균값을 해당 픽셀뿐만 아니라 적용한 주변 모든 픽셀(예를 들어 3×3 범위의)에 대체한다는 점이 다름
- 추가정보 : 원본 정보, 필터 사이즈 등

〈블러링과 픽셀화 처리 시 참고 사항〉

- 표준 QCIF(Quarter Common Intermediate Format) 화상회의 이미지 크기(176×144 픽셀, 픽셀당 24비트)에서 초당 24프레임의 프레임 속도로 변환하고 PC에서 재생하기에 적합한 AVI 파일로 저장한 5개 이미지 스틸컷에 대하여 설문을 통한 실험 결과, 식별가능 임계값을 블러링시 최소 레벨 5이상을 픽셀화시 최소 레벨 6이상을 임계값으로 제시

※ 출처 : Michael Boyle 외 2인, the effects of filtered video on awareness and privacy(2000)

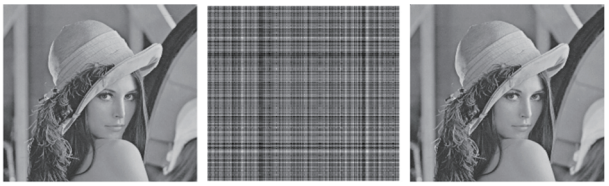
- 또한 현재 인터넷 상에 이미지 블러링 및 픽셀화 처리가 가능한 OpenCV 등 라이브러리나 소스코드 등이 공개되어 있어 쉽게 처리가 가능

❸ 이미지 마스킹(Masking) (=블랙박스(Blackbox))

② 이미지 암호화

▶ 원본 이미지의 일부를 암호화하여 데이터 주체를 알아볼 수 없도록 하는 기법

▶ 이미지 암호화로 처리된 영상 이미지



원본 이미지 암호화된 이미지 복호화된 이미지

※ 출처 : Lihua Gong 외 3인, Image compression-encryption algorithms by combining hyper-chaotic system with discrete fractional random transform(Optics and Laser Technology, 2018)

▶ 이미지 암호화 기술 : 이산코사인변환 기반 암호화, 픽셀 위치 기반 암호화

❶ 이산코사인변환 기반 암호화(DCT, Discrete Cosine Transform)

- 영상을 주파수 영역으로 바꾸어 특정 부분만 암호화

❷ 픽셀 위치 기반 암호화(Pixel location)

- 픽셀의 위치를 일정한 규칙에 따라 바꾸는 방식으로 구현하는 암호화

③ 얼굴 합성(프라이버시 보존형 데이터 마이닝)

▶ K-익명성 프라이버시보호 모델을 확장하여 K명의 얼굴을 합성한 기술로 K-same 모델로도 부름

※ 출처 : E.M.Newton 외 2인, preserving privacy by de-identifying face images(2005)

▶ K-same 기법으로 처리된 영상 이미지

Original $k=2$ $k=3$ $k=5$ $k=10$ $k=20$

※ 출처 : R. Gross 외 2인, Integrating utility into face de-identification(2005)

▶ K-same 모델의 개선

- K-same 모델을 보다 개선하여 개인정보 보호와 유용성의 균형*을 맞추기 위한 K-Same-Select 모델 등장

* 공개 위험(즉, 이미지 난독화 수준)과 분류 정확도 간의 균형

※ 출처 : R. Gross 외 2인, Integrating utility into face de-identification(2005)

- K-same 모델의 정확도를 유지하면서 데이터 셋의 민감한 정보를 숨기기 위해 기존 샘플에 무작위 노이즈를 추가하거나 새 샘플을 생성하는 난독화(obfuscate) 기능을 추가로 설계

* 안전성은 높으나 성별이 바뀌는 등 이미지 왜곡이 심함

※ 출처 : T. Zhang, Privacy-preserving machine learning through data obfuscation(2018)

④ 인페인팅(Inpainting)

▶ 영상 내 개인 식별 영역을 제거한 후 다른 물체 또는 배경으로 대체하여 신원을 보호하는 기술

▶ 인페이팅 기법으로 처리된 영상 이미지

원본 이미지 원본의 플래그 이미지 인페인팅 이미지

※ 출처 : Takahiro Ogawa 외 1인, Image inpainting based on sparse representations with a perceptual metric(2013)

▶ 인페인팅 기술 : 패치 기반 인페인팅, 객체 기반 인페인팅

❶ 패치 기반 인페인팅 기술

- 영상 프레임 내 공백과 가장 비슷한 영역을 찾아 채우는 방식

❷ 객체 기반 인페인팅 기술

- 영상을 배경과 객체로 구분해 객체를 제거 후 남은 부분은 배경으로 채우는 방식

⑤ AI 이용 영상정보 가명처리

▶ 얼굴 보존형 가명·익명처리 기술(De-identification without losing faces)

- 원본 얼굴의 요소를 변경하거나 얼굴을 완전히 합성하는 대신, 훈련된 얼굴 속성 전달 모델을 사용하여 동의한 대상의 소수(보통 2~3명)인 기증자의 얼굴에 비 신원 관련 얼굴 속성을 매핑

원본 이미지 K-same 이미지 제안 이미지

※ 출처 : Yuezun Li, De-identification Without Losing Faces(2019)

▶ AnonymousNet 프레임워크

- 얼굴 보존형 가명·익명처리 기술의 프레임과 비슷해 보이지만 다음의 4단계 절차를 통하여 아이덴티티가 완전히 다른 이미지를 생성

※ (1단계) 얼굴 특징 추출, (2단계) 의미 기반 속성 난독화, (3단계) 익명화된 얼굴 생성, (4단계) 적대적 교란 절차

원본 이미지 AnonymousNet

※ 출처 : Tao Li 외 1인, AnonymousNet: Natural Face De-Identification with Measurable Privacy (CV-COPS 2019)

▶ AI 딥러닝 기반의 알고리즘을 활용

- 얼굴 및 차량 번호판 등을 추출한 후 각종 SW 라이브러리를 활용하여 블러링 처리

※ 미검출이나 오검출 얼굴 등에 대해서는 수작업을 통하여 추가 보정

▶ AI 적대적 생성 신경망 기반 모델(GAN, Generative Adversarial Networks)

- 이미지에서 보존해야할 영역을 수치화하여 연속형 데이터로 처리한 후 해당 개별 데이터에 노이즈를 추가하거나 재현 처리

2. 음성정보

① 음성정보 자체에 대한 가명처리

▪ 대부분 텍스트로 변환하여 처리하고 있으며 그 외 발화자의 개인정보(비속어, 성적, 정치적 발언 등 포함)에 대한 규칙을 정하여 해당 구간을 단순히 삭제하거나 혹은 음성 변형의 원리에 기반하여 음조를 변형하여 처리

▶ 규칙기반 개인정보 단순 삭제

- 음성 데이터 상에서 규칙*을 정하여 개인식별가능 정보를 단순히 삭제

* (예) 인명, 지명 등 개인식별가능 정보가 포함된 카드번호, 주민등록번호, 전화번호 등 삭제하고, 비속어, 성적 및 정치적 발언의 경우도 삭제하거나 다른 언어로 대체하는 등

※ 이 방법은 완전 자동으로 수행이 어려운 측면이 있으므로 추가적인 검수가 필요할 수 있음

- 음성 변형의 원리에 기반하여 음성 데이터 내용에 영향을 주지 않으면서 주어진 발화의 비언어적 특징들을 수정

❶ 음성 변형(VT, Voice Transformation)

- 원본 정보 소스, 필터, 소스와 필터의 조합 등을 통하여 음성을 변형하는 기법

※ 소스로는 시간량(time-sclae), 음조(pitch) 또는 음량(energy)을 변조할 수 있으며, 필터로는 음성 트랙 시스템에 기반하여 음폭을 변형 가능

종류	설명
시간량 (time-scale) 수정	• 원래 음성의 지각적(perceptual) 품질에 영향을 주지 않으면서 조음 속도를 변경
음조 (pitch, 음높이, 고저) 수정	• 짧은 시간 스펙트럼 엔벨로프(포먼트의 위치 및 대역폭)와 시간을 보존하면서 스펙트럼의 하모닉 구성 요소 사이의 간격을 압축하거나 확장하기 위해 기본 주파수를 변경
음량(energy) 변형	• 입력 음성의 인지된 크기를 수정
필터 수정	• 성도(vocal tract) 시스템의 주파수 응답의 크기 스펙트럼 수정(예 : 여성의 목소리를 수정하여 아이처럼 들리게 함)
소스 수정과 필터 수정의 조합	• 일반적으로 사용되는 기법으로 예를 들어 화자의 음성을 다른 화자의 음성처럼 들리도록 수정하려면 운율(prosody)과 성도(vocal tract) 수정을 결합해야 함 → 이때 특정 대상 화자가 제공되는 경우 이를 음성 변환(voice conversion)이라 부름

❷ 음성 변환(VC, Voice Conversion)

- 음성 변형(VT)의 특별한 형태로 발화자의 음성 특성을 특정(대상) 화자의 음성 특성으로 매핑하는 것으로 원래의 목소리를 다른 사람의 목소리로 변환하는 기법
- 음성 변환(VC)은 '텍스트 종속적'이거나 '텍스트 독립적'일 수 있음

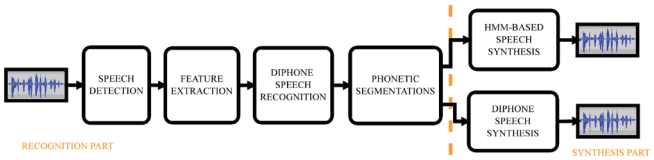
종류	설명
텍스트 종속적	• AI 학습 단계에서 병렬 말뭉치(동일한 텍스트를 발화하는 소스 및 대상 화자의 훈련 자료)가 필요 ※ 실제 응용 프로그램에서 음성 비식별화에 이러한 방식을 사용하는 것은 주요 제약 사항임
텍스트 독립적	• 짧은 시간 스펙트럼 엔벨로프(포먼트의 위치 및 대역폭)와 시간을 보존하면서 스펙트럼의 하모닉 구성 요소 사이의 간격을 압축하거나 확장하기 위해 기본 주파수를 변경

❸ GMM(Gaussian Mixture Model) 매핑기반의 음성 변환 방식

- 영상정보에서의 K-same 방식과 유사한 텍스트 독립적인 GMM 매핑기반의 음성 변환 기법
- 화자 인식 시스템이 비식별화된 음성에서 우연보다 더 나은 성능으로 실제 화자의 신원을 인식할 수 없는 반면, 비식별화된 음성은 대부분의 경우 이해할 수 있음
- 모든 음성이 동일한 대상 화자에 매핑되기 때문에 서로 다른 화자를 구별할 수 있는 (비식별화된) 음성을 생성하는 기능은 부족

❹ HMM(Hidden Markov Model) 기반과 TD-PSOLA(diphone Time-Domain Pitch Synchronous Overlap and Add) 기술을 기반으로 한 음성 변형 방식(DROPSY)

- HMM과 TD-PSOLA 기술 기반 DROPSY 알고리즘



※ 출처 : T. Justin 외 5인, Speaker de- identification using diphone recognition and speech synthesis(2015)

- PitchScale SOLA(Synchronous Overlap and Add) 알고리즘(2단계)

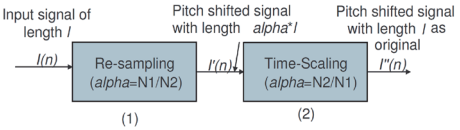


Figure 4.4: Pitch Distortion Algorithm (two processes)

※ 출처 : Y. Stylianou, Voice transformation : a survey(2009)

② 음성을 텍스트로 변환(STT, Speech To Text) 후 가명처리

▶ 개인식별정보가 포함된 음성을 텍스트로 변환 후 변환한 텍스트에서 개인식별정보를 가명처리하고 다시 음성으로 변환하는 방식

3. 텍스트정보

① 규칙기반 개인정보 단순 삭제 혹은 마스킹

▶ 사전에 텍스트 내 개인식별(가능)정보들을 정의하고 정의된 형태(포맷)에 기반하여 해당 정보를 삭제하거나 마스킹, 대체 처리 등의 방법으로 제거

② 스크러빙(Scrubbing)

▶ 원 텍스트의 내용과 구조를 보존하면서 즉석해서 파싱을 통하여 혹은 파싱 이후 개인식별(가능)정보만을 제거(마스킹 혹은 대체)하는 것으로 이 경우 다수의 정보 주체와 해당 속성들 사이의 명확한 연관성이 없어질 수도 있음

- 즉, 특정 정보 주체가 어느 속성을 지칭하는지 알 수가 없게 될 수도 있음
- 단순 삭제 혹은 마스킹 방법과 유사하며 수작업이 아닌 자동화 SW를 이용한다는 점이 다름

③ 정규표현식(Regular Expression)

▶ 문자나 혹은 문자열의 일정한 패턴을 표현하는 일종의 형식 언어

	항목	텍스트 예시	정규표현식
1	전화번호	010-1111-2222	^d{3}\-d{4}\-d{4}$
2	이메일주소	aa@bb.net	^[0-9a-zA-Z]([-_.]?[0-9a-zA-Z])*@[0-9a-zA-Z]([-_.]?[0-9a-zA-Z])*.[a-zA-Z]{2,3}$/i;
3	IP주소	111.111.111.111	^\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}$
4	주민등록번호	220101-1111111	^\d{6}\-[1-4]\d{6}$

※ 사용시 오타나 혹은 문자 숫자 혼합 등에 따라 누락되는 경우가 존재하므로 추가 검수가 필요할 수 있음

④ 주석 달기(Annotation)

▶ 주어진 텍스트를 논리적으로 분할한 후 분할된 단어(들)에 주석을 첨가하는 기법

종류	설명
규칙(Rule) 기반	• 구문 문법의 규칙에 따라 텍스트를 토큰(예 : 사전 정의된 수의 단어)으로 나누는 것으로 고급 규칙의 경우 정규표현식을 사용하여 정의
사전(Dictionary) 기반	• 개인식별(가능)정보들을 미리 사전으로 정의한 후 개체명 인식(NER, Named Entity Recognition) 기술을 이용하여 주어진 텍스트(이름, 주소, 전화번호 등)와 일치시킴 * 사전기반의 경우 AI 기술을 활용하여 각 단어들을 자동으로 인식하지만 모든 단어(들)을 완벽하게 인식하는 것은 어려우며, 따라서 맥락(context)에 의한 개인식별 가능성에 대한 조치는 어려울 수 있음

⑤ AI 기반 텍스트정보 가명처리

▶ 규칙, 정규표현식 등을 통한 개인정보 검출 및 마스킹은 정확도 측면에서 한계가 있을 수 있으며, 이를 보완하기 위해 딥러닝 기술 등을 적용한 자연어 처리 언어 모델을 통해 사전에 정의되지 않은 패턴의 개인정보를 검출하고 마스킹할 수 있음

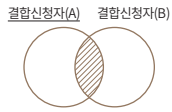
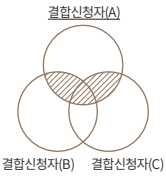
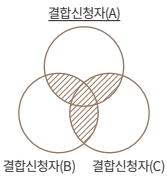
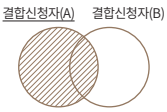
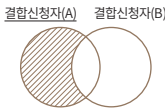
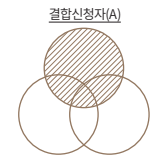
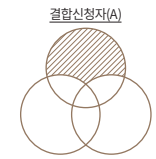
▶ 학습방법에 따라 다양한 형태의 인공지능 기반 개인정보 검출 기법 존재 (HMM, MEM, CRFs, structural SVM, Deep-Learning)

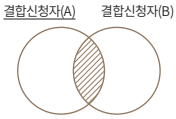
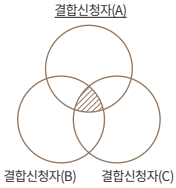
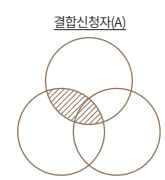
▶ 규칙기반의 유연성 부족을 해결하기 위해 패턴이나 규칙을 수동이나 반자동으로 작성하고 인공지능을 통해 사전(dictionary)을 확장하여 규칙에 적용되는 개인식별가능정보를 새롭게 검출할 수 있음(단계별 규칙을 순차적으로 적용하여 가중치에 따라 인명, 지명, 조직명 등으로 범주를 결정하는 등)

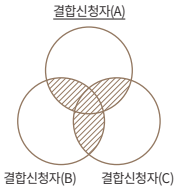
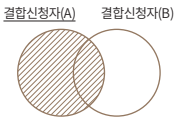
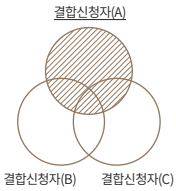
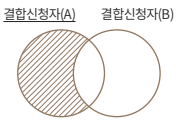
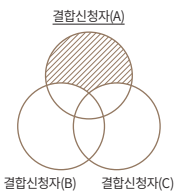
⑥ 텍스트를 테이블 형식으로 변환

▶ 주어진 텍스트를 구문 문법의 규칙에 따라 파싱(parsing, 정해진 규칙에 따라 문장의 구문을 분할)한 다음 분할된 각 세그먼트들을 열과 행이 있는 테이블 형태로 정렬한 후 나머지 데이터들은 삭제
- 변환된 테이블은 기존 정형 데이터에 대한 가명처리를 적용
- 실제로는 텍스트 자체의 복잡성으로 인하여 구조화된 테이블로의 변환이 불가능할 수도 있음

참고3 결합의 다양한 유형

공통결합 (INNER JOIN)			확대결합 (OUTER JOIN)	잔여결합 (ANTI-INNER JOIN)
단일 (INNER SINGLE)	다중 (INNER MULTI)	완전 (INNER FULL)	단일 (OUTER SINGLE)	단일 (ANTI-INNER SINGLE)
				
				

유형		설명
공통단일결합 (INNER SINGLE JOIN)		결합신청자(A) 결합키 | 일련번호 dac…387ea | A1 c2e…7067c | A2 cd1…78ccd | A3 89c…df12d | A4 9a0…a5ddf | A5 … | … 결합신청자(B) 결합키 | 일련번호 abc…2cabd | B1 dac…387ea | B2 c2e…7067c | B3 9a0…a5ddf | B4 rew…nm3d0 | B5 … | … 결합형태 일련번호 A1 | B2 A2 | B3 A5 | B4 … | … … | …
		결합신청자(A) 결합키 | 일련번호 dac…387ea | A1 c2e…7067c | A2 cd1…78ccd | A3 89c…df12d | A4 9a0…a5ddf | A5 … | … 결합신청자(B) 결합키 | 일련번호 abc…2cabd | B1 dac…387ea | B2 c2e…7067c | B3 9a0…a5ddf | B4 rew…nm3d0 | B5 … | … 결합신청자(C) 결합키 | 일련번호 o23…ere12 | C1 dac…387ea | C2 c2e…7067c | C3 89c…df12d | C4 rew…nm3d0 | C5 … | … 결합형태 일련번호 A1 | B2 | C2 A2 | B3 | C3 … | … | … … | … | … … | … | … … | … | …
공통다중결합 (INNER MULTI JOIN)		결합신청자(A) 결합키 | 일련번호 dac…387ea | A1 c2e…7067c | A2 cd1…78ccd | A3 … | … 89c…df12d | A4 9a0…a5ddf | A5 … | … 결합신청자(B) 결합키 | 일련번호 abc…2cabd | B1 dac…387ea | B2 c2e…7067c | B3 9a0…a5ddf | B4 … | … rew…nm3d0 | B5 … | … 결합신청자(C) 결합키 | 일련번호 o23…ere12 | C1 dac…387ea | C2 c2e…7067c | C3 … | … 89c…df12d | C4 rew…nm3d0 | C5 … | … 결합형태 일련번호 A1 | B2 | C2 A2 | B3 | C3 … | … | … A5 | B4 | - A4 | - | C4 … | … | … … | … | …
		결합신청자(A) 결합키 | 일련번호 dac…387ea | A1 c2e…7067c | A2 cd1…78ccd | A3 … | … 89c…df12d | A4 9a0…a5ddf | A5 … | … 결합신청자(B) 결합키 | 일련번호 abc…2cabd | B1 dac…387ea | B2 c2e…7067c | B3 9a0…a5ddf | B4 … | … rew…nm3d0 | B5 … | … 결합신청자(C) 결합키 | 일련번호 o23…ere12 | C1 dac…387ea | C2 c2e…7067c | C3 … | … 89c…df12d | C4 rew…nm3d0 | C5 … | … 결합형태 일련번호 A1 | B2 | C2 A2 | B3 | C3 … | … | … A5 | B4 | - … | … | … … | … | …

유형	설명
공통완전결합 (INNER FULL JOIN)	
확대단일결합 (OUTER SINGLE JOIN)	
	
잔여단일결합 (ANTI-INNER SINGLE JOIN)	
	

결합신청자(A)

결합키	일련번호
dac…387ea	A1
c2e…7067c	A2
cd1…78ccd	A3
…	…
89c…df12d	A4
9a0…a5ddf	A5
…	…

결합신청자(B)

결합키	일련번호
abc…2cabd	B1
dac…387ea	B2
c2e…7067c	B3
9a0…a5ddf	B4
…	…
rew…nm3d0	B5
…	…

결합신청자(C)

결합키	일련번호
o23…ere12	C1
dac…387ea	C2
c2e…7067c	C3
…	…
89c…df12d	C4
rew…nm3d0	C5
…	…

결합형태

일련번호		
A1	B2	C2
A2	B3	C3
…	…	…
A5	B4	-
A4	-	C4
-	B5	C5
…	…	…

결합신청자(A)

결합키	일련번호
dac…387ea	A1
c2e…7067c	A2
…	…
cd1…78ccd	A3
89c…df12d	A4
…	…
9a0…a5ddf	A5

결합신청자(B)

결합키	일련번호
abc…2cabd	B1
dac…387ea	B2
c2e…7067c	B3
…	…
…	…
9a0…a5ddf	B4
rew…nm3d0	B5

결합형태

일련번호	
A1	B2
A2	B3
A3	-
A4	-
A5	B4
…	…
…	…

결합신청자(A)

결합키	일련번호
dac…387ea	A1
c2e…7067c	A2
…	…
cd1…78ccd	A3
…	…
89c…df12d	A4
9a0…a5ddf	A5

결합신청자(B)

결합키	일련번호
abc…2cabd	B1
dac…387ea	B2
c2e…7067c	B3
…	…
9a0…a5ddf	B4
rew…nm3d0	B5
…	…

결합신청자(C)

결합키	일련번호
o23…ere12	C1
dac…387ea	C2
c2e…7067c	C3
…	…
89c…df12d	C4
rew…nm3d0	C5
…	…

결합형태

일련번호		
A1	B2	C2
A2	B3	C3
…	…	…
A3	-	-
A5	B4	-
A4	-	C4
…	…	…

결합신청자(A)

결합키	일련번호
dac…387ea	A1
c2e…7067c	A2
cd1…78ccd	A3
89c…df12d	A4
9a0…a5ddf	A5
…	…
…	…

결합신청자(B)

결합키	일련번호
abc…2cabd	B1
dac…387ea	B2
…	…
…	…
c2e…7067c	B3
9a0…a5ddf	B4
rew…nm3d0	B5

결합형태

일련번호	
…	…
…	…
A3	-
A4	-
…	…
…	…
…	…

결합신청자(A)

결합키	일련번호
dac…387ea	A1
c2e…7067c	A2
cd1…78ccd	A3
89c…df12d	A4
9a0…a5ddf	A5
…	…
…	…

결합신청자(B)

결합키	일련번호
abc…2cabd	B1
dac…387ea	B2
…	…
c2e…7067c	B3
9a0…a5ddf	B4
rew…nm3d0	B5
…	…

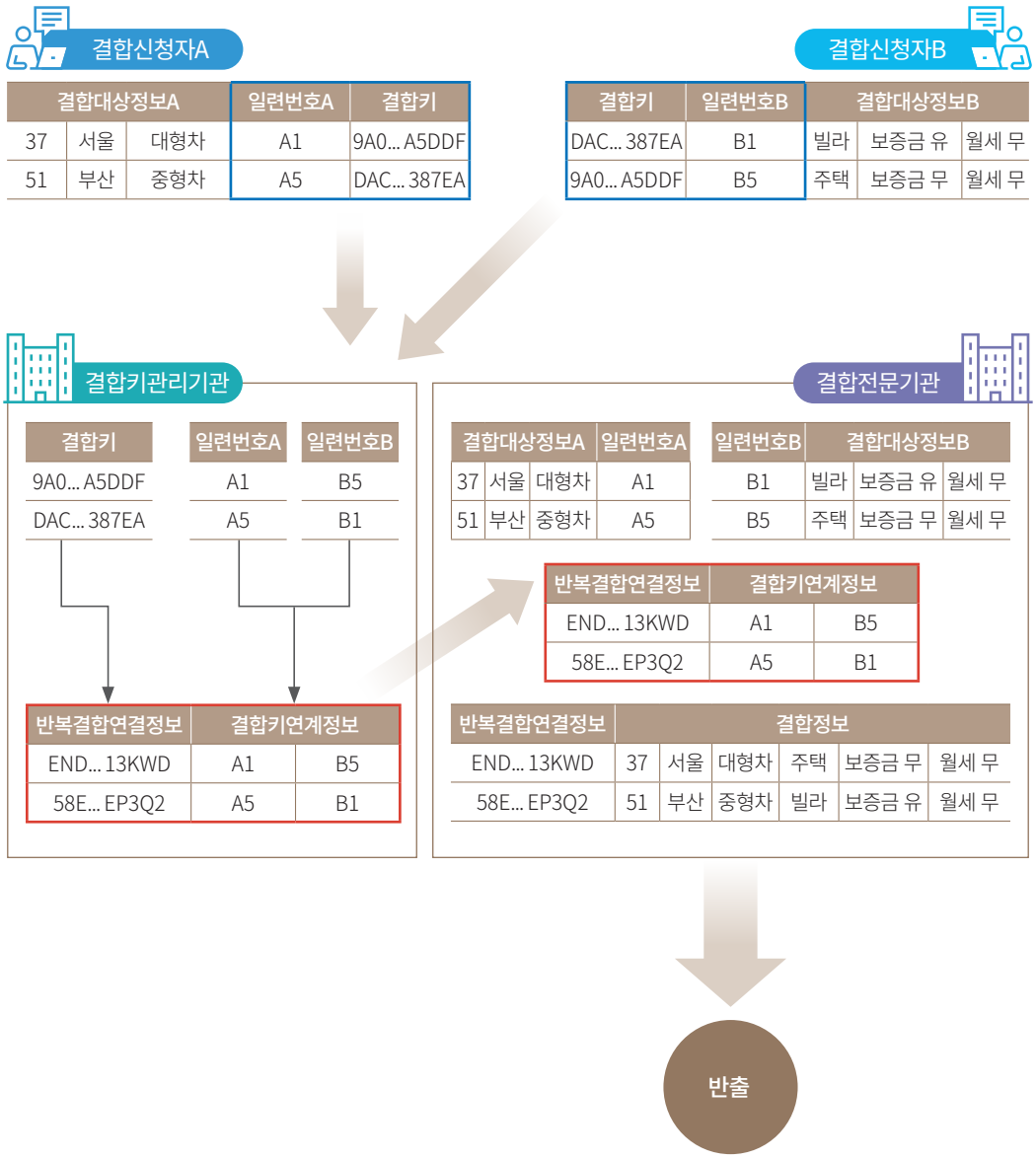
결합신청자(C)

결합키	일련번호
o23…ere12	C1
dac…387ea	C2
…	…
c2e…7067c	C3
89c…df12d	C4
rew…nm3d0	C5
…	…

결합형태

일련번호		
…	…	…
…	…	…
A3	-	-
…	…	…
…	…	…
…	…	…
…	…	…

참고4 시계열 분석을 위한 반복결합 절차



① 결합신청자는 가명처리 대상 정보에 정보주체별로 일련번호를 생성하고, 결합키관리기관과 협의한 방법에 따라 결합키를 생성

- 생성된 결합키와 일련번호는 결합키관리기관으로 송신

② 결합키관리기관은 결합신청자로부터 수신받은 결합키를 활용하여 1) 시계열 분석에 필요한 키(반복결합연결정보)와 2) 결합키연계정보를 생성

※ 일반적인 결합의 경우 추가적인 시계열 분석에 필요한 키를 생성하지 않음

③ 결합키관리기관은 생성한 1), 2)의 정보를 결합전문기관에 송신하고, 결합전문기관은 수신받은 2)를 활용하여 결합신청자의 가명정보를 결합

④ 결합전문기관은 결합정보에 1)을 포함하여 반출하고, 결합신청자는 반출정보에 대한 안전조치 의무 수행

※ 유의사항 : 반복결합 신청자는 추후 반출되는 정보와의 연계·분석을 위하여 결합키에 사용된 결합키 생성항목, 인코딩 방식, 알고리즘(Salt값 제외)을 보관하여야 함

추가 반복결합 신청 및 활용 방법

결합신청자가 추가 시계열 분석을 위한 반복결합을 신청하는 경우 결합신청자는 최초 반복결합 신청시 사용한 방식에 따라 결합키를 생성(Salt값은 결합키관리기관이 제공)하여 일련번호와 함께 결합키관리기관에 전달하고, 결합 후 반출된 반출정보에 포함된 1)의 정보를 활용하여 내부에서 연계하여 활용

※ 시계열 분석이 완전히 종료된 경우 이를 결합키관리기관에 통지하여야 하며, 결합키관리기관은 해당 결합 후 보관하고 있는 1)의 생성방법을 삭제

참고5 가명처리 및 결합 목적 증빙 자료 예시

통계작성 계획서

통계명		
대표 참여진	소속	
	담당자명	
통계작성 배경 및 목적		
통계작성 대상자 수		
통계작성 계획 및 방법		
기대효과 및 활용방안		
붙임. 상세 통계작성 계획서 등		

과학적 연구 계획서

연구명		
연구진	소속	
	연구책임자	
연구 배경 및 목적		
예상 연구 기간		
연구 대상자 수		
연구 방법		
연구내용		
기대효과 및 활용방안		
붙임. 상세 연구계획서 등		

공익적 기록보존 계획서

공익적 기록보존명		
대표 참여진 (기록보관 기관)	보관기관명	
	담당자명	
공익적 기록보존 목적		
보존기간		
공익적 기록보존 방법		
내용		
기대효과 및 활용방안		
붙임. 상세 계획서 등		

참고6 결합신청서 작성 방법

결합신청서

신청번호	
접수번호	

① 결합신청자			
기관명		사업자등록번호 또는 법인등록번호	
주소		대표자명	
담당자		담당자 연락처 (전화, e-mail)	
유형	[ ] 개인 [ ] 공공기관 [ ] 비영리법인 [ ] 민간기관		

② 결합 개요		
전체 결합신청기관명	총 기관 수	[ ]개 기관
	기관명	
반복결합	[ ] 해당없음 [ ] 최초 [ ] 추가(결합접수번호 :)	
추가절차 신청	결합률 확인 [ ] 가명정보 추출 [ ] 모의결합 [ ]	

③ 가명정보 제공자		해당없음 [ ]
제출 방법	[ ] 온라인 [ ] 오프라인	
지원 요청 사항	[ ] 결합 신청에 필요한 가명처리	
가명정보 제공 담당자	이름	연락처 (전화, e-mail)

④ 결합정보 이용자		해당없음 [ ]
결합 목적	[ ] 통계작성 [ ] 과학적 연구 [ ] 공익적 기록보존 등	
세부 결합 목적		
분석공간 이용	[ ] 추가 가명처리만 수행 [ ] 결합정보 분석 [ ] 이용안함	
지원 요청 사항	[ ] 반출 전 처리 [ ] 분석	
결합정보 이용 담당자	이름	연락처 (전화, e-mail)

「개인정보 보호법」 제28조의3제1항 및 같은 법 시행령 제29조의3제1항에 따른 결합을 위하여 결합전문기관에 결합신청서를 위와 같이 제출합니다.

년 월 일

⑤ 결합신청기관 (서명 또는 인)

(결합전문기관명)의 장 귀하

첨부 서류	1. 사업자등록증, 법인등기부등본 등 결합신청자 관련 서류 1부 2. 결합 목적을 증명할 수 있는 서류 1부(결합된 정보를 반출하려는 자에 한함) 3. 결합 대상 가명정보에 관한 서류(전체 항목명, 가명처리 대상 항목명*, 가명처리 내역 등**) 1부(가명정보 제공자에 한함) * 결합키 생성에 사용된 항목 제외 ** 결합대상정보가 확정된 이후에 제출

① 결합신청자

▶ 기관명, 사업자등록번호 또는 법인등록번호, 대표자명 : 결합신청자가 개인인 경우 공란 허용

▶ 담당자: 결합전문기관과 협의를 담당하는 자*

* 예시) 가명정보 제공 : 데이터 보유부서 책임자, 결합정보 이용 : 연구 책임자

※ 결합신청서의 '담당자'와 반출신청서의 '담당자'는 다른 사람으로 작성해도 무관

예시) '가명정보 제공 + 결합정보 이용'의 경우 결합신청서에는 데이터 보유부서 책임자를, 반출신청서에는 연구책임자를 담당자로 작성 가능

② 결합 개요

▶ 전체 가명정보 제공자명(총수): 해당 결합을 신청하는 가명정보 제공자의 전체 기관명 및 전체 기관수(총 OO개)를 작성

※ 결합신청자간 서로 협의한 내용이 정확한지 확인 및 해당 결합 건의 전체 결합신청자 수를 파악하기 위한 목적이므로, 기관이 너무 많은 경우 대표기관명만 나열 후 전체 기관수를 기재할 수 있음

▶ 반복결합: 추가일 경우 결합접수번호가 누락되지 않도록 주의

▶ 추가절차 신청: 모의결합을 지원하지 않는 경우, 다른 결합전문기관에 신청하도록 안내하거나 모의결합 진행 없이 결합으로 진행할 수 있도록 안내

③ 가명정보 제공자

▶ 지원 요청 사항: '결합 전 가명처리'를 지원하지 않는 경우, 다른 결합전문기관에 신청하도록 안내하거나 컨설팅 안내 등 추가적으로 협의를 통해 결합을 진행할 수 있도록 안내

▶ 가명정보 제공 담당자: 데이터 보유부서 책임자

※ 전체 담당자와 동일한 경우 공란 허용하며, '가명정보 제공 + 결합정보 이용'의 경우 데이터 보유부서 책임자 및 연구 책임자를 분리하여 작성 필요

④ 결합정보 이용자

▶ 세부 결합 목적: 결합 목적 및 필요 정보를 알 수 있을 정도로 구체적으로 작성

▶ 분석공간 이용: 분석공간 이용기간에 대해서 결합 후 별도로 협의하여 이용할 수 있도록 안내

▶ 지원요청사항: '반출 전 처리', '분석'을 지원하지 않는 경우, 다른 결합전문기관에 신청하도록 안내하거나 컨설팅 안내 등 추가적으로 협의를 통해 결합을 진행할 수 있도록 안내

▶ 결합정보 이용 담당자: 연구 책임자

※ 전체 담당자와 동일한 경우 공란 허용하며, '가명정보 제공 + 결합정보 이용'의 경우 데이터 보유부서 책임자 및 연구 책임자를 분리하여 작성 필요

⑤ 결합신청자의 서명

▶ 결합신청자가 개인이 아닌 경우 개인정보처리자 직인이 원칙, 다만 서면(직인 날인) 외에 전자문서로 제출할 수 있음을 안내

유의사항

- 결합신청자가 개인이 아닌 경우 전자문서로 제출시 결합신청자 본인이 직접결재가 불가하며, 차상위 의사결정권자 이상의 결재 필요
  * 1인 사업장인 경우 허용
- 전자문서 내 직인이 없는 경우 보완 요청해야 함

참고7 반출신청서 작성 방법

① 반출신청서

②

반출접수번호	
결합접수번호	

결합신청자			
기관명	A사	사업자등록번호 또는 법인등록번호	○○○-○○-○○○○○
주소	○○시 ○○구 ○○○	대표자명	○○○
담당자	홍길동	담당자 연락처 (전화, e-mail)	010-○○○○-○○○○ ○○○○○@○○○○○.○○.○○

③ 결합 유형	
반복결합	[ ] 최초 [ ] 추가

④ 반출 개요	
파일명	ccdab
반출 목적	[ ] 통계작성 [∨] 과학적 연구 [ ] 공익적 기록보존 등
세부 반출 목적	구체적 목적 설명
반출정보 유형	[∨] 가명정보 [ ] 법 제58조의2에 해당하는 정보(익명정보)
제공 받는 방법	[∨] 온라인 [ ] 오프라인 [ ] 결합전문기관 내 분석공간
지원 요청 사항	[∨] 반출된 정보의 분석 [∨] 개인정보 보호 교육

「개인정보 보호법」 제28조의3제2항 및 같은 법 시행령 제29조의3제3항·제6항,
「가명정보의 결합 및 반출 등에 관한 고시」 제10조제3항에 따라 결합된 정보를 반출하기 위하여 결합전문기관에 반출신청서를 위와 같이 제출합니다.

○○○○년 ○○월 ○○일

결합신청자 (서명 또는 인)

결합전문기관의 장 귀하

⑤ 첨부 서류	1. 반출 대상 정보에 관한 서류 1부(추가적인 서류 제출이 필요한 경우에 한함) 2. 반출 목적을 증명할 수 있는 서류 1부(추가적인 서류 제출이 필요한 경우에 한함) 3. 반출정보의 안전조치계획 및 이를 증빙할 수 있는 서류 1부

① 반출신청서 작성

▶ (작성 주체) 결합정보 또는 분석결과 등을 결합전문기관 외부로 반출하고자하는 자는 반출 신청서를 작성하여 결합전문기관에 제출

※ 가명정보를 제공만 하는 자와 결합정보를 결합전문기관 내의 분석공간에서 분석만을 수행하는 자는 반출신청서를 작성하지 않아도 됨

② 반출접수번호 및 결합접수번호

▶ (반출접수번호) 결합전문기관이 제출된 반출신청서를 접수할 때 발행하는 번호, 반출신청서 제출 시 공란으로 제출

▶ (결합접수번호) 결합전문기관이 제출된 결합신청서를 접수할 때 발행하였던 번호

③ 결합 유형

▶ 반복결합을 신청한 자는 최초/추가 여부를 체크하며, 반복결합이 아닌 경우 공란

④ 반출 개요

▶ 결합정보를 반출하려는 결합신청자는 파일명(반출할 결합 결과물), 반출 목적, 반출정보 유형 등을 작성

※ 결합신청자는 반출심사 전인 경우 결합 목적과 반출 목적 변경 가능

- (반출정보 유형) 반출을 신청하려는 자가 정보의 형태 등을 고려하여 가명 또는 익명으로 판단하여 표기
- (제공받는 방법) 결합전문기관등이 제공하는 시스템을 통해 제공받는 경우는 온라인, USB등 저장장치를 이용해 반출하는 경우는 오프라인, 결합전문기관이 제공하는 분석공간을 이용하는 경우로 구분하여 표기
- (지원 요청사항) 정보의 분석을 위해 결합전문기관의 지원이나 가명정보의 처리에 관한 교육이 필요하면 표기

⑤ 첨부 서류

▶ 추가적인 서류 제출이 필요한 경우에 한하여 모든 서류를 제출

- (반출 대상 정보에 관한 서류) 분석공간을 통해 추가 가명처리가 수행되어 반출 대상 정보가 당초 제출한 결합 대상 가명정보와 상이한 경우에만 제출하고 동일한 경우에는 생략 가능
- (반출 목적 관련 서류) 반출 목적이 당초의 결합 목적과 달라진 경우(결합 목적과 반출 목적의 양립 가능성 검토 필요)만 제출하고 동일한 경우에는 생략 가능
- (안전조치 계획) 개인정보 처리방침, 내부 관리계획, 운영 지침 등 반출정보의 안전조치와 관련된 자료를 제출

참고8 내부 관리계획 작성 예시

제○○조(가명정보 및 추가정보 관리책임자 지정) ① 개인정보 보호책임자는 다음과 같은 역할을 수행한다.

1. 가명정보에 대한 내부 관리계획의 수립·시행
2. 내부 관리계획의 이행실태 점검 및 관리
3. 가명처리 및 적정성 검토 현황 관리
4. 가명정보 및 추가정보에 대한 관리·감독
5. 가명정보 처리 현황 및 관련 기록 관리
6. 가명정보를 처리하는 자 교육계획의 수립 및 시행
7. 가명처리 및 가명정보 처리 위탁 사항에 대한 관리·감독(해당 시)
8. 가명정보에 대한 재식별 모니터링 및 재식별 시 처리 방안의 수립·시행
9. 그 밖의 가명정보 처리에 대한 보호에 관한 사항

제○○조(가명정보 및 추가정보의 분리보관) ① 가명정보는 가명처리가 완료되면 가명처리 전 개인정보와 분리·보관하여야 한다.

② 가명처리의 과정에서 발생하는 추가정보는 가명정보와 분리·보관하여야 한다.

③ 가명처리 전 개인정보, 가명정보 및 추가정보는 물리적으로 분리 보관하는 것을 원칙으로 하며 물리적 보관이 어려운 경우 논리적인 분리를 시행할 수 있다.

④ 논리적으로 분리·보관하는 경우 엄격한 접근통제를 적용해야 한다.

제○○조(가명정보 및 추가정보에 대한 접근권한 분리) ① 가명처리가 완료되면 가명정보 또는 추가정보의 접근권한은 최소한의 인원으로 엄격하게 통제하여야 하며, 업무에 따라 차등적으로 부여 하여야 한다.

② 추가정보에 대한 접근권한과 가명정보에 대한 접근권한은 분리하여 관리해야 한다.

③ 가명정보 또는 추가정보에 대한 접근권한 부여, 변경 또는 말소에 대한 내역을 기록하도록 하고 이 기록은 최소 3년간 보관하여야 한다.

제○○조(가명정보 및 추가정보의 안전성 확보조치) ① 가명정보와 추가정보는 개인정보보호법 및 동법 시행령에서 요구하는 안전성 확보조치를 수행하여야 한다.

② 추가정보에 특별한 이유가 없는 한 생성 즉시 삭제하도록 한다. 단, 시계열 분석 등의 이유로 추가정보가 필요한 경우 저장 시 암호화하여 저장하여야 한다.

제○○조(가명정보를 처리하는 자의 교육) ① 가명정보 관리책임자는 가명정보를 처리하는 자에게 필요한 가명정보 보호 교육계획을 수립하고 실시하여야 한다.

② 가명정보 보호 교육은 다음과 같은 내용을 포함하여 시행하여야 한다.

1. 가명정보 처리에 관한 사항
2. 가명정보 및 추가정보의 안전조치에 관한 사항
3. 재식별 금지에 관한 사항

③ 가명정보를 처리하는 자에 대한 교육은 개인정보 보호교육과 함께 수행할 수 있으며 교육을 실시한 결과 또는 이를 입증할 수 있는 관련 자료 등을 기록·보관하여야 한다.

제○○조(가명정보 처리 기록 작성 및 보관) ① 가명정보의 처리 시 다음과 같은 사항에 대해 가명정보 처리 대장에 기록을 작성하여 보관하여야 한다.

1. 가명정보의 처리 목적
2. 가명처리한 개인정보의 항목
3. 가명정보의 이용내역
4. 제3자 제공 시 제공받는 자
5. 그 밖에 가명정보의 처리 내용을 관리하기 위하여 개인정보보호위원회가 필요하다고 인정하여 고시하는 사항

제○○조(개인정보 처리방침 공개) ① 가명정보 처리와 관련하여 아래와 같은 내용을 개인정보 처리방침에 포함하여 공개하여야 한다.

1. 가명정보의 처리 목적
2. 가명정보 처리기간(선택)
3. 가명정보 제3자 제공에 관한 사항(해당 시)
4. 가명정보 처리 위탁에 관한 사항(해당 시)
5. 처리하는 가명정보의 항목
6. 가명정보의 안전성 확보조치에 관한 사항

제○○조(가명정보의 재식별 금지) ① 가명정보를 처리하는 자의 가명정보에 대한 재식별 행위는 엄격하게 금지한다.

② 가명정보를 처리하는 자는 가명정보를 처리하는 중 특정 개인에 대한 재식별이 발생하는 경우 즉시 처리를 중단하고 이를 가명정보 관리책임자에게 통보한 후 수립된 재식별 시 처리 방안에 따라 즉시 조치하여야 한다.

참고9 적정성 검토 관련 서식 예시

※ 아래 참고자료는 적정성 검토 시 필요한 사항들에 대한 이해를 돕기 위해 작성한 자료로써 표준이 아니며, 자체적으로 제작하여 다른 자료로 활용할 수 있음

가명정보 이용·제공 신청서

접수번호		접수일	

신청자	조직/부서명			
	담당자 직위		담당자 성명	
	전화번호		이메일 주소	

처리 목적		☐ 통계작성 ☐ 과학적 연구 ☐ 공익적 기록보존
		세부내용 :
활용 형태		☐ 내부이용 ☐ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체		☐ 동일 개인정보처리자 ☐ 제3자(제공받는 자) 제공
처리 환경		☐ 내부 ☐ 외부
처리 장소		☐ 폐쇄 환경 ☐ 제한 없음
반복 제공 여부		☐ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☐ 온라인 ☐ 오프라인
제공 받는 자		(내부) ☐ 동일부서 ☐ 타부서 (외부) ☐ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	
	데이터 내역	
	가명정보 이용 기간	년 월 일~ 년 월 일(개월)

위와 같이 가명정보 이용·제공을 신청합니다.

년 월 일

신청인(부서장명) (서명 또는 인)

첨부서류	1. 개인정보 유형 분류표 2. 활용데이터 요구 수준표(선택) 3. 데이터 처리 위탁 계약서(위탁 계약의 경우) / 추후 제출 4. 데이터 제공 계약서(제3자 제공의 경우) / 추후 제출

개인정보 유형 분류표

순번	항목명	개인정보유형	비고
1		개인식별정보/ 개인식별가능정보	민감성정보 / 비민감성정보 등
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

활용데이터 요구 수준표

순번	항목명	요구 수준	비고
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

가명정보 처리 기초자료 명세서

신청 기관 정보			
기관명			
주소			
데이터명		평가 목적	
데이터 수집			
이용 방법			
이용기간	년 월 일~ 년 월 일		

데이터 명세		
번호	구분	검토사항
1	데이터 특징	
2	데이터 생성 방법	
3	데이터 제공방법	
4	데이터 관리 환경	

적정성 검토 결과서(위원용)

접수번호						
검토위원 정보	성명		소속		직위	

검토 대상	☐ 신규 ☐ 보완	
검토 일자	년 월 일 ~ 년 월 일	
최종검토결과	☐ 적정(승인) ☐ 조건부 승인 ☐ 부적정(반려)	
세부결과	가명정보 목적 적합성	☐ 적합 ☐ 미흡
	가명정보 이용항목 적합성	☐ 적합 ☐ 미흡
	식별위험성 검토 결과 보고서 적정성	☐ 적합 ☐ 미흡
	가명처리 방법 및 수준 정의표 적정성	☐ 적합 ☐ 미흡
	처리 수준에 따른 처리 결과의 정확성	☐ 적합 ☐ 미흡
	처리 결과의 목적 달성 가능성	☐ 적합 ☐ 미흡
종합검토의견	※ 검토 결과가 조건부 승인인 경우 보완사항을 부적정인 경우 사유를 상세히 기재	

위와 같이 적정성 검토 결과를 통지합니다.

년 월 일

서명란	

적정성 검토 종합결과서

이용신청 접수번호	
검토 대상	☐ 신규 ☐ 보완
검토 일자	년 월 일 ~ 년 월 일
최종검토결과	☐ 적정(승인) ☐ 조건부 승인 ☐ 부적정(반려)
종합검토의견	

위와 같이 적정성 검토 결과를 통지합니다.

년 월 일

서명란	위원장	검토위원	검토위원	사내 개인정보보호 책임자
	이름 (인)	이름 (인)	이름 (인)	이름 (인)

비밀유지의무 서약서

본인은 가명처리 적정성 검토와 관련한 활동으로 얻어진 모든 정보에 대하여 ○○○○○○○의 허락 없이 외부에 공개하지 않을 것을 서약합니다. 본 서식에 서명함으로써, 본인은 정보의 비밀을 지키기 위해 합당한 역할과 완전한 책임을 다 할 것에 동의합니다.

서명일 : _____

소　속 : _____

성　명 : _____

서　명 : _____

○○○○○○○○○장 귀하

이해상충 서약서

접수번호	
가명정보 이용신청자명	
적정성 검토 회의명	

본인은 상기 적정성 검토와 관련하여 아래와 같이 적정성 검토 대상 가명정보 및 가명정보 이용신청자과 이해관계가 없음을 서약합니다.

	이해 관계 내용	예	아니오
1	적정성 검토 대상 가명정보를 이용할 예정이 있는지		
2	적정성 검토 대상 가명정보 활용에 대한 경제적·비경제적 이익을 가지고 있는지		
3	가명정보 이용신청자와 고용관계(상근, 비상근/ 공식, 비공식 등)에 있는지		
4	가명정보 이용신청자로부터 본 적정성 검토 비용 외에 검토 결과에 영향을 미칠 수 있는 경제적·비경제적 이익을 제공받은 사실이 있는지		
5	본인 또는 배우자의 직계가족이 소속된 회사가 위에서 기술된 것과 같은 관계를 가지고 있는지		
6	그 밖에 적정성 검토 대상 가명정보 또는 가명정보 이용신청자와 이해관계가 있는지		

본인이 확인한 모든 내용은 정확히 기술되었으며 만약 평가 진행 중에 의뢰기관에 대한 이해관계가 변동되는 이해상충이 생기는 경우 이를 인지한 날로부터 5영업일 이내에 XXX에 통지하겠습니다.

년 월 일

서약자 : (인)

가명정보에 대한 안전조치 이행 확약서

본 기관은 「개인정보 보호법」에서 규정하고 있는 가명정보에 대한 안전조치의무 등(제28조의4)및 가명정보에 대한 안전성 확보 조치(시행령 제29조의5)를 성실히 이행하고 기타 관련 법령을 준수하였습니다.

아울러, 이를 이행·준수하지 아니하여 발생하는 관련 법적 책임을 부담할 것을 확약합니다.

년 월 일

신청기관 (직인)

참고10 가명정보 처리 관련 실무 서식 예시

가명정보 관리대장

(여러 가명처리 기록 서식)

일련번호	이용 신청 접수번호	이용신청자 (부서/이름)	이용신청기간	결합 유무	처리목적	제공 받는자 (제3자제공 포함)	파기| 일자	대장 기록자	기록 확인자
1		/	~						
	가명처리 개인정보항목				가명정보 파일명				
2		/	~						
	가명처리 개인정보항목				가명정보 파일명				
3		/	~						
	가명처리 개인정보항목				가명정보 파일명				
4		/	~						
	가명처리 개인정보항목				가명정보 파일명				
5		/	~						
	가명처리 개인정보항목				가명정보 파일명				

추가정보 관리대장

신청서 접수번호	이용·제공 신청기관	신청자	보유 기간	생성		이용			파기		분리보관 확인자	최종 확인자
				생성일	담당자	이용일	사유	담당자	파기일	담당자		

가명정보 접근 권한 관리대장

신청접수번호	담당자	책임자
가명정보 관리대장 번호 동일		

번호	소속(부서)	사용자명	사용자 IP	가명정보 관련 업무목적	권한 범위	권한 부여일자	권한 말소일자	비고
				분석/내부결합 등	열람/수정/삭제 등			

추가정보 접근 권한 관리대장

신청접수번호	담당자	책임자
가명정보 관리대장 번호 동일		

번호	소속(부서)	사용자명	사용자 IP	추가정보 관련 업무목적	권한 범위	권한 부여일자	권한 말소일자	비고
					열람/삭제 등			

가명정보 파기대장

일련 번호	가명정보 이용 신청번호	파기 사유	가명정보 파기일자	가명정보 파기 방법	파기 신청자	파기 수행자	파기 확인자
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

추가정보 파기대장

일련 번호	가명정보 이용 신청번호	파기 사유	추가정보 파기일자	추가정보 파기 방법	파기 신청자	파기 수행자	파기 확인자
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

비정형데이터 대상 가명처리 결과에 대한 자체 검증 결과서

<table>
<tr><td rowspan="10">검증 대상
데이터 명세</td><td colspan="3">개요</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td>데이터 유형</td><td colspan="2">이미지, 영상, 음성, 텍스트 등</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td colspan="2">JPG, MP4, TXT 등</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td colspan="2">JPG, MP4, TXT 등</td></tr>
<tr><td>데이터 규모</td><td colspan="2">이미지 장수, 동영상 클립수, 발화정보 개수, 글자 수,
기타(러닝타임, 해상도, FPS 등) 등</td></tr>
<tr><td>데이터 크기(용량)</td><td colspan="2">00GB, 00MB 등</td></tr>
<tr><td>대상 데이터 항목명</td><td colspan="2">복수 기재 가능</td></tr>
<tr><td>가명처리 적용 기술</td><td colspan="2">데이터 필터링(블러링) 등 항목이 여러 개인 경우
항목별로 기술</td></tr>
<tr><td>자체 검증 기간</td><td colspan="3">20 년 월 일 ~ 20 년 월 일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="3"></td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="3">육안 전수 검사 등 자체 검증 과정과 방법 기록</td></tr>
<tr><td>자체 검증 결과</td><td colspan="3">확인 결과 이상 없음 등</td></tr>
<tr><td rowspan="4">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

부록 2

정형데이터 가명처리 시나리오 예시

유통분야

A유통업체는 상품군별 판매추이에 대한 통계작성을 위한 데이터 제공을 요청받아
B통계전문업체에게 다음과 같이 판매 데이터를 제공하려 한다.

☑ 데이터의 이용 목적

- 코로나 이전과 코로나 이후의 상품군별 판매추이에 대한 통계작성으로 지속적인 분석을 위해 2019년 1월 부터 12월까지의 상품 매출 데이터와 2021년 1월부터 12월까지의 상품 매출 데이터를 제공하기로 함

☑ 데이터 특징

- 동일 기간의 고객 구매 데이터를 제공 목적에 맞게 원본데이터를 가공하여 제공
- 전체 820만명의 고객 중 50%를 샘플링하여 410만건을 제공

☑ 데이터의 이용 환경

- 데이터 이용자는 B통계전문업체 임
- B통계전문업체와는 데이터 제공에 대한 계약을 맺어 제공함
- B통계전문업체는 다양한 통계분석을 통해 통계정보를 만들어 판매하는 회사임
- B통계전문업체는 기존에 개인정보의 유출사고로 인해 과태료 처분을 받은 적이 있으며 그 이후 개인정보의 보호를 위한 부분에 많은 투자를 하여 시스템적으로 안전한 환경을 구축하였으며 ISMS-P인증을 취득한 상태임
- 개인정보의 분석환경은 인터넷과 분리된 환경에서 별도의 네트워크를 운영하고 있음
- B통계전문업체는 가명정보에 대한 내부 관리계획을 기존의 개인정보 내부 관리계획에 추가하는 방식으로 수립하여 운용하고 있음
- 모든 연구원에 대해 보안서약서를 제출받았으며 B통계전문업체 명의의 보안 확약서도 제출받았음
- 관련 연구원들은 모두 10년이상 통계를 생성하는 분석업무를 수행한 인력으로 구성되어 있음
- 제공하는 데이터에는 직장이 B통계전문업체 인력도 포함되어 있는 것이 확인되었음

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	ZZANG-20220522001	접수일	2022년 5월 22일

신청자	조직/부서명	A 유통회사 구매부		
	담당자 직위	과장	담당자 성명	이유통
	전화번호	02-1234-5678	이메일 주소	gamyeongzzang@utong.com

처리 목적	☑ 통계작성 ☐ 과학적 연구 ☐ 공익적 기록보존
	코로나 이전과 코로나 이후의 상품군별 판매 추이에 대한 통계 작성을 위한 가명정보의 제공
활용 형태	☐ 내부이용 ☑ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체	☐ 동일 개인정보처리자 ☑ 제3자(제공받는 자) 제공
처리 환경	☐ 내부 ☐ 외부
처리 장소	☑ 폐쇄 환경 ☐ 제한 없음
반복 제공 여부	☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법	☐ 온라인 ☑ 오프라인
제공 받는 자	(내부) ☑ 동일부서 ☐ 타부서 (외부) ☐ 제3자, 결합전문기관 제공

신청 명세	가명처리 대상 명칭	자사의 고객 구매정보
	데이터 내역	구매부에서 보관하고 있는 자사 고객 중 회원가입이 2019년 1월 이전인 고객의 구매정보 중 2019년 1월 ~ 2019년 12월, 2021년 1월 ~ 2021년 12월 까지의 구매정보데이터 전체 222컬럼, 전체 대상인원 820만명 중 50% 무작위 샘플링한 410만명의 정보
	가명정보 이용 기간	2022년 4월 1일 ~ 2022년 10월 31일(7개월)

위와 같이 가명정보 이용·제공을 신청합니다.

2022년 5월 22일

신청인(부서장명) 이 공 재 (서명 또는 인)

첨부서류	1. 개인정보 유형 분류표 2. 활용데이터 요구 수준표(선택) 3. 데이터 처리 위탁 계약서(위탁 계약의 경우) / 추후 제출 4. 데이터 제공 계약서(제3자 제공의 경우) / 추후 제출

가명정보에 대한 안전조치 이행 확약서

본 기관은 「개인정보 보호법」에서 규정하고 있는 가명정보에 대한 안전조치의무 등(제28조의4)및 가명정보에 대한 안전성 확보 조치(시행령 제29조의5)를 성실히 이행하고 기타 관련 법령을 준수하였습니다.

아울러, 이를 이행·준수하지 아니하여 발생하는 관련 법적 책임을 부담할 것을 확약합니다.

2022년 5월 22일

신청기관 B통계회사 (직인)

② 보호법에서 정한 목적 중에서 가명정보 처리 목적을 명확히 설정하였는지 검토

통계작성 계획서

통계명	2022_COVID19_SalesTrend_410	
대표 참여진	소속	B 통계회사
	담당자명	이통재
통계작성 배경 및 목적	COVID19는 전세계적으로 펜데믹을 불러왔다. 이에 따라 각종 산업이 위축되고 개인에 대한 거리두기로 인해 다양한 산업에서 변화를 가지고 왔으며 특히 개인의 소비 생활에도 많은 영향을 미치게 되었다. 이에 따라 코로나 이전과 코로나 이후의 개인별 판매 데이터를 분석하여 COVID19로 인해 개인에게 미치는 영향에 대한 파악 등 다양한 방면에 사용하기 위한 통계를 작성하려 한다.	
통계작성 대상자 수	410만명(2019년 1월~2021년 12월까지 구매내역이 있는 고객 중 50%를 무작위 샘플링)	
통계작성 계획 및 방법	동일 집단의 COVID19발병 이전과 발병이후의 제품의 판매 추이를 통해 COVID19가 개인의 소비 패턴의 변화에 어떤 영향을 주었는지를 파악하기 위한 통계를 작성	
기대효과 및 활용방안	코로나와 같은 질병이 발생할 때 소비의 변화에 따라 지원금의 처리 방법, 지원 규모와 생산 산업군에 대한 영향 등을 파악하여 이를 보완하기 위한 정책 연구 등에 사용	
붙임. 상세 통계작성 계획서 등		

③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

개인정보 유형 분류표

순번	항목명	개인정보유형	비고
1	고객ID	개인식별정보	고객의 개인별 구분값
2	나이	개인식별가능정보	생년월일에서 추출한 정보
3	주소	개인식별가능정보	배송지 주소
4	성별	개인식별가능정보	
5	2019년 1월 여행용품구매액	개인식별가능정보	2019년 1월 여행용품 군의 구매 총액
6	2019년 1월 식품류 구매액	개인식별가능정보	2019년 1월 식품 군의 구매 총액
7	2019년 1월 의류 구매액	개인식별가능정보	2019년 1월 의복 군의 구매 총액
8	2019년 1월 취미용품구매액	개인식별가능정보	2019년 1월 취미용품 군의 구매 총액
9	2019년 1월 생활용품구매액	개인식별가능정보	2019년 1월 생활용품 군의 구매 총액
10	2019년 1월 유아용품구매액	개인식별가능정보	2019년 1월 유아용품 군의 구매 총액
11	2019년 1월 기타 구매액	개인식별가능정보	2019년 1월 6가지 주요 제품군이 아닌 제품의 구매 총액
12	2019년 1월 구매 총금액	개인식별가능정보	2019년 1월 구매 총 금액
13	2019년 1월 선호 제품군	개인식별가능정보	2019년 1월 가장 많이 구매한 제품군
~	~		
217	2021년 12월 생활용품구매액	개인식별가능정보	2021년 12월 생활용품 군의 구매 총액
218	2021년 12월 유아용품구매액	개인식별가능정보	2021년 12월 유아용품 군의 구매 총액
219	2021년 12월 기타 구매액	개인식별가능정보	2021년 12월 6가지 주요 제품군이 아닌 제품의 구매 총액
220	2021년 12월 구매 총금액	개인식별가능정보	2021년 12월 구매 총 금액
221	2021년 12월 선호 제품군	개인식별가능정보	2021년 12월 가장 많이 구매한 제품군
222	2021년 고객 등급	개인식별가능정보	2020년 까지의 구매내역을 통해 산정한 2021년 고객의 등급 P, G, S, B, F

활용데이터 요구 수준표

순번	항목명	요구 수준	비고
1	고객ID	각 고객이 서로 다른 사람이라는 구분만 가능하면 됨	
2	나이	중학생 이상 90세 미만까지 분석 예정 단 90세 이상의 경우에도 다른 대조군과의 비교를 위해 90세 이상 표기 요청	
3	주소	시군구 단위의 분석 예정이며 이에 따라 시군구 단위의 주소 필요	
4	성별	분석목적에 남녀의 차이에 대한 분석이 포함되어 필요	
5	2019년 1월 여행용품구매액	최소 1만단위 이상의 값에 대해서는 정확한 값이 필요	
6	2019년 1월 식품류 구매액		
7	2019년 1월 의류 구매액		
8	2019년 1월 취미용품구매액		
9	2019년 1월 생활용품구매액		
10	2019년 1월 유아용품구매액		
11	2019년 1월 기타 구매액		
12	2019년 1월 구매 총금액		
13	2019년 1월 선호 제품군	제품군이 명확하게 명시되어야 함	
~	~		
217	2021년 12월 생활용품구매액	최소 1만단위 이상의 값에 대해서는 정확한 값이 필요	
218	2021년 12월 유아용품구매액		
219	2021년 12월 기타 구매액		
220	2021년 12월 구매 총금액		
221	2021년 12월 선호 제품군	제품군이 명확하게 명시되어야 함	
222	2021년 고객 등급	5단계의 고객등급이 구분되어야 함	

식별 위험성 검토 결과보고서

<table>
<tr><td>가명정보 활용목적</td><td colspan="2">▪ B통계업체는 코로나 이전과 코로나 이후의 상품군별 판매 추이에 대한 통계작성을 위해 당사의 구매정보 데이터 중 2019년 1월부터 12월까지의 주요 상품군별 판매액 정보와 2021년 1월부터 12월까지의 주요 상품군별 판매액 정보를 나이와 성별, 시군구 단위의 주소별 데이터를 분석</td></tr>
<tr><td>가명처리 대상 데이터 항목</td><td colspan="2">▪ 고객ID, 나이, 주소, 성별, 2019년 1월~12월, 2021년 1월~12월까지의 여행용품, 식품류, 의류, 취미용품, 생활용품, 유아용품, 기타의 7개 범주의 구매금액의 월별 합계액, 월별 구매 총 금액, 월별 선호 제품군, 각 년도의 고객 등급(전체 222개의 컬럼)
▪ 전체 고객 820만명 중 50%를 무작위 샘플링하여 구성한 410만명에 대한 데이터</td></tr>
<tr><td rowspan="3">데이터 위험성</td><td>식별성 유무</td><td>▪ ‘고객 ID’는 개인식별정보임
▪ ‘나이’, ‘주소’, ‘성별’은 조합했을 때 개인의 식별이 가능한 개인식별 가능정보임</td></tr>
<tr><td>특이정보 유무</td><td>▪ 각 범주별 구매금액의 경우 특이정보로 인한 개인 식별성이 발생할 수 있음</td></tr>
<tr><td>재식별시 영향도</td><td>▪ 단순 고객의 구매데이터로 재식별 시 영향도는 크지 않을 것으로 판단됨</td></tr>
<tr><td rowspan="3">처리 환경 검토</td><td>이용 및 제공 형태</td><td>▪ 제3자 제공
- 데이터 제공 계약을 체결하여 데이터를 제공
- 데이터 제공 계약에는 재제공 금지, 목적 달성 후 삭제, 재식별 금지 및 재식별 시 조치에 관한 사항들이 포함되어 있음
▪ B통계업체는 개인정보(가명정보)처리시스템에 대한 ISMS-P인증을 취득하고 있음</td></tr>
<tr><td>처리 장소</td><td>▪ B통계업체에서 가명정보는 인터넷에 접근할 수 없는 차단된 별도의 분석 PC에서 분석 예정
▪ 분석PC가 있는 환경은 별도의 분석실로 내부적인 출입통제를 적용하는 것으로 파악됨</td></tr>
<tr><td>다른 정보와의 결합 가능성</td><td>▪ B통계업체는 다양한 통계를 생성하는 업체로 유사 업종에 대한 통계정보 등 결합 가능성이 있는 정보를 보유하고 있음</td></tr>
<tr><td>최종 검토의견*</td><td colspan="2">▪ 해당 연구는 자사의 데이터를 B통계업체에 제공하는 것으로 데이터 자체 위험성과 처리 환경 위험성을 검토할 때 다음과 같은 조치가 필요함
- 통계전문업체의 특성 상 다른 정보의 결합가능성이 있으나 처리 장소와 개인정보 보호 수준을 검토할 때 결합에 대한 시도는 거의 없을 것으로 판단됨
- ‘고객ID’는 개인식별정보로 개인식별 가능성이 매우 높으며 이에 따라 다시 원래의 정보로 대체할 수 없는 Salt값이 포함된 해시처리 등의 기법의 적용이 필요
- ‘나이’, ‘주소’, ‘성별’은 그대로 사용하는 경우 조합에 의한 개인식별 가능성이 있으며 이에 따라 다음과 같은 처리가 필요
• ‘나이’ : 물품의 주 구매대상이 아닌 20세 미만의 경우 삭제처리가 필요하며 그 외의 나이에 대해서는 일반적인 구매 분석 통계에 사용되는 10살 단위로 제공하며 80세 이상의 나이에 대해서는 80세 이상으로 처리하는 것이 필요
• ‘주소’ : 동단위와 상세 주소의 경우 통계목적에 필요하지 않기 때문에 삭제하며 시군구 단위의 주소까지만 사용하는 것이 필요
• ‘성별’ : 성별은 분석목적에 필요하므로 그대로 사용
▪ 구매액 관련 정보들은 구매금액별 특이정보를 검토하여 구매 금액에 대한 적절한 수준의 상단 코딩을 적용(19년 8월 취미용품 114,562,000원 → 1억원 이상)하고 금액에 대해서 라운딩을 적용하는 것이 필요
▪ 고객등급의 경우 식별성이 높은 VIP, S, A를 하나로 B, C를 하나로 D, E, F를 하나로 묶을 필요가 있음</td></tr>
</table>

* 최종 검토의견은 외부전문가에게 자문 및 작성을 요청할 수 있음

④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

항목별 가명처리계획

순번	항목명	개인정보유형	제3자 제공	
			처리 방법	처리 수준
1	고객ID	개인식별정보	대체	- 일련번호 대체
2	나이	개인식별가능정보	범주화	- 10살 단위 범주화
			상하단 코딩	- 20세 미만 삭제 - 80세 이상은 80세 이상 경계치 입력
3	주소	개인식별가능정보	부분삭제	- 동단위 이하 삭제
4	성별	개인식별가능정보	처리 없음	
5	2019년 1월 여행용품구매액	개인식별가능정보	범주화	- 상단 99.9%를 초과하는 경우 경계치로 변경 - 금액은 다음과 같이 범주화 적용 - 0원 : 0원 - 10만단위 미만 : 1만단위로 라운드 업 - 1,000만 단위 미만 : 10만 단위로 라운딩 - 1,000만 단위 이상 100만 단위로 라운딩
6	2019년 1월 식품류 구매액	개인식별가능정보		
7	2019년 1월 의류 구매액	개인식별가능정보		
8	2019년 1월 취미용품구매액	개인식별가능정보		
9	2019년 1월 생활용품구매액	개인식별가능정보		
10	2019년 1월 유아용품구매액	개인식별가능정보		
11	2019년 1월 기타 구매액	개인식별가능정보		
12	2019년 1월 구매 총금액	개인식별가능정보	범주화	- 각 구매액과 동일한 처리
13	2019년 1월 선호 제품군	개인식별가능정보	그대로 사용	
~	~	~	~	
214	2021년 12월 식품류 구매액	개인식별가능정보	범주화	- 상단 1.5IQR을 넘는 값은 평균값으로 상단 코딩 - 금액은 다음과 같이 범주화 적용 - 0원 : 0원 - 10만단위 미만 : 1만단위로 라운드 업 - 1,000만 단위 미만 : 10 만 단위로 라운딩 - 1,000만 단위 이상 100만 단위로 라운딩
215	2021년 12월 의류 구매액	개인식별가능정보		
216	2021년 12월 취미용품구매액	개인식별가능정보		
217	2021년 12월 생활용품구매액	개인식별가능정보		
218	2021년 12월 유아용품구매액	개인식별가능정보		
219	2021년 12월 기타 구매액	개인식별가능정보		
220	2021년 12월 구매 총금액	개인식별가능정보	범주화	- 각 구매액과 동일한 처리
221	2021년 12월 선호 제품군	개인식별가능정보	그대로 사용	
222	2021년 고객 등급	개인식별가능정보	범주화	- 다음과 같이 범주화 처리 - VIP, S, A → 1등급 - B, C → 2등급 - D, E, F → 3등급

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

가명정보 처리 기초자료 명세서

신청 기관 정보			
기관명	B통계전문회사		
주소	서울시 은평구 통일로 2368		
데이터명	D2468-20210325.txt	평가 목적	통계작성목적의 데이터 제3자 제공
데이터 수집	2019년 1월부터 12월, 2021년 1월부터 12월까지의 고객의 구매정보를 수집		
이용 방법	코로나 이전과 코로나 이후의 상품군별 판매 추이에 대한 통계작성을 위해 당사의 구매정보 데이터 중 2019년 1월부터 12월까지의 주요 상품군별 판매액 정보와 2021년 1월부터 12월까지의 주요 상품군별 판매액 정보를 나이와 성별, 시군구 단위의 주소별 데이터를 분석		
이용기간	2022년 2월 1일 ~ 2022년 3월 31일		

데이터 명세		
번호	구분	검토사항
1	데이터 특징	전체 고객의 구매정보를 7가지 제품군(여행용품, 식품류, 의류, 취미용품, 생활용품, 유아용품, 기타)으로 분류하여 각 군별 구매금액을 계산한 데이터
2	데이터 생성 방법	전체 고객의 구매정보 820만명 중 50%를 무작위 샘플링하여 410만명을 추출하여 생성 각 제품의 구매금액을 위의 7가지 제품군으로 분류하고 각 제품군의 월별 구매금액의 합계를 계산하여 생성
3	데이터 제공방법	제3자 제공 계약을 통해 제공에 대한 근거를 마련 데이터 제공은 1회성으로 제공하며 제공 시 암호화와 관련 보호조치를 적용한 HDD에 데이터를 복사하여 제공할 예정, 암호화는 한국인터넷진흥원의 안전한 암호 가이드라인에 따라 안전한 방식으로 암호화 하며 암호화에 사용한 키는 별도의 경로를 통해 제공할 예정임 데이터 제공 후 당사의 데이터는 B통계업체의 초기 확인을 거쳐 문제가 없는 경우 바로 절차에 따라 파기할 예정
4	데이터 관리 환경	제공한 가명정보는 인터넷에 접근할 수 없는 차단된 별도의 분석 PC에서 분석 예정 분석 PC가 있는 환경은 별도의 분석실로 내부적인 출입통제를 적용 B통계전문업체는 개인정보(가명정보)처리 시스템에 대한 ISMS-P 인증을 취득 B통계전문업체는 개인정보보호법의 관련 조항에서 명시한 개인정보 보호 조치를 수행하고 있으며 내부 관리계획에 가명정보 관련 항목을 추가하여 안전한 관리를 수행하고 있음

예시 원본 데이터 세부 항목별 명세

컬럼명	명세내용
고객ID	EC-XXXXXX
나이	10살 ~ 97살
주소	시(도)군구 도로명 번지
성별	M/F
2019년 1월 여행용품구매액	2019년 1월 여행용품 구매 금액 합계
2019년 1월 식품류 구매액	2019년 1월 식품류 구매 금액 합계
2019년 1월 의류 구매액	2019년 1월 의류 구매 금액 합계
2019년 1월 취미용품구매액	2019년 1월 취미용품 구매 금액 합계
2019년 1월 생활용품구매액	2019년 1월 생활용품 구매 금액 합계
2019년 1월 유아용품구매액	2019년 1월 유아용품 구매 금액 합계
2019년 1월 기타 구매액	2019년 1월 위의 범주에 포함되지 않는 구매 금액 합계
2019년 1월 구매 총금액	2019년 1월 구매 금액 합계
2019년 1월 선호 제품군	위의 7가지 범주 중 가장 높은 구매 금액 합계액의 제품군
~	~
2021년 12월 식품류 구매액	2021년 12월 식품류 구매금액 합계
2021년 12월 의류 구매액	2021년 12월 의류 구매 금액 합계
2021년 12월 취미용품구매액	2021년 12월 취미용품 구매 금액 합계
2021년 12월 생활용품구매액	2021년 12월 생활용품 구매 금액 합계
2021년 12월 유아용품구매액	2021년 12월 유아용품 구매 금액 합계
2021년 12월 기타 구매액	2021년 12월 위의 범주에 포함되지 않는 구매 금액 합계
2021년 12월 구매 총금액	2021년 12월 구매 금액 합계
2021년 12월 선호 제품군	위의 7가지 범주 중 가장 높은 구매 금액 합계액의 제품군
2021년 고객 등급	2021년 고객 등급(VIP, S, A, B, C, D, E, F)

예시 원본 데이터

컬럼명	예시
고객ID	EC-232578
나이	36
주소	서울시 은평구 연서로 29길 245
성별	M
2019년 1월 여행용품구매액	0
2019년 1월 식품류 구매액	248,240
2019년 1월 의류 구매액	180,000
2019년 1월 취미용품구매액	740,000
2019년 1월 생활용품구매액	562,500
2019년 1월 유아용품구매액	360,000
2019년 1월 기타 구매액	715,000
2019년 1월 구매 총금액	2,805,740
2019년 1월 선호 제품군	취미용품
~	~
2021년 12월 식품류 구매액	362,150
2021년 12월 의류 구매액	0
2021년 12월 취미용품구매액	24,300
2021년 12월 생활용품구매액	478,130
2021년 12월 유아용품구매액	452,300
2021년 12월 기타 구매액	782,350
2021년 12월 구매 총금액	2,099,230
2021년 12월 선호 제품군	기타
2021년 고객 등급	S

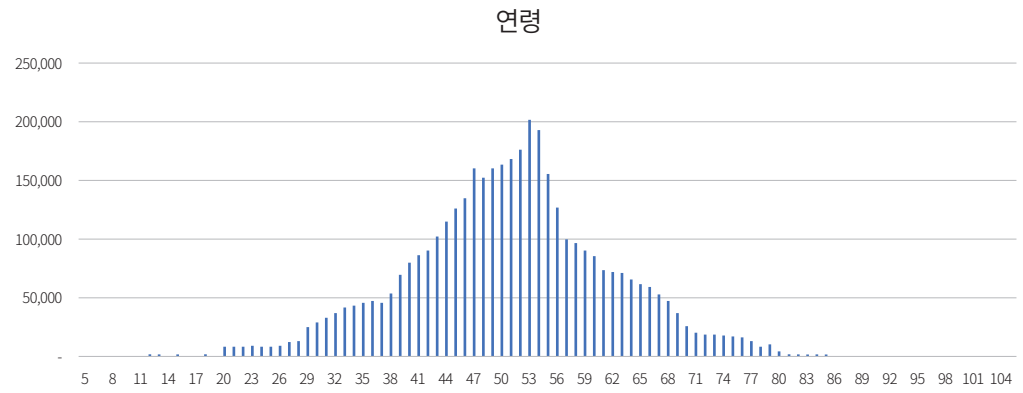
예시 개인정보 속성

컬럼명	예시
고객ID	개인식별정보(ID)
나이	개인식별가능정보(QI)
주소	개인식별가능정보(QI)
성별	개인식별가능정보(QI)
2019년 1월 여행용품구매액	개인식별가능정보(NSA)
2019년 1월 식품류 구매액	개인식별가능정보(NSA)
2019년 1월 의류 구매액	개인식별가능정보(NSA)
2019년 1월 취미용품구매액	개인식별가능정보(NSA)
2019년 1월 생활용품구매액	개인식별가능정보(NSA)
2019년 1월 유아용품구매액	개인식별가능정보(NSA)
2019년 1월 기타 구매액	개인식별가능정보(NSA)
2019년 1월 구매 총금액	개인식별가능정보(NSA)
2019년 1월 선호 제품군	개인식별가능정보(NSA)
~	~
2021년 12월 식품류 구매액	개인식별가능정보(NSA)
2021년 12월 의류 구매액	개인식별가능정보(NSA)
2021년 12월 취미용품구매액	개인식별가능정보(NSA)
2021년 12월 생활용품구매액	개인식별가능정보(NSA)
2021년 12월 유아용품구매액	개인식별가능정보(NSA)
2021년 12월 기타 구매액	개인식별가능정보(NSA)
2021년 12월 구매 총금액	개인식별가능정보(NSA)
2021년 12월 선호 제품군	개인식별가능정보(NSA)
2021년 고객 등급	개인식별가능정보(NSA)

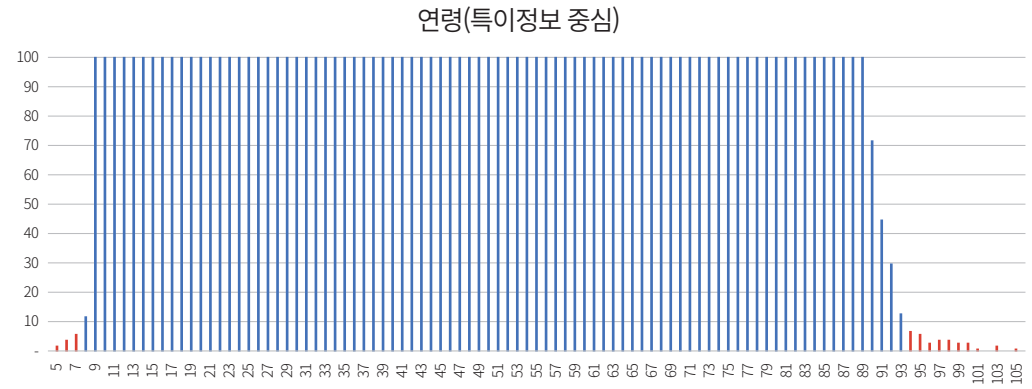
예시 원본데이터 분포

※식별 가능성이 있는 컬럼 위주로 분포표를 작성

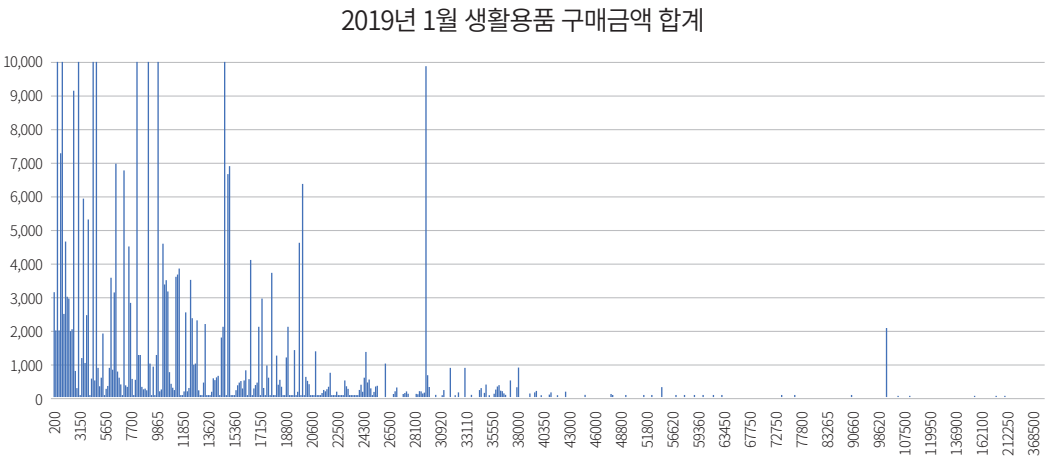
- 연령의 일반 분포



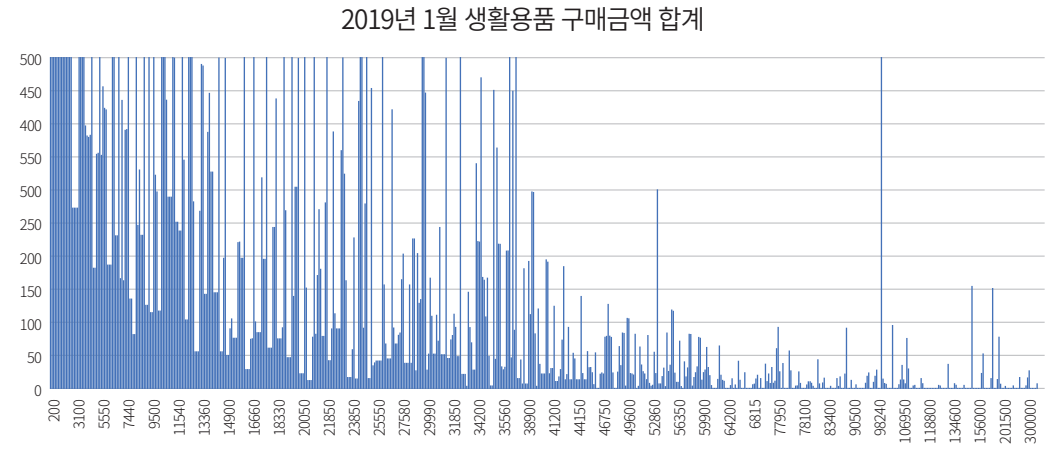
- 연령의 특이정보에 대한 분포



- 2019년 1월 생활용품 구매액의 분포



- 2019년 1월 생활용품 구매액의 특이정보에 대한 분포



예시 평가 대상 데이터 세부 항목별 명세(집계 데이터)

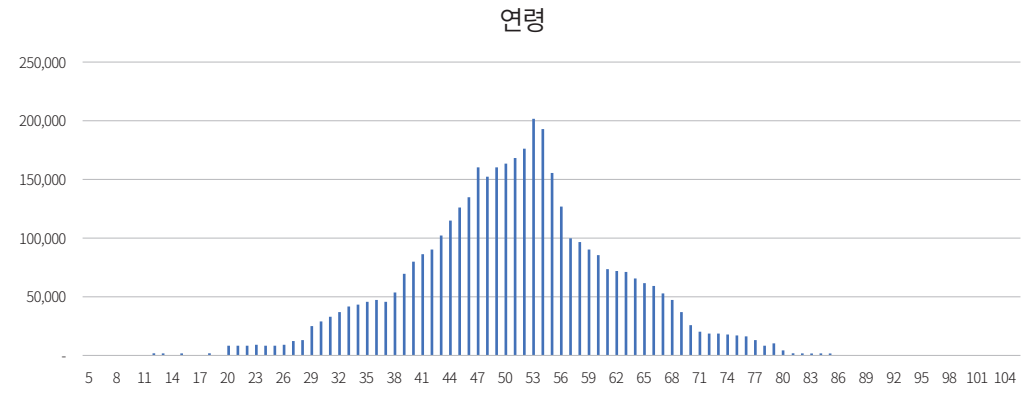
컬럼명	예시
고객ID	1~2,000,000까지의 일련번호
나이	10살단위의 나이 20살 미만 제거, 80세 이상 80세로 상단코딩
주소	시군구 단위의 주소로 부분 삭제
성별	M/F
2019년 1월 여행용품구매액	2019년 1월 여행용품 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2019년 1월 식품류 구매액	2019년 1월 식품류 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2019년 1월 의류 구매액	2019년 1월 의류 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2019년 1월 취미용품구매액	2019년 1월 취미용품 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2019년 1월 생활용품구매액	2019년 1월 생활용품 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2019년 1월 유아용품구매액	2019년 1월 유아용품 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2019년 1월 기타 구매액	2019년 1월 위의 범주에 포함되지 않는 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2019년 1월 구매 총금액	위의 7개 값의 합계
2019년 1월 선호 제품군	위의 7가지 범주 중 가장 높은 구매 금액의 제품군
~	~
2021년 12월 식품류 구매액	2021년 12월 식품류 구매금액 합계에서 1만단위로 라운딩 적용된 금액
2021년 12월 의류 구매액	2021년 12월 의류 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2021년 12월 취미용품구매액	2021년 12월 취미용품 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2021년 12월 생활용품구매액	2021년 12월 생활용품 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2021년 12월 유아용품구매액	2021년 12월 유아용품 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2021년 12월 기타 구매액	2021년 12월 위의 범주에 포함되지 않는 구매 금액 합계에서 1만단위로 라운딩 적용된 금액
2021년 12월 구매 총금액	위의 7개 값의 합계
2021년 12월 선호 제품군	위의 7가지 범주 중 가장 높은 구매 금액의 제품군
2021년 고객 등급	범주화된 고객 등급 VIP, S, A → 1 B, C → 2 D, E, F → 3

예시 평가대상 데이터

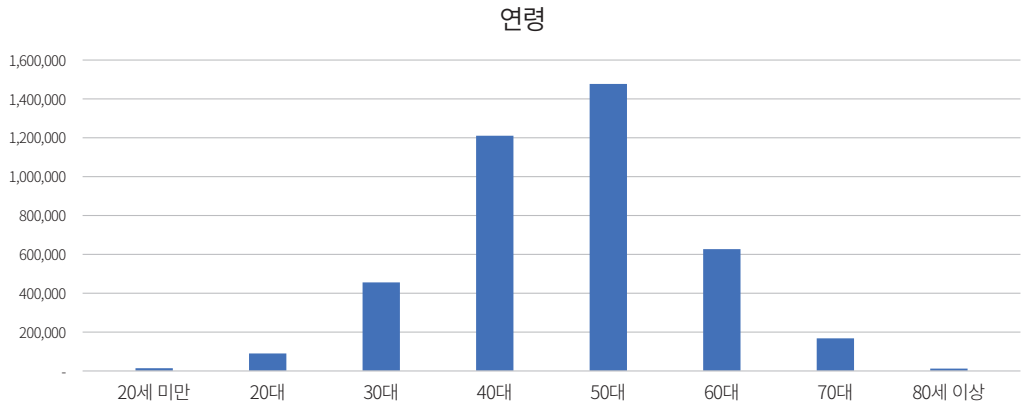
컬럼명	예시
고객ID	2428
나이	30
주소	서울시 은평구
성별	M
2019년 1월 여행용품구매액	0
2019년 1월 식품류 구매액	250,000
2019년 1월 의류 구매액	180,000
2019년 1월 취미용품구매액	740,000
2019년 1월 생활용품구매액	560,000
2019년 1월 유아용품구매액	360,000
2019년 1월 기타 구매액	710,000
2019년 1월 구매 총금액	2,800,000
2019년 1월 선호 제품군	취미용품
~	~
2021년 12월 식품류 구매액	360,000
2021년 12월 의류 구매액	0
2021년 12월 취미용품구매액	30,000
2021년 12월 생활용품구매액	480,000
2021년 12월 유아용품구매액	450,000
2021년 12월 기타 구매액	780,000
2021년 12월 구매 총금액	2,100,000
2021년 12월 선호 제품군	기타
2021년 고객 등급	1

예시 평가대상 데이터 분포

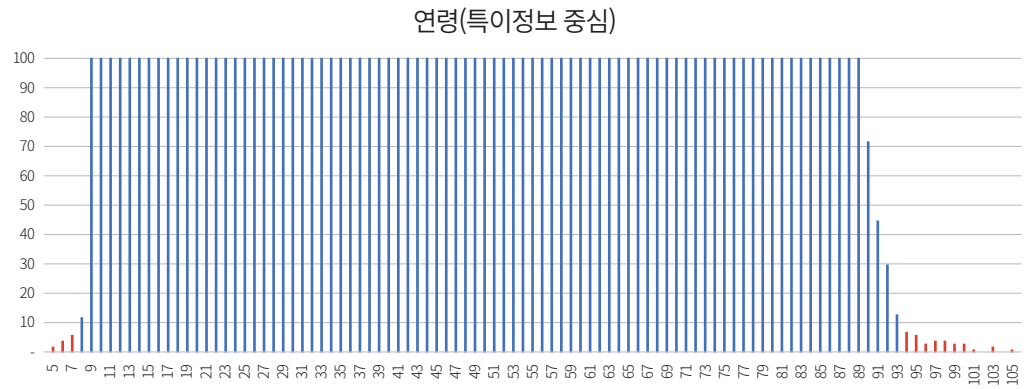
- 연령의 일반 분포(원본)



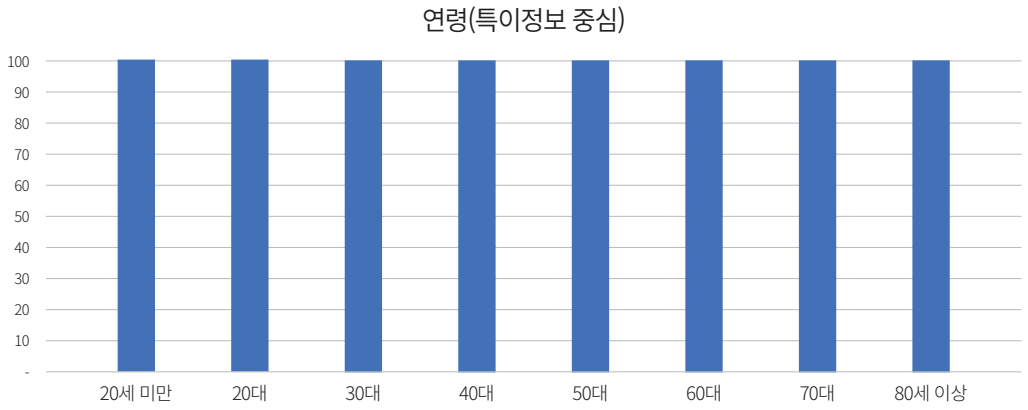
- 연령의 일반 분포(가명처리 후)



- 연령의 특이정보에 대한 분포(원본)



- 연령의 특이정보에 대한 분포(가명처리 후)



부록 3

비정형데이터 가명처리 시나리오 예시

※ 아래 가명처리 시나리오는 실제 비정형데이터를 가명처리하여 활용했던 사례를 관련 기업·기관 및 전문가 논의를 통해 재구성한 것으로 단순 참고용이며, 처리자 및 적정성 검토위원회 등의 판단에 따라 데이터 활용 분야·상황에 맞게 가명처리 방법·수준 등을 자유롭게 적용할 수 있음

시나리오 ❶ : 유방암·골밀도 감소 여부 진단 AI 개발 사례

의료 분야 (이미지, 영상, 텍스트)

한국대학교병원은 병원 내부 연구자에게 과학적 연구 수행을 위한 데이터 제공을 요청받아 유방암 환자의 이미지·영상 데이터(병리조직, 흉부CT, 골밀도검사(DEXA) 기록 등)과 임상 데이터(외과병리 검사결과(텍스트), 정형데이터)를 제공하려 한다.

☑ 데이터의 이용 목적

▶ 유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구

☑ 데이터 특징

▶ (이미지·영상 데이터) 암 진단 및 치료 관련 병리조직이미지, 흉부CT, 골밀도 검사(DEXA) 기록

▶ (임상 데이터) 암 진단 시 발생한 암 관련 외과병리 검사결과(텍스트), 암등록정보 등 정형데이터

☑ 데이터의 이용 환경

▶ (폐쇄연구분석환경 활용) 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실에서 데이터 활용, 승인된 사용자 외에는 접근 불가

▶ (자료 반입) 자료 반입시 한국대학교병원 관리자에게 요청(관리자가 자료 확인 후 반입)

▶ (자료 반출) 분석결과 반출 시 한국대학교병원 관리자에게 요청(관리자 자료 확인 후 제공)

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	SG-20230110001	접수일	2023년 1월 10일

신청자	조직/부서명	A대학교병원 외과		
	담당자 직위	교수	담당자 성명	한의료
	전화번호	03-1234-5678	이메일 주소	research_han@corea.ac.kr

처리 목적		☐ 통계작성 ☑ 과학적 연구 ☐ 공익적 기록보존
		유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구
활용 형태		☑ 내부이용 ☐ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체		☑ 동일 개인정보처리자 ☐ 제3자(제공받는 자) 제공
처리 환경		☑ 내부 ☐ 외부
처리 장소		☑ 폐쇄 환경 ☐ 제한 없음
반복 제공 여부		☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☑ 온라인* ☐ 오프라인 * 폐쇄형연구분석환경에서만 접속 가능한 클라우드 서버에 업로드
제공 받는 자		(내부) ☐ 동일부서 ☑ 타부서 (외부) ☐ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	임상데이터, 병리 및 CT, DEXA 영상 이미지
	데이터 내역	2012년부터 2019년까지 유방암 병리 및 CT, DEXA 검사를 받은 유방암 환자 (500명)의 임상데이터, 병리 및 CT, DEXA 영상 이미지
	가명정보 이용 기간	2023년 5월 1일 ~ 2025년 4월 30일(2년)

위와 같이 가명정보 이용·제공을 신청합니다.

2023년 3월 2일

신청인(부서장명) 한 의 료 (서명 또는 인)

첨부서류	1. 연구계획서 2. 개인정보 유형 분류표 3. 가명정보에 대한 안전조치 이행 확약서 4. 보안 서약서 등출

② 보호법에서 정한 목적 중 가명정보 처리 목적을 명확히 설정하였는지 검토

과학적 연구 계획서

연구명	유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구
연구진	한국대학교병원 외과 한의료
연구 배경 및 목적	▪ 유방암은 2020년 기준 우리나라 여성에서 가장 많이 발생하는 암으로 정확한 진단을 위해 조직병리검사를 수행하고 있으며, 암 수술 후 전이, 재발 등을 검진하기 위해 정기적으로 CT 등을, 치료 중 골밀도 감소 여부 확인을 위해 DEXA 검사를 수행하고 있음 ▪ 유방암 병리조직이미지를 이용한 AI 분석을 통해 유방암 조직학적진단의 민감도를 높이고자 함 ▪ 또한 유방암 치료 중 골밀도 감소증 발생이 흔하여 주기적인 추적 검사가 필요한 바, CT 및 DEXA 이미지 학습을 통해 CT 검사를 통한 골밀도 감소 사전 예측을 수행하고자 함
예상 연구 기간	2023. 5. 1. ~ 2025. 4. 30.(2년)
연구 대상자 수	한국대학병원에서 유방암으로 진단받고 수술한 여성 환자 500명 - 연구대상자 선정기간 (2012.1.1. ~ 2019.12.31.)
연구 방법	유방암 병리 및 CT, DEXA 검사 등 임상데이터에서 유방암 발생 여부, 골밀도 감소 여부 등을 판별 및 분류하는 AI 모듈 개발을 위한 학습에 활용
연구내용	▪ 병리 이미지와 조직 병리 검사결과를 이용한 AI 학습을 통해 병리 이미지를 이용한 진단 결과 도출 - 450명은 학습용 데이터셋으로, 50명은 테스트 데이터셋으로 활용 ▪ 흉부CT 이미지와 DEXA 검사결과를 이용한 AI 학습을 통해 CT에서 골밀도 감소 여부 결과 도출 - 450명은 학습용 데이터셋으로, 50명은 테스트 데이터셋으로 활용
기대효과 및 활용방안	▪ 조직 병리 진단 자동화를 통해 유방암 진단의 정확도를 높이고 예후와의 관련성을 확인 ▪ 유방암 환자의 재발 등을 확인하기 위해 검사한 CT 이미지를 이용하여 골밀도 감소 진단을 수행함으로써 유방암 환자들에게 필요한 DEXA 검사 등의 부담을 감소시킬 수 있음

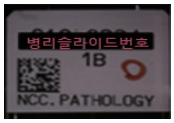
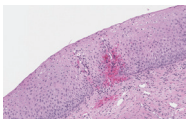
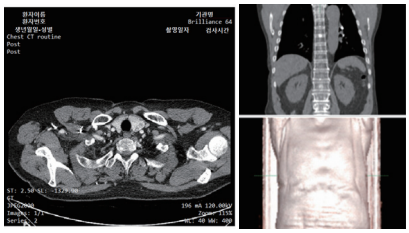
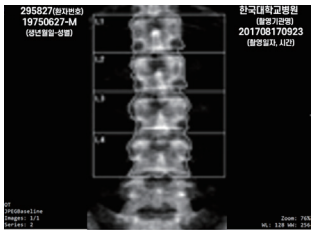
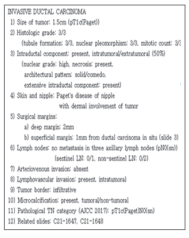
③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

개인정보 유형 분류표 (요약)

연번	항목명	데이터 유형	데이터 규모	비고
1	병리조직 데이터	이미지	2,000장 (500명*4장)	비정형데이터
2	흉부CT 데이터	영상·이미지 (DICOM*)	100,000장 (500명*100장*2회 촬영)	
3	골밀도 검사 (DEXA) 데이터	영상·이미지 (DICOM*)	1,000장 (500명*2장*1회 촬영)	
4	외과병리 검사결과	텍스트(비정형, 관찰입력정보)	500건	
4-1	Organ(장기명)	텍스트(정형)	500건	비정형데이터 ⇒ 정형데이터 변환
4-2	Location(유방 수술의 위치)	텍스트(정형)	500건	
~	~	~	500건	
4-17	Intraductal_comp (상피내암 유무)	텍스트(정형)	500건	
5	gender(성별)	텍스트(정형)	500건	정형데이터 (환자 임상데이터)
6	death(사망일자)	숫자(정형)	500건	
~	~	~	500건	
32	MIEX_YMD(영상검사 일자)	숫자(정형)	500건	

* DICOM(Digital Imaging and Communications in Medicine) : 의료용 디지털 영상 및 통신 표준

개인정보 유형 분류표 (상세 : 비정형데이터)

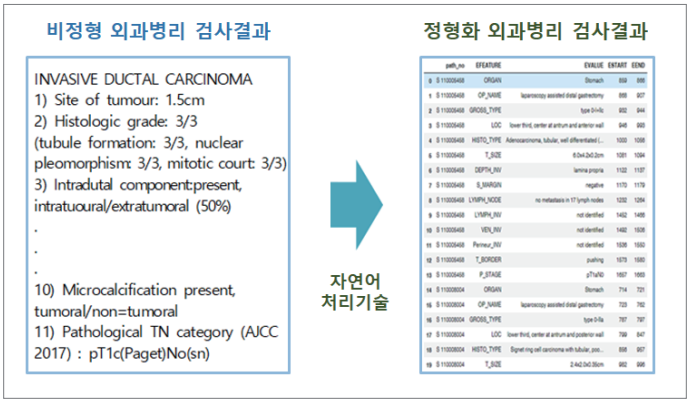
연번	항목명	데이터 유형	데이터 규모	구분	예시
1	병리조직 데이터	이미지	2,000장 (500명* 4장)	① 비율슬라이드 사진	
				② 병리슬라이드 번호 사진	
				③ 검체 슬라이드 사진	
				④ 검체 확대 사진	
2	흉부CT 데이터	영상·이미지 (DICOM)	100,000장 (500명 * 100장 * 2회 촬영)		 * 이미지 내 환자이름, 생년월일, 성별, 환자번호 존재
3	골밀도 검사 (DEXA) 데이터	영상·이미지 (DICOM)	1000장 (500명 * 2장 * 1회 촬영)		 * 이미지 내 환자 생년월일, 성별, 환자번호, 촬영기관명, 촬영일자·시간 존재
4	외과병리 검사결과	텍스트 (관찰입력정보, 자유입력텍스트)	500명 (11.5MB)		

⇒4. 외과병리 검사결과(비정형 텍스트데이터)는 자연어 처리기술 및 수기작업을 통해 정형데이터로 변경하여 활용하고자 함

예시 비정형 외과병리 검사결과의 정형데이터 변환

외과병리 검사결과는 암의 치료, 예후 예측을 위해 매우 중요한 정보를 포함하나, 반정형의 검사결과 형태로 되어 있어 정형자료와 비정형자료의 애매한 경계선에 있는 자료로 그대로는 활용이 쉽지 않고, 병리보고서 내에 연구에 필요하지 않은 다양한 개인식별가능정보들이 정제되지 않은 형태로 활용될 위험이 있으므로, 비정형 검사결과 활용을 위해서는 정형화된 형태로의 작업을 통해 활용이 가능하도록 처리가 필요할 수 있음
기존에는 병리과 수작업 등의 방법으로 추출 및 활용하였으나, 최근에는 자연어처리기술의 발달로 학습에 의한 정형화된 정보로 추출 및 관리가 가능해짐

비정형 외과병리 검사결과의 정형화 데이터 변환 예시



개인정보 유형 분류표 (상세: 비정형데이터 ⇒ 정형데이터 변환)

연번	항목명(영문)	항목명(한글)	예시	비고
〈외과병리 정보〉				
4-1	Organ	장기명	Breast	
4-2	Location	유방 수술의 위치	Right, left	
4-3	Tumor_size	종양의 크기	2.2cm	
4-4	Histologic_type	조직학적 유형	Invasive ductal carcinoma	
4-5	Histologic_grade	조직학적 등급	2/3	
4-6	Surgical_margins	수술절제면	deep margin	
~	~	~	~	~
4-17	Intraductal_comp	상피내암 유무	present	

개인정보 유형 분류표 (상세: 정형데이터)

연번	항목명(영문)	항목명(한글)	예시	비고
〈기본정보〉				
5	gender	성별	F	
6	death	사망일자	19910525	
〈암 등록정보〉				
7	fdx	초진연월일	20200505	
8	age	암진단당시 나이	50	
~	~	~	~	
〈수술 정보〉				
20	sur_Date	수술일자	20200605	
21	sur_Date1	처방일자	20200605	
~	~	~	~	
〈외과병리 정보〉				
25	hist_date1	병리검사일자	20200605	
26	hist_date2	처방일자	20200605	
〈DEXA 검사 결과〉				
27	MIEX_YMD	검사일자	2017-04-21	
28	MIEX_CD	검사코드	RX211	
29	MIEX_NM	검사명	Dexa Bone Densitometry	
30	RGN_NM	검사부위명	L1-L4	
〈CT영상 검사 결과〉				
31	ORD_YMD	영상검사 처방일자	2020-08-24	
32	MIEX_YMD	영상검사 일자	2020-09-27	

활용데이터 요구 수준표

※ 가명정보 처리 목적 달성에 필요한 데이터 항목과 가명처리 요구수준을 검토

비정형데이터

연번	항목명	요구 수준	비고
1	병리조직데이터 (이미지)	1) 비율슬라이드 사진 : 분석에 필요 없어 삭제·대체 가능 2) 병리슬라이드 번호 사진 : 분석에 필요 없어 삭제·대체 가능 3) 검체 슬라이드 사진 : 연구목적 달성을 위해 그대로 사용 필요 4) 검체 확대 사진 : 연구목적 달성을 위해 그대로 사용 필요	
2	흉부CT 데이터 (DICOM)	1) 흉부 촬영부분: 연구목적 달성을 위해 그대로 사용 필요 2) 이미지 내 환자관련정보 2-1) 환자이름 : 분석에 필요 없으므로 삭제·대체 가능 2-2) 생년월일 : 분석에 필요 없어 삭제·대체 가능 2-3) 환자성별 : 여성환자로 한정하여 분석하므로, 남성환자 정보는 모두 삭제 삭제·대체해도 연구에 영향 없음 2-4) 환자번호 : 환자구분만 되면 되므로 일련번호로 대체 가능	
3	골밀도 검사 데이터 (DICOM)	1) 골밀도 촬영부분: 연구목적 달성을 위해 그대로 사용 필요 2) 이미지 내 환자관련정보 2-1) 생년월일 : 분석에 필요 없으므로 삭제·대체 가능 2-2) 환자성별 : 분석에 필요 없으므로 삭제·대체 가능 2-3) 환자번호 : 분석에 필요 없으므로 삭제·대체 가능 2-4) 촬영기관명 : 분석에 필요 없으므로 삭제·대체 가능 2-5) 촬영일자·시간 : 분석에 필요 없으므로 삭제·대체 가능	
4	외과병리 검사결과	연구수행을 위해 정형화된 검사결과만 필요하므로, 정형데이터로 변환 후 삭제 가능	

비정형데이터 → 정형데이터

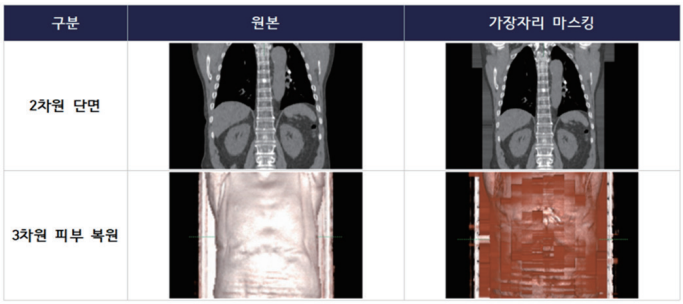
연번	항목명(영문)	항목명(한글)	요구 수준	비고
〈외과병리 정보〉				
4-1	Organ	장기명	연구목적 달성을 위해 그대로 사용 필요	
4-2	Location	유방 수술의 위치	연구목적 달성을 위해 그대로 사용 필요	
4-3	Tumor_size	종양의 크기	연구목적 달성을 위해 그대로 사용 필요	
4-4	Histologic_type	조직학적 유형	연구목적 달성을 위해 그대로 사용 필요	
4-5	Histologic_grade	조직학적 등급	연구목적 달성을 위해 그대로 사용 필요	
4-6	Surgical_margins	수술절제면	연구목적 달성을 위해 그대로 사용 필요	
~	~	~	~	~
4-17	Intraductal_comp	상피내암 유무	연구목적 달성을 위해 그대로 사용 필요	

정형데이터

연번	항목명(영문)	항목명(한글)	요구 수준	비고
〈기본정보〉				
5	gender	성별	여성환자 대상 분석으로, 여성환자 정보만 필요 남성환자 정보는 삭제	
6	death	사망일자	연별 분석 수행 예정으로, 연도 자료만 필요	
〈암 등록정보〉				
7	fdx	초진연월일	월단위 분석 수행 예정으로, 연월만 필요	
8	age	암진단당시 나이	통계 분석에 무리가 없는 선에서 일정 수준의 범주화 가능(협의 필요)	
~	~	~	~	
〈수술 정보〉				
20	sur_Date	수술일자	월단위 분석 수행 예정으로, 연월만 필요	
21	sur_Date1	처방일자	월단위 분석 수행 예정으로, 연월만 필요	
~	~	~	~	
〈외과병리 정보〉				
25	hist_date1	병리검사일자	월단위 분석 수행 예정으로, 연월만 필요	
26	hist_date2	처방일자	월단위 분석 수행 예정으로, 연월만 필요	
〈DEXA 검사 결과〉				
27	MIEX_YMD	검사일자	월단위 분석 수행 예정으로, 연월만 필요	
28	MIEX_CD	검사코드	연구목적 달성을 위해 그대로 사용 필요	
29	MIEX_NM	검사명	연구목적 달성을 위해 그대로 사용 필요	
30	RGN_NM	검사부위명	연구목적 달성을 위해 그대로 사용 필요	
〈CT영상 검사 결과〉				
31	ORD_YMD	영상검사 처방일자	월단위 분석 수행 예정으로, 연월만 필요	
32	MIEX_YMD	영상검사 일자	월단위 분석 수행 예정으로, 연월만 필요	

식별 위험성 검토 결과보고서

※ 식별 위험성 검토 점검표를 기반으로 식별 위험성 검토 결과보고서 작성

<table>
<tr><td>가명정보
활용목적</td><td colspan="2">▪ 유방암 진단 자동화 및 골밀도 감소 여부 확인을 위한 AI 개발 연구</td></tr>
<tr><td>가명처리 대상
데이터 항목</td><td colspan="2">▪ 병리조직데이터(이미지) 〈연번 1〉
▪ 흉부CT데이터(영상·이미지) 〈연번 2〉
▪ 골밀도 검사(DEXA) 데이터(영상·이미지) 〈연번 3〉
▪ 환자 임상데이터(외과병리 검사결과, 텍스트) 〈연번 4 (4-1~4-17)〉
▪ 환자 임상데이터(환자 기본정보, 암 등록정보, 수술정보, 외과병리정보, DEXA검사결과, CT영상 검사결과) 〈연번 5~32〉</td></tr>
<tr><td rowspan="3">데이터
위험성</td><td>식별성 유무</td><td>〈비정형데이터〉
▪ (병리이미지) 병리슬라이드 번호사진 외에는 개인식별 가능성 거의 없음
▪ (CT 영상·이미지) 영상·이미지 자체로는 개인식별 가능성 거의 없음
- DICOM 영상·이미지에 포함된 환자관련정보(환자이름, 생년월일, 환자성별, 환자번호)는 개인식별 가능성이 있어 가명처리 필요
▪ (DEXA 영상·이미지) 영상·이미지 자체로는 개인식별 가능성 거의 없음
- DICOM 영상·이미지에 포함된 환자관련정보(환자이름, 생년월일, 환자성별, 환자번호 등)는 개인식별 가능성이 있어 가명처리 필요
▪ (외과병리 검사결과) 외과병리 결과 데이터를 정형데이터로 변환하여 활용할 예정이며, 외과병리 검사결과 내 외과병리정보만으로는 개인식별 가능성 거의 없음
〈정형데이터〉
▪ 한국대학교병원 암관련 환자의 사망일자, 초진연월일, 암진단 연령, 병리검사일자, 처방일자, CT영상검사 처방일자, CT영상검사 일자 등은 조합되었을 때 개인의 식별이 가능한 개인식별가능정보임</td></tr>
<tr><td>특이정보 유무</td><td>〈비정형데이터〉
▪ (CT 영상·이미지) 개인당 200장의 이미지가 촬영되었기 때문에 3차원 재건 등의 기술을 활용하면 신체 이미지를 입체적으로 복원가능하며, 복원시 특이한 외형·흉터 등이 있는 환자의 경우 낮은 확률로 식별위험이 생길 수 있음
* 가장자리 마스킹 기법을 활용하여 3차원 재건 공격 위험을 막을 수 있음
〈정형데이터〉
▪ '암진단 연령'의 경우 상당히 적거나 높은 경우 특이치로 인한 개인식별성이 발생할 수 있음
 
</td></tr>
<tr><td>재식별시 영향도</td><td>▪ 개인에 대한 진료·진단 정보로, 재식별시 영향도는 높은 편</td></tr>
</table>

<table>
<tr><td rowspan="3">처리 환경
위험성</td><td>이용 및 제공 형태</td><td>▪ 기관내 연구자에게 제공</td></tr>
<tr><td>처리 장소</td><td>▪ 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실
▪ 한국대학교병원은 개인정보(가명정보)처리시스템에 대한
ISMS-P인증 취득</td></tr>
<tr><td>다른 정보
결합 가능성</td><td>▪ 폐쇄환경분석실 관리자 승인하에 제한된 데이터·프로그램(프로그램 패키지,
라이브러리, 코드설명서 등)만 반입 가능
▪ 분석대상 가명정보와 결합가능성 있는 데이터는 반입 제한</td></tr>
<tr><td>최종 검토의견</td><td colspan="2">▪ 해당 가명정보는 기관내 연구자에게 제공될 뿐만 아니라, 클라우드 기반의 폐쇄연구 분석환경(외부 인터넷 이용, 다른 데이터의 반입 및 결합, 데이터 외부반출 등이 제한)에서만 접속하여 분석 가능하므로 식별 가능성이 낮은 편이며 안전한 처리 환경을 고려할 때 다음과 같은 조치가 필요함
- 병리슬라이드 번호 사진, 개인 신상정보 중 '성별' 등 연구 수행에 필요없는 정보는 삭제
- 병리조직·CT·DEXA 영상·이미지는 연구목적에 필요한 범위 내에서 그대로 활용하여도 식별 가능성이 거의 없으나 영상·이미지에 포함된 환자관련정보(환자이름, 생년월일, 환자성별, 환자번호 등)는 삭제·대체 등 필요
- 개인당 200장의 이미지가 촬영된 CT의 경우 3차원 재건 등을 통한 식별위험이 존재하나, 연구자가 기관 내 폐쇄연구 분석환경에서 연구를 수행하고 외부 데이터·프로그램 활용이 제한되기 때문에 관련 식별위험이 없을 것으로 판단되므로, 가장자리 마스킹 등 별도 가명처리 없이 그대로 활용 가능
- '사망일자', '초진연월일', 연령', '병리검사일자', '처방일자', 'CT영상검사 처방일자', 'CT영상검사 일자', 'DEXA영상검사 일자'는 그대로 사용하는 경우 조합에 의한 개인식별 가능성이 있으며, 연구목적상 연·월단위 분석까지만 수행하므로 필요에 맞게 사망일자는 연단위까지, 나머지 데이터는 월단위까지만 활용 필요
- '암진단 연령' 또한 '39세 이하', '40~84세: 1세단위', '85세 이상'으로 범주화하여 연구 필요
- 자유입력데이터인 외과병리 검사결과는 그대로 활용이 쉽지 않으므로 자연어 처리 기술 등을 통해 정형데이터로 변환 후 활용하는 것이 적절</td></tr>
</table>

④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

항목별 가명처리 계획

비정형데이터

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	병리조직 데이터	1) 비율슬라이드 사진		분석에 필요 없으므로 삭제
		2) 병리슬라이드 번호 사진	☑ 삭제	분석에 필요 없고, 식별 가능성 있으므로 삭제
		3) 검체 슬라이드 사진	☑ 그대로 사용	별도 처리하지 않음
		4) 검체 확대 사진	☑ 그대로 사용	별도 처리하지 않음
2	흉부CT 데이터	1) 흉부 촬영부분	개인당 200장의 사진이 촬영되어, 3차원 재건 등의 기술을 활용하면 신체 이미지를 복구할 수 있고 복원시 특이한 외형·흉터 등이 있는 환자의 경우 낮은 확률로 식별가능성 존재	
			☐ 표면 가장자리 삭제	3차원 재건으로 인한 개인식별 가능성을 더욱 낮출 수는 있으나 이 경우, 데이터 손실로 인해 연구목적 달성이 어려움
			☑ 그대로 사용	이미지를 그대로 사용하되, 3차원 재건 기술을 적용할 수 없도록 처리 환경을 통제 * 클라우드 기반 폐쇄연구 분석환경에서만 활용, 철저한 외부 데이터·프로그램 반입 관리 수행
		2) 이미지 내 환자관련정보 * DICOM 형식 그대로 연구에 활용이 필요하기 때문에 블랙마스킹 처리가 불가하며, 이미지 내 표시된 DICOM 메타데이터를 손쉽게 일괄 변경할 수 있도록 자체개발한 도구를 활용하여 가명처리 수행		
		2-1) 환자이름	☑ 대체	연구에 필요 없는 정보이므로, 자체개발한 DICOM 데이터 변경도구를 활용하여 Anonymized 값으로 대체 (블랙마스킹으로 삭제해도 무방)
		2-2) 생년월일	☑ 대체	연구에 필요 없으므로 자체개발한 DICOM 데이터 변경도구를 활용하여 임의값인 “1900-00-00”으로 대체(블랙마스킹으로 삭제해도 무방)
		2-3) 환자성별	☑ 삭제·대체	여성환자만 분석에 사용하므로 남성 환자가 있다면 데이터를 삭제하고 이후 성별값은 구분에 의미가 없으므로, 자체개발한 DICOM 데이터 변경도구를 활용하여, “F”(Female) 값을 “S”(Sex)로 대체 (블랙마스킹으로 삭제해도 무방)
		2-2) 환자번호	☑ 대체	환자구분만 되면 되므로 자체개발한 DICOM 데이터 변경도구를 활용하여, 단순일련번호로 대체

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
3	골밀도검사 (DEXA) 데이터	1) 골밀도 촬영부분	개인당 2장의 사진만 촬영되었을 뿐만 아니라, 골밀도 촬영 사진의 특성상 3차원 재건 등의 기술을 활용한 신체 이미지 복구는 어려움	
			☑ 그대로 사용	3차원 재건 등을 통한 식별 위험성이 없으므로, 별도 처리하지 않음
		2) 이미지 내 환자관련정보 * PNG 포맷으로 연구에 활용할 계획으로, 파일 포맷 변환(DICOM→PNG) 후, 블랙마스킹 기법을 통해 이미지 내 환자관련정보 삭제		
		2-1) 생년월일	☑ 마스킹	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
		2-2) 환자성별	☑ 마스킹	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
		2-3) 환자번호	☑ 마스킹	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
		2-4) 촬영기관명	☑ 마스킹	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
		2-5) 촬영일자·시간	☑ 마스킹	분석에 필요 없으므로, 블랙마스킹 기법으로 삭제 처리
4	외과병리 검사결과	검사결과 내용	☑ 삭제	정형데이터로 변환시켜 분석에 필요한 주요 검사결과만 추출 (아래 4-1~17 참조)한 뒤 삭제

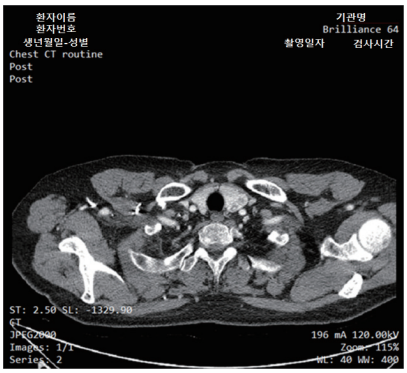
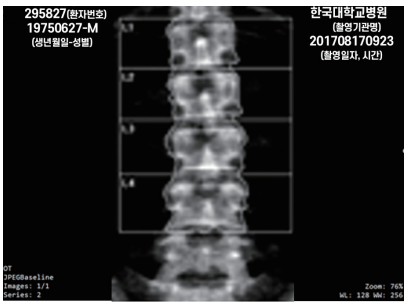
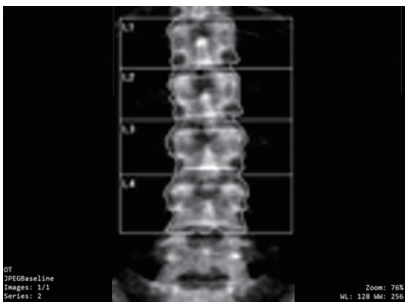
비정형데이터 → 정형데이터

연번	항목명(영문)	항목명(한글)	처리방법	세부방법 및 처리수준
〈외과병리 검사결과〉				
4-1	Organ	장기명	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-2	Location	유방 수술의 위치	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-3	Tumor_size	종양의 크기	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-4	Histologic_type	조직학적 유형	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-5	Histologic_grade	조직학적 등급	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
4-6	Surgical_margins	수술절제면	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음
~	~	~	~	~
4-17	Intraductal_comp	상피내암 유무	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음

정형데이터

연번	항목명(영문)	항목명(한글)	처리방법	세부방법 및 처리수준	가명처리 전	가명처리 후
〈개인 신상정보〉						
5	gender	성별	삭제	분석에 필요 없으므로 삭제 처리		
6	death	사망일자	부분삭제	연도만 필요하므로 월일은 삭제 처리	20230404	2023
〈암 등록정보〉						
7	fdx	초진연월일	부분삭제	연월만 필요하므로 일자는 삭제 처리		
8	age	암진단당시 나이	범주화	39세 이하, 40-84세(1세단위), 85세 이상으로 구분	29세 50세 93세	39세 이하 50세 85세 이상
~	~	~	~	~		
〈수술 정보〉						
20	sur_Date	수술일자	부분삭제	연월만 필요하므로 일자는 삭제 처리		
21	sur_Date1	처방일자	부분삭제	연월만 필요하므로 일자는 삭제 처리		
~	~	~	~	~		
〈외과병리 정보〉						
25	hist_date1	병리검사일자	부분삭제	연월만 필요하므로 일자는 삭제 처리		
26	hist_date2	처방일자	부분삭제	연월만 필요하므로 일자는 삭제 처리		
〈DEXA 검사 결과〉						
27	MIEX_YMD	검사일자	부분삭제	연월만 필요하므로 일자는 삭제 처리		
28	MIEX_CD	검사코드	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음		
29	MIEX_NM	검사명	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음		
30	RGN_NM	검사부위명	그대로 사용	개인 식별가능성이 낮아 별도 처리하지 않음		
〈CT영상 검사 결과〉						
31	ORD_YMD	영상검사 처방일자	부분삭제	연월만 필요하므로 일자는 삭제 처리	20200610	202006
32	MIEX_YMD	영상검사 일자	부분삭제	연월만 필요하므로 일자는 삭제 처리	20200927	202009

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

연번	항목명	가명처리 전	가명처리 후
2	흉부CT 이미지	 1) 흉부촬영부분 2) DICOM 이미지 내 환자관련정보 2-1) 환자이름 2-2) 생년월일 2-3) 환자성별 2-4) 환자번호	 1) 흉부촬영부분 : 그대로 유지 2) DICOM 이미지 내 환자관련정보 2-1) 환자이름 : Anonymized로 대체 2-2) 생년월일 : 정해진 임의 값으로 대체 2-3) 환자성별 : 성별값을 “S”로 대체 2-4) 환자번호 : 환자 구별을 위한 단순 일련번호로 대체 * 자체개발한 DICOM 데이터 변경도구 활용
3	골밀도 검사 DEXA 이미지	 1) 골밀도 촬영부분 2) 이미지 내 환자관련정보 2-1) 생년월일 2-2) 환자성별 2-3) 환자번호 2-4) 촬영기관명 2-5) 촬영일자·시간	 1) 골밀도 촬영부분 : 그대로 유지 2) 이미지 내 환자관련정보 2-1) 생년월일 : 블랙마스킹 2-2) 환자성별 : 블랙마스킹 2-3) 환자번호 : 블랙마스킹 2-4) 촬영기관명 : 블랙마스킹 2-5) 촬영일자·시간 : 블랙마스킹 * PNG 포맷으로 연구에 활용할 계획으로, 파일 포맷 변환(DICOM→PNG) 후, 이미지 블랙마스킹 기법을 통해 이미지 내 환자관련정보 삭제

가명처리 결과 자체검증

비정형데이터 가명처리 기술의 적절성·신뢰성 관련 근거(예시)

<table>
<tr><th>연번</th><th>대상 항목</th><th>데이터 유형</th><th>처리 기술명</th><th>처리 기술의 적절성·신뢰성 관련 근거 또는 배경</th><th>비고</th></tr>
<tr><td>2</td><td>흉부CT 데이터</td><td>이미지</td><td>- DICOM 데이터 변경도구를 사용한 환자정보 변경·대체</td><td>▪ 한국대학교병원에서 자체 개발한 DICOM 데이터 변경도구를 활용하여 이미지 내 환자관련정보(환자이름, 생년월일, 환자성별, 환자번호) 대체
- 자체 개발한 DICOM 데이터 변경도구는 CT사진에 대해 DICOM 형식 그대로 연구에 활용이 필요할 경우(블랙마스킹 처리불가) 등을 위해 이미지 내에 표시된 DICOM 메타데이터를 손쉽게 일괄 변경할 수 있도록 개발한 솔루션
- 개발용역사의 테스트 결과, 객체 인식률 99.2%, 처리 정확도 100%로 측정
※ 객체 인식률, 처리 정확도(오류율) 증빙자료 별첨
- 환자관련정보가 특이한 경우(환자이름이 너무 길어 도구 인식가능 영역을 넘어가는 경우 등) 인식이 안되는 경우가 존재하므로, 내부정책상 DICOM 데이터 변경도구를 이용한 가명처리 후 추가적인 자체 전수검사 수행</td><td></td></tr>
<tr><td>3</td><td>골밀도 검사 (DEXA) 데이터</td><td>이미지</td><td>- 영상이미지 블랙 마스킹</td><td>▪ 파일 포맷 변환(DICOM→PNG) 후, 이미지 블랙마스킹 프로그램을 통해 이미지 내 환자관련정보 삭제
- 파일 포맷 변환(DICOM→PNG) 후에도 이미지 내 환자관련정보가 동일한 위치에 존재하므로 환자관련정보가 위치한 영역을 지정하여 블랙마스킹 프로그램 적용
- 프로그램 개발사 테스트 결과, 객체 인식률 98%, 처리 정확도 98.4%로 측정
※ 객체 인식률, 처리 정확도(오류율) 증빙자료 별첨
- 환자관련정보가 특이한 경우(환자이름이 너무 길어 도구 지정 영역을 넘어가는 경우 등), 인식·처리가 정확히 안되는 경우가 존재하므로, 내부정책상 블랙마스킹 프로그램을 이용한 가명처리 후 추가적인 자체 전수검사 수행</td><td></td></tr>
<tr><td>4</td><td>외과병리 검사결과</td><td>텍스트</td><td>- 자연어 처리 기술(NLP)을 통해 정형데이터로 변환
* NLP(Natural Language Processing)는 인간의 언어를 해석, 조작 및 이해하는 능력을 컴퓨터에 부여하는 기계학습 기술로서, 과거 규칙기반 처리에서 최근 머신러닝 알고리즘 기반으로 발전 중</td><td>▪ 한국대병원은 EMR 내 암종별 외과병리 검사결과 현황을 파악하여 암종별 특성에 따라 자연어 처리가 가능한 학습모델을 자체개발
▪ 사전에 검토한 자연어처리 모델 중 외과병리 검사결과 처리에 가장 적합하다고 판단한 BioBERT 모델을 선정하고, 의료정보관리사가 암종별로 주석처리한 외과병리 검사결과를 바탕으로 학습모델을 개발하여 검증
▪ 한국대병원의 외과병리 검사결과 중 4개 암종(간암, 대장암, 위암, 유방암)을 대상으로 검증을 진행하였으며, 유방암은 95.8%의 처리 정확도를 보임
※ 관련 증빙자료 별첨
<table>
<tr><th>암종명</th><th>건수</th><th>변수 개수</th><th>암종별 평균 처리 정확도</th></tr>
<tr><td>간암</td><td>2,312</td><td>22</td><td>97.8%</td></tr>
<tr><td>대장암</td><td>7,869</td><td>24</td><td>94%</td></tr>
<tr><td>위암</td><td>5,772</td><td>15</td><td>95.3%</td></tr>
<tr><td>유방암</td><td>18,492</td><td>24</td><td>95.8%</td></tr>
</table>
▪ 검증된 모델을 통해 실제 암 외과병리 검사결과의 텍스트 정보를 추출하여 정형데이터로 변환해 활용
- 다만, 처리 정확도가 100%가 아니므로 변환 이후 추가적인 자체 전수검사를 수행하여 보완작업 진행</td><td></td></tr>
</table>

비정형데이터 가명처리 결과에 대한 자체 검증 결과서 (1)

<table>
<tr><td rowspan="9">검증 대상
데이터 명세</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">한국대학교병원에서 수집한 흉부CT 데이터에 대해 영상이미지 내 환자관련정보
(환자이름, 생년월일, 환자성별, 환자번호)를 DICOM 데이터 변경 도구를 활용하여 대체처리</td></tr>
<tr><td>데이터 유형</td><td>영상·이미지</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td>DICOM</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td>DICOM</td></tr>
<tr><td>데이터 규모</td><td>100,000장 (500명 * 100장 * 2회)</td></tr>
<tr><td>데이터 크기(용량)</td><td>80GB</td></tr>
<tr><td>대상 데이터 항목명</td><td>흉부CT데이터 〈연번 2〉</td></tr>
<tr><td>가명처리 적용 기술</td><td>- 한국대학교병원에서 자체 개발한 DICOM 데이터 변경 도구를 활용하여 이미지 내 환자관련정보(환자이름, 생년월일, 환자성별, 환자번호) 대체처리</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 4월 1일 ~ 2023년 4월 7일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">회의실 (원격으로 분석실의 가상컴퓨터 접속)</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">(검증방법)
- 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행
- 실장 및 직원이 데이터를 나누어 환자관련정보 가명처리 정상 수행 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행

(검증 시 확인사항)
1. DICOM 이미지 내 환자관련정보가 다음과 같이 처리되었는지 전수 확인
① 환자이름 : Anonymized 로 대체
② 생년월일 : 임의 연월일로 대체 (1900-00-00)
③ 환자성별 : 환자성별값('M', 'F')을 'S'로 대체
④ 환자번호 : 환자구별을 위한 단순일련번호로 대체
2. 위 4가지 사항 중 하나라도 만족하지 않았을 경우 DICOM 데이터 변경도구 또는 수작업을 통해 다시 한 번 추가처리를 수행
3. 처리가 제대로 되지 않는 표본은 최종 연구 표본에서 삭제처리</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">확인 결과 이상 없음
(검수 결과, 724장이 인식되지 않았으나, 수작업을 통해 최종 가명처리 완료)</td></tr>
<tr><td rowspan="6">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>한국대학교 개인정보보호부장</td><td>가관리</td><td></td></tr>
<tr><td>한국대학교 의료정보관리실장</td><td>나정보</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>다보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>라보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>마보호</td><td></td></tr>
</table>

비정형데이터 가명처리 결과에 대한 자체 검증 결과서 (2)

<table>
<tr><td rowspan="9">검증 대상
데이터 명</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">한국대학교병원에서 수집한 골밀도 검사(DEXA) 데이터를 DICOM 포맷에서 PNG 포맷으로 변환 후,
영상이미지 내 환자관련정보(생년월일, 환자성별, 환자번호, 촬영기관명, 촬영일자·시간)을 블랙마스킹
프로그램을 활용하여 마스킹처리</td></tr>
<tr><td>데이터 유형</td><td>영상·이미지</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td>DICOM</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td>PNG</td></tr>
<tr><td>데이터 규모</td><td>1,000장 (500명 * 2장 * 1회)</td></tr>
<tr><td>데이터 크기(용량)</td><td>1GB</td></tr>
<tr><td>대상 데이터 항목명</td><td>골밀도 검사(DEXA) 데이터 〈연번 3〉</td></tr>
<tr><td>가명처리 적용 기술</td><td>- DICOM 데이터 삭제
- 파일 포맷 변환(DICOM ⇒ PNG)을 통한 메타데이터 제거
- 영상이미지 내 환자관련정보 블랙마스킹</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 4월 10일 ~ 2023년 4월 11일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">회의실 (원격으로 분석실의 가상컴퓨터 접속)</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">(검증방법)
- 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행
- 실장 및 직원이 데이터를 나누어 환자관련정보 가명처리 정상 수행 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행

(검증 시 확인사항)
1. PNG 이미지 내 환자관련정보가 문제 없이 블랙마스킹 처리되었는지 전수 확인
- 생년월일, 환자성별, 환자번호, 촬영기관명, 촬영일자·시간
2. 처리가 완벽히 수행되지 않았을 경우 수작업으로 블랙마스킹 처리 수행</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">확인 결과 이상 없음
(검수 결과, 22장이 제대로 처리되지 않아, 수작업을 통해 최종 가명처리 완료)</td></tr>
<tr><td rowspan="6">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>한국대학교 개인정보보호부장</td><td>가관리</td><td></td></tr>
<tr><td>한국대학교 의료정보관리실장</td><td>나정보</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>다보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>라보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>마보호</td><td></td></tr>
</table>

비정형데이터 가명처리 결과에 대한 자체 검증 결과서 (3)

<table>
<tr><td rowspan="9">검증 대상
데이터 명세</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">암 조직 관련 외과병리 검사결과를 자연어 처리 기술을 적용하여 정형데이터로 변환하여 활용</td></tr>
<tr><td>데이터 유형</td><td>텍스트</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td>DB</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td>DB</td></tr>
<tr><td>데이터 규모</td><td>500건</td></tr>
<tr><td>데이터 크기(용량)</td><td>11.5MB</td></tr>
<tr><td>대상 데이터 항목명</td><td>외과병리 검사결과 내 아래 데이터 항목을 추출하여 정형데이터로 변환
- Organ, Location, Tumor_size, Histologic_type, Histologic_grade, Surgical_margins, Lymph_node, P_stage, Necrosis, Nuclear grade, Microcalcification, Skin/nipple invasion, Arteriovenous_invasion, Lymphovascular invasion, Tumor Border, Architectural pattern, Intraductal_comp</td></tr>
<tr><td>가명처리 적용 기술</td><td>자연어처리기술을 통해
텍스트정보를 정형데이터로 변환</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 4월 13일 ~ 2023년 4월 14일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">회의실 (원격으로 분석실의 가상컴퓨터 접속)</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">(검증방법)
- 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 개인정보보호부 직원 2인이 함께 진행
- 직원이 데이터를 나눠 정형데이터로의 정상 변환 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 검수·처리 수행

(검증 시 확인사항)
1. 17가지의 데이터 항목이 문제없이 인식되어 정형화되었는지 확인
2. 처리가 완벽히 수행되지 않을 경우 수작업으로 DB에 데이터 입력</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">확인 결과 이상 없음
(검수 결과, 57건의 표본이 제대로 처리되지 않아, 수작업을 통해 최종 가명처리 완료)</td></tr>
<tr><td rowspan="4">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>한국대학교 개인정보보호부장</td><td>가관리</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>다보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>라보호</td><td></td></tr>
</table>

시나리오❷ : 구강질환 진단 AI 개발 사례

의료 분야 (이미지)

한국대학교병원 임상시험센터는 헬스케어 분야 AI개발 스타트업인 ㈜헬스케어소프트로부터 과학적 연구 수행을 위한 데이터 제공을 요청받아 구강 건강검진 촬영 이미지 데이터를 제공하려 한다.

☑ 데이터의 이용 목적

▶ 구강 내 충치 및 치주염 관련 질환을 진단하는 AI 개발 연구

☑ 데이터 특징

▶ (이미지 데이터) 구강 건강검진 촬영 사진

- 5,000명의 구강 건강검진 촬영영상에서 각 10장씩 이미지를 추출하여 저장한 데이터 (JPEG포맷)
- AI 모델 학습을 위해 각 구강 이미지의 구개, 혀, 위 아래 잇몸, 상악치, 하악치, 절치, 구치, 충치영역, 치주염 영역을 각각 라벨링처리하였음
- 각 이미지에 촬영된 자에 대한 메타데이터(성별, 이름, 나이, 촬영 날짜) 포함

☑ 데이터의 이용 환경

▶ (폐쇄연구분석환경 활용) 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실에서 데이터 활용, 승인된 사용자 외에는 접근 불가

▶ (자료 반입) 자료 반입시 한국대학교병원 관리자에게 요청(관리자가 자료 확인 후 반입)

▶ (자료 반출) 분석결과 반출 시 한국대학교병원 관리자에게 요청(관리자 자료 확인 후 제공)

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	SG-20230110002	접수일	2023년 1월 5일

신청자	조직/부서명	㈜헬스케어소프트 데이터연구팀		
	담당자 직위	팀장	담당자 성명	홍길동
	전화번호	03-1234-5678	이메일주소	kildong@healthcaresoft.com

처리목적		☐ 통계작성 ☑ 과학적 연구 ☐ 공익적 기록보존 구강 내 충치와 치주염 관련 질환을 진단하는 AI 모듈 개발
활용 형태		☐ 내부이용 ☑ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체		☐ 동일 개인정보처리자 ☑ 제3자(제공받는 자) 제공
처리 환경		☑ 내부 ☐ 외부
처리 장소		☑ 폐쇄 환경 ☐ 제한없음
반복 제공여부		☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☑ 온라인* ☐ 오프라인 * 폐쇄형연구분석환경에서만 접속 가능한 클라우드 서버에 업로드
제공 받는 자		(내부) ☐ 동일부서 ☐ 타부서 (외부) ☑ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	구강 건강검진 촬영 영상 및 이미지 데이터
	데이터 내역	2022년 구강 건강검진 및 질환 치료를 위해 촬영한 영상 데이터 5만건을 대상으로 학습 데이터로 필요한 단층 이미지로 변환해둔 데이터를 가명처리·제공하여 구강 질환 진단 AI 모듈을 개발하는데 활용
	가명정보 이용기간	2023년 3월 16일 ~ 2024년 3월 15일(1년)

위와 같이 가명정보 이용·제공을 신청합니다.

2023년 1월 5일

신청인(부서장명) 홍길동 (서명 또는 인)

첨부서류	1. 연구계획서 2. 개인정보 유형 분류표 3. 가명정보에 대한 안전조치 이행 확약서 4. 보안 서약서 등

② 보호법에서 정한 목적 중 가명정보 처리 목적을 명확히 설정하였는지 검토

과학적 연구 계획서

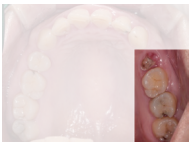
연구명	구강 내 충치와 치주염 관련 질환을 진단하는 AI 모듈 개발 연구
연구진	㈜헬스케어소프트 데이터연구팀 (책임자: 홍길동 팀장)
연구 배경 및 목적	▪ 구강검사 시 단순한 기초 검사에도 대기와 진행시간이 소요되고 질환 치료의 의사결정까지 오래 걸림 ▪ 대형병원의 충치 및 치주염 등 기초검사 진행시간 단축 및 정확도 향상 필요 ▪ 다양한 구강질환의 진단, 초기 발견 등 치료에 필요한 구강검진 기록입력 자동화 등을 지원 가능 ▪ 국민의 Self 구강검진을 통한 조기검진 및 예방관리 지원 기술 개발 필요
예상 연구 기간	2023년 3월 16일 ~ 2024년 3월 15일(1년)
연구 대상자 수	한국대학병원에서 구강 건강검진을 받은 5,000명의 내원자 연구대상자 선정기간 (2012.1.1. ~ 2022.12.31.)
연구 방법	▪ (데이터 전처리) 데이터를 전처리하여, 이미지 크기 조정, 노이즈 제거, 이미지 보정 등의 과정 수행 ▪ (데이터 라벨링) 데이터에 대해 라벨링 작업 수행, 라벨링 작업은 전문가들이 수행하며, 충치와 치주염 여부, 질환의 위치, 크기 등을 라벨링 ▪ (모델 개발) 데이터를 바탕으로 구강 질환 판별 모델 개발 ▪ (모델 평가) 개발한 모델을 평가(정확도, 재현율, F1-score 등의 지표 검증)
연구 내용	구강 건강검진 및 질환 치료를 위해 촬영한 영상 데이터 5,000건(단층 이미지 50,000장)을 학습 데이터로 학습하여 구강질환 진단 모듈 개발
기대효과 및 활용방안	구강진단 AI 모듈 개발로 진료 정확성 및 효과성 향상
붙임. 상세 연구계획서 등	

③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

개인정보 유형 분류표 (요약)

연번	항목명	데이터 유형	데이터 규모	예시	비고
1	구강 촬영데이터	이미지 (JPEG 포맷)	이미지 50,000장 (5,000명*10장) * 개인 영상 당 단층 촬영물 10개의 이미지 파일 추출 * 개인당 평균 약 6MB, 총 300GB		비정형데이터 (메타데이터 포함)

개인정보 유형 분류표 (상세)

연번	항목명	구분	설명	예시(해당 부분 강조)	비고
1	구강 촬영데이터	① 구개	입천장		이미지 (비정형데이터)
		② 혀	구강의 바닥에 위치한 근육 조직		
		③ 윗잇몸	위잇몸(상악점막)		
		④ 아래잇몸	아랫잇몸(하악점막)		
		⑤ 상악치	윗니		
		⑥ 하악치	아랫니		
		⑦ 충치 영역	절치와 구치 중 일부		
		⑧ 치주염 영역	상악점막과 하악점막 중 일부		
	1-1	성별	‘남성’, ‘여성’으로 구분	여성	구강 촬영데이터의 메타데이터
	1-2	이름	내원자의 이름	홍길동	
	1-3	나이	12세~89세까지 분포	27세	
	1-4	촬영날짜	2012.1.1.~2022.12.31.까지 분포	2021.8.28.	

활용데이터 요구 수준표

연번	항목명	구분	요구 수준	비고
1	구강 촬영데이터	① 구개	▪ 충치 및 치주염 관련 진단에 직접적으로 활용되지 않으므로, 블러링 기법 등을 통해 가명처리해도 상관 없음 ▪ 다만, 다른 영역과 구분하여 인식하여야 하므로 삭제하거나 마스킹 등으로 완전히 삭제하거나 대체하면 안 됨	이미지 (비정형데이터)
		② 혀		
		③ 윗잇몸		
		④ 아래잇몸		
		⑤ 상악치		
		⑥ 하악치		
		⑦ 충치 영역	▪ 연구목적 달성을 위해 그대로 사용 필요(학습에 직접적으로 활용)	
		⑧ 치주염 영역		
1-1	성별		▪ 분석에 필요 없으므로 삭제·대체 가능	메타데이터
1-2	이름			
1-3	나이			
1-4	촬영날짜			

식별 위험성 검토 결과보고서

※ 식별 위험성 검토 점검표를 기반으로 식별 위험성 검토 결과보고서 작성

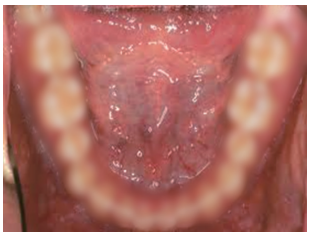
가명정보 활용목적	구강 내 충치와 치주염 관련 질환을 진단하는 AI 모듈 개발 연구	
가명처리 대상 데이터 항목	▪ 구강촬영데이터(이미지)〈 연번 1 〉 ▪ 구강촬영데이터의 메타데이터(성별,이름,나이,촬영날짜)〈 연번 1-1~1-4 〉	
데이터 위험성	식별성 유무	**〈 구강촬영이미지 〉** ▪ (구개, 혀, 윗잇몸, 아래잇몸) 해당 부위는 개인식별 가능성 거의 없으며, 타 촬영부위와 연계되어도 식별위험은 높아지지 않을 것으로 판단됨 ▪ (상악치, 하악치) 구강구조 등을 통한 개인식별 가능성은 상당히 낮으나, 메타데이터와 결합되어 분석될 시 개인식별 가능성 존재 ▪ (충치 영역, 치주염 영역) 충치·치주염 이미지를 통한 개인식별 가능성은 상당히 낮으나, 메타데이터와 결합되어 분석될 시 개인식별 가능성 존재 **〈 메타데이터 〉** ▪ 한국대학교병원 진료환자의 성별, 이름, 나이, 촬영날짜 등은 조합되었을 때 개인식별이 가능한 개인식별가능정보이며, 구강촬영이미지와 연계되면 개인식별 가능성이 더욱 높아질 수 있음

	특이정보 유무	**〈 구강촬영이미지 〉** ▪ (구개, 혀, 윗잇몸, 아래잇몸) 특이사항이 발생하기 어려운 부위임 ▪ (상악치, 하악치) 구강구조, 치열 등이 특이한 경우 개인식별 가능성 존재 ▪ (충치 영역, 치주염 영역) 심각한 구강질환이 다수 존재하는 경우, 특이한 시술 등이 이뤄진 경우 등엔 개인식별 가능성 존재 **〈 메타데이터 〉** ▪ '나이'의 경우 상당히 적거나 높은 경우 특이치로 인한 개인식별성이 발생할 수 있음
	재식별시 영향도	▪ 개인 구강에 대한 촬영정보는 재식별시 영향도는 낮은 편
처리 환경 위험성	이용 및 제공 형태	▪ 기관외부 기업에게 제공 ▪ 환자를 진료한 자와 관련 없는 제3자가 데이터를 활용하므로, 처리자가 보유한 경험·데이터로 인한 특정 환자 추정 위험성은 낮음
	처리 장소	▪ 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실 ▪ 한국대학교병원은 개인정보(가명정보)처리시스템에 대한 ISMS-P인증 취득
	다른 정보 결합 가능성	▪ 폐쇄환경분석실 관리자 승인하에 제한된 데이터·프로그램(프로그램 패키지, 라이브러리, 코드설명서 등)만 반입 가능 ▪ 분석대상 가명정보와 결합가능성 있는 데이터는 반입 제한
최종 검토의견	▪ 해당 연구는 그 자체로는 개인식별 위험성이 낮고, 재식별시 영향도가 미미한 구강촬영데이터를 다루는 연구로 전반적인 위험성이 낮은 편임 ▪ 구강구조, 특이한 시술 등으로 인한 특이정보가 발생할 수는 있으나, 해당 연구는 타 데이터의 반입 및 결합이 불가능한 클라우드 기반의 폐쇄연구 분석환경에서 연구될 뿐만 아니라, 데이터셋에 포함된 환자와 전혀 관련이 없는 제3자에 의해 연구되므로 환자에 대한 재식별·추정 가능성은 상당히 낮은 편임 ▪ 다만, 연구에는 충치·치주염 영역만 학습데이터로 활용되고, 촬영된 영역이 이미 라벨링을 통해 구분되어 있어 가명처리도 용이한 편이므로, 다음과 같이 최소한의 가명처리 후 제공 필요 - 구개, 혀 윗잇몸, 아랫잇몸 부분은 개인식별 가능성이 거의 없으므로 그대로 활용 가능 - 상악치, 하악치 부분은 연구에 필요 없고 구강구조, 치열 등이 특이한 경우 개인식별 가능성이 존재하므로 충치 영역을 제외한 부분만 블러링 처리하여 활용 - 충치 영역, 치주염 영역은 연구목적 달성을 위해 그대로 활용 가능 - 이미지의 메타데이터는 연구에 전혀 필요 없으며, 이미지와 연계되어 개인식별 위험성이 높아지므로 전체 삭제하여 활용	

④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	구강 촬영데이터	① 구개	☑ 그대로 사용	별도 처리하지 않음
		② 혀	☑ 그대로 사용	별도 처리하지 않음
		③ 윗잇몸	☑ 그대로 사용	별도 처리하지 않음
		④ 아래잇몸	☑ 그대로 사용	별도 처리하지 않음
		⑤ 상악치	☑ 블러링 처리	충치 영역 외 부분 블러링 처리
		⑥ 하악치	☑ 블러링 처리	충치 영역 외 부분 블러링 처리
		⑦ 충치 영역	☑ 그대로 사용	별도 처리하지 않음
		⑧ 치주염 영역	☑ 그대로 사용	별도 처리하지 않음
1-1	성별		☑ 삭제	연구에 필요 없으므로 삭제
1-2	이름		☑ 삭제	연구에 필요 없으므로 삭제
1-3	나이		☑ 삭제	연구에 필요 없으므로 삭제
1-4	촬영날짜		☑ 삭제	연구에 필요 없으므로 삭제

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

<table>
<tr><th>연번</th><th>항목명</th><th>세부 항목</th><th>가명처리 전</th><th>가명처리 후</th></tr>
<tr><td rowspan="3">1</td><td rowspan="3">구강
촬영데이터</td><td>⑤ 상악치</td><td>  </td><td>  </td></tr>
<tr><td colspan="2">▪ 충치 영역 외 부분 블러링 처리
- A사의 '이미지 가명처리 솔루션' 활용 (블러링 수준 3단계 적용)</td></tr>
<tr><td>⑥ 하악치</td><td>  </td><td>  </td></tr>
<tr><td colspan="3">▪ 충치 영역 외 부분 블러링 처리
- A사의 '이미지 가명처리 솔루션' 활용 (블러링 수준 3단계 적용)</td></tr>
</table>

비정형데이터 가명처리 기술의 적절성·신뢰성 관련 근거(예시)

<table>
<tr><th>연번</th><th>항목명</th><th>세부 항목</th><th>처리 기술명</th><th>처리 기술의 적절성·신뢰성 관련 근거 또는 배경</th></tr>
<tr><td rowspan="2">1</td><td rowspan="2">구강
촬영데이터</td><td>⑤ 상악치</td><td>블러링(부분)</td><td rowspan="2">▪ 한국대학교병원은 A사에서 구매한 이미지 가명처리 솔루션을 활용하여 상악치·하악치에서 충치 영역으로 라벨링되어 있는 영역 외의 부분을 부분 블러링 처리
- A사 이미지 가명처리 솔루션은 설정한 객체를 인식하여 블러링 처리 수준을 1~5단계로 설정 가능함
* A사의 솔루션 내부 테스트 결과, 객체 인식률 92%, 처리 정확도 97%로 측정
※ 객체 인식률, 처리 정확도(오류율) 증빙자료 별첨
- 현재 복원기술의 발전 수준 및 데이터 처리 환경(타 정보·복원기술 활용 불가) 등을 고려하여 블러링 수준을 3단계로 설정
* 블러링 수준의 적정성은 외부전문가를 과반 이상 포함한 적적성 평가 위원회를 구성하여 검토 완료
※ 적정성 검토위원회 결과보고서 및 회의록 별첨
- A사 가명처리 솔루션이 상악치·하악치를 인식하지 못하는 경우, 블러링 처리에 오류가 있는 경우 등이 존재하므로, 가명처리 솔루션 적용 후 처리결과에 대해 추가적인 자체 전수검사 수행</td></tr>
<tr><td>⑥ 하악치</td><td>블러링(부분)</td></tr>
</table>

비정형데이터 가명처리 결과에 대한 자체 검증 결과서

<table>
<tr><td rowspan="9">검증 대상
데이터 명세</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">한국대학교병원에서 수집한 구강검진 촬영데이터 중 상악치·하악치 부분(충치영역 이외)을
A사의 '이미지 가명처리 솔루션'을 활용하여 블러링 처리</td></tr>
<tr><td>데이터 유형</td><td>이미지</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td>JEPG</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td>JEPG</td></tr>
<tr><td>데이터 규모</td><td>50,000장 (5,000명 * 10장)</td></tr>
<tr><td>데이터 크기(용량)</td><td>300GB</td></tr>
<tr><td>대상 데이터 항목명</td><td>구강검진 촬영데이터 〈연번 1〉</td></tr>
<tr><td>가명처리 적용 기술</td><td>- 한국대학교병원에서 구매한 A사의 '이미지 가명처리 솔루션'을
활용하여 각 이미지의 상악치·하악치로 라벨링된 부분을 블러링
처리(3단계 수준)</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 1월 20일 ~ 2023년 1월 27일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">병원 내 임상시험센터 분석PC</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">(검증방법)
- 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행
- 실장 및 직원이 데이터를 나누어 블러링 처리 정상 수행 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행

(검증 시 확인사항)
1. 구강검진 이미지 내 상악치·하악치가 다음과 같이 처리되었는지 전수 확인
① 충치영역으로 라벨링된 부분을 제외한 상악치·하악치 모든 영역이 블러링처리 되었는지?
② 충치영역으로 라벨링된 부분까지 블러링처리되지는 않았는지?
2. 위 사항 중 하나라도 만족하지 않았을 경우 수작업으로 영역을 지정한 뒤, 솔루션을 활용하여 블러링 처리를 수행
3. 처리가 제대로 되지 않는 표본은 최종 연구 표본에서 삭제처리</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">확인 결과 이상 없음
(검수 결과, 최초 1,241장이 제대로 처리되지 않았으나, 수작업을 통해 최종 가명처리 완료)</td></tr>
<tr><td rowspan="6">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>한국대학교 개인정보보호부장</td><td>가관리</td><td></td></tr>
<tr><td>한국대학교 의료정보관리실장</td><td>나정보</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>다보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>라보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>마보호</td><td></td></tr>
</table>

시나리오 ❸ : 안면골 골절 진단 AI 개발 사례

의료 분야 (이미지, 영상)

한국대학교병원은 임상시험센터는 한국대학병원 내 영상의학과 연구자와 헬스케어 분야 AI개발 스타트업인 ㈜퓨처비전데이터(공동연구 컨소시움)로부터 과학적 연구 수행을 위한 데이터 제공을 요청받아 안면골(얼굴뼈) 골절 관련 Facial CT 이미지·영상 등의 데이터를 제공하려 한다.

☑ 데이터의 이용 목적

▶ 안면골 골절을 진단하는 AI 솔루션 개발 연구

☑ 데이터 특징

▶ (영상·이미지 데이터) 안면골 골절 증상으로 CT검사를 받은 80,000명(2007년~2020년)의 환자에 대한 Facial CT 영상·이미지 파일 (DICOM포맷)

- DICOM 헤더정보 포함 (환자번호, 환자이름, 성별, 생년월일 등)

▶ (증례기록지 데이터[정형데이터]) 안면골 골절 증상으로 Facial CT를 촬영한 환자에 대한 기본정보(차트번호, 검사일자, 성별, 나이)와 진단명(진단코드)

☑ 데이터의 이용 환경

▶ (폐쇄연구분석환경 활용) 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실에서 데이터 활용, 승인된 사용자 외에는 접근 불가

▶ (자료 반입) 자료 반입시 한국대학교병원 관리자에게 요청(관리자가 자료 확인 후 반입)

▶ (자료 반출) 분석결과 반출 시 한국대학교병원 관리자에게 요청(관리자 자료 확인 후 제공)

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	SG-20230110003	접수일	2023년 2월 1일

신청자	조직/부서명	한국대학교병원 영상의학과		
	담당자 직위	연구원	담당자 성명	강진단
	전화번호	03-1234-5678	이메일주소	kjd@hankook.ac.kr

처리목적		☐ 통계작성 ☑ 과학적 연구 ☐ 공익적 기록보존 안면골 골절에 대한 진단 AI 및 영상분석 솔루션 개발
활용 형태		☑ 내부이용* ☐ 제3자 제공 ☐ 결합전문기관을 통한 결합 * 한국대병원 영상의학과-(주)퓨처비전데이터(처리위탁) 컨소시움을 통한 공동연구
이용 주체		☑ 동일 개인정보처리자 ☐ 제3자(제공받는 자) 제공
처리 환경		☑ 내부 ☐ 외부
처리 장소		☑ 폐쇄 환경 ☐ 제한없음
반복 제공여부		☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☑ 온라인* ☐ 오프라인 * 폐쇄형연구분석환경에서만 접속가능한 클라우드 서버에 업로드
제공 받는 자		(내부) ☐ 동일부서 ☑ 타 부서 (외부) ☐ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	Facial CT 영상·이미지(비정형) 및 증례기록지(정형)
	데이터 내역	2007년부터 2020년까지 안면골절의 진단을 위해 Facial CT 검사를 받은 사람 중 안면골절 환자 40,000명과 정상인자 40,000명의 CT 영상·이미지 및 증례기록
	가명정보 이용기간	2023년 3월 1일 ~ 2025년 3월 1일 (3년)

위와 같이 가명정보 이용·제공을 신청합니다.

2023년 2월 1일

신청인(부서장명) 유관순 (서명 또는 인)

첨부서류	1. 연구계획서 2. 개인정보 유형 분류표 3. 가명정보 처리 위탁 계약서 등 4. 가명정보에 대한 안전조치 이행 확약서 5. 보안 서약서 등

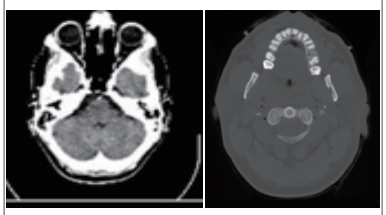
② 보호법에서 정한 목적 중 가명정보 처리 목적을 명확히 설정하였는지 검토

과학적 연구 계획서

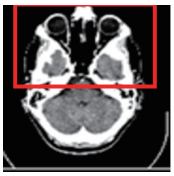
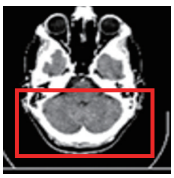
연구명	안면골 골절을 진단하는 AI 솔루션 개발 연구
연구진	한국대학교병원 영상의학과 강진단 연구원(연구책임자) ㈜ 퓨처비전데이터 데이터연구팀(공동연구 참여, 처리위탁 : AI모듈 개발)
연구 배경 및 목적	▪ 2020년 심사평가원에서 제공한 보건의료 통계에 따르면 연간 70만명에 이르는 안면골 골절 환자가 발생하고 있음 ▪ 안면골 골절은 사고나 상해 등의 이유로 매우 빈번하게 일어나므로 CT 촬영이 가능한 모든 의료기관에서의 진단이 필요하나 많은 병원에서 실시간 판독이 가능한 전문의가 없는 경우가 많음 ▪ 안면골 골절 진단이 가능한 영상의학과, 성형외과, 구강외과 등의 전문의가 없더라도 당직 의사에게 빠른 의사결정을 위한 안면골 골절을 진단할 수 있는 AI솔루션 개발 필요
예상 연구 기간	2023년 3월 1일 ~ 2025년 3월 1일 (3년)
연구 대상자 수	2007년부터 2020년까지 한국대학교병원에서 안면골 골절 의심 증상으로 Facial CT검사를 받은 환자의 CT촬영 영상 80,000건
연구 방법	▪ 본 연구는 후향적 연구로, 안면골 골절에 대한 AI를 이용한 판독을 지원하는 AI솔루션을 개발하는 것을 목표로 함 ▪ AI 기계학습과 대조군을 이용한 검증 수행
연구 내용	▪ 안면골 골절 증상으로 CT검사를 받은 80,000건의 환자 데이터 활용 - 안면골 골절진단을 받은 40,000건을 학습데이터 사용하고 대조군의 40,000건을 대상으로 검증 시행 예정 ▪ 기계학습을 통해 S02의 다양한 진단에 대해 학습하고 검증분석할 예정임
기대효과 및 활용방안	▪ Facial CT영상 분석을 통해 성형외과, 정형외과, 영상의학과 전문의가 없는 CT를 보유하고 있는 병원에서 안면골 골절에 대한 진단에 도움을 받을 수 있음 ▪ 안면골 골절의 정확한 영상 판독을 통한 적절한 치료로 후유증 발생률 최소화 가능
붙임. 상세 연구계획서 등	

③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

개인정보 유형 분류표 (요약)

연번	항목명	데이터 유형	데이터 규모	예시	비고
1	Facial CT 데이터	영상·이미지 (DICOM*)	80,000건 (97TB) - 안면골 골절 판정 데이터 (40,000건) - 정상 판정(안면골 골절X) 데이터 (40,000건)		비정형데이터 (DICOM 헤더정보 포함)
2	증례기록지	텍스트·숫자 등 (정형)	80,000건 (52MB)	차트번호, 검사일자, 환자성별, 환자나이, 진단코드 포함	정형데이터

개인정보 유형 분류표 (상세)

연번		항목명	구분	설명	예시
1		Facial CT 데이터	① 안면부	▪ 안면골 골절 여부 판단에 필요한 안면부 촬영부분	
			② 후두부 (뇌 뒷부분)	▪ 안면골 골절 여부 판단에 필요없는 후두부(뇌 뒷부분) 촬영부분	
	1-1	DICOM 헤더정보	환자 번호	▪ 환자 구분을 위해 고유하게 부여되는 번호	P158687
	1-2		환자 이름	▪ 환자의 이름	김철수
	1-3		환자 성별	▪ 1: 남자, 2: 여자	1
	1-4		환자 생년월일	▪ 환자의 생년월일 정보	1968.5.26.

2-1	증례기록지	① 차트번호	▪ 환자진료기록 차트에 고유하게 부여되는 번호로, 각 환자의 Facial CT 사진에 매핑하여 활용	01768062
2-2		② 검사일자	▪ 안면골 골절에 대한 Facial CT 검사일자	2019-06-23
2-3		③ 환자 성별	▪ 1: 남자, 2: 여자	2
2-4		④ 환자 나이	▪ 1세 단위, 12세~98세까지 분포	58
2-5		⑤ 진단명 (진단코드)	▪ S02 두개골 및 안면골의 골절 ▪ S02.0 두개원개의 골절 ▪ S02.1 두개저의 골절 ▪ S02.2 비골의 골절 ▪ S02.3 안와바닥의 골절 ▪ S02.4 광대뼈 및 상악골의 골절 ▪ S02.5 치아의 파절 ▪ S02.6 하악골의 골절 ▪ S02.7 두개골 및 안면골을 침범한 다발골절 ▪ S02.8 기타 두개골 및 안면골의 골절 ▪ S02.9 두개골 및 안면골의 상세불명 부분의 골절	S02

활용데이터 요구 수준표

※ 가명정보 처리 목적 달성에 필요한 데이터 항목과 가명처리 요구수준을 검토

연번	항목명	구분	요구 수준	비고
1	Facial CT 데이터	① 안면부	연구목적 달성에 필수적인 정보로 그대로 활용 필요	이미지 (비정형데이터)
		② 후두부 (뇌뒷부분)	연구목적 달성에 필요하지 않으므로, 마스킹 기법 등을 통해 가명처리해도 상관 없음	
1-1	DICOM 헤더정보	환자 번호	연구에 필요 없는 정보로 삭제·대체 가능	메타데이터
1-2		환자 이름	연구에 필요 없는 정보로 삭제·대체 가능	
1-3		환자 성별	연구에 필요한 정보이나, 증례기록지를 통해 확인 가능한 정보이므로 삭제·대체 가능	
1-4		환자 생년월일	연구에 필요한 정보이나, 증례기록지를 통해 확인 가능한 정보이므로 삭제·대체 가능	
2-1	증례기록지	① 차트번호	환자 구분을 위해 필요한 정보로, 단순일련번호로 대체 가능	정형데이터
2-2		② 검사일자	연구에 반드시 필요한 정보이며, 연 단위까지 범주화 가능	
2-3		③ 환자 성별	연구에 반드시 필요한 정보이며, 그대로 사용 필요	
2-4		④ 환자 나이	연구에 반드시 필요한 정보이며, 10세 단위까지 범주화 가능	
2-5		⑤ 진단명 (진단코드)	연구에 반드시 필요한 정보이며, 그대로 사용 필요	

식별 위험성 검토 결과보고서

※ 식별 위험성 검토 점검표를 기반으로 식별 위험성 검토 결과보고서 작성

<table>
<tr><td>가명정보
활용목적</td><td colspan="2">안면골 골절에 대한 진단 AI 및 영상분석 솔루션 개발</td></tr>
<tr><td>가명처리 대상
데이터 항목</td><td colspan="2">▪ Facial CT 영상·이미지(비정형데이터) 〈연번 1〉
▪ Facial CT 영상·이미지의 메타데이터(DICOM헤더) 〈연번 1-1~1-4〉
▪ 증례기록지(정형데이터) 〈연번 2〉</td></tr>
<tr><td rowspan="3">데이터
위험성</td><td>식별성 유무</td><td>〈Facial CT 영상·이미지〉
▪ 데이터 그 자체로는 개인식별 위험성이 상당히 낮은 편이나, 개인에 대한 대용량의 CT 영상·이미지에 대해 3차원 재건 등의 기술을 활용하면 얼굴 외형을 입체적으로 복원가능하며, 복원시 특이한 얼굴·외형 등이 있는 경우, 연예인 등 유명인인 경우 등엔 낮은 확률로 식별 위험이 생길 수 있음
* 가장자리 마스킹 기법을 활용하여 3차원 재건 공격 위험을 막을 수 있음
〈Facial CT DICOM 헤더정보〉
▪ 환자번호, 환자이름, 생년월일 정보는 다른 항목, 다른 정보와 결합될 시 개인식별 가능성이 있어 가명처리 필요
〈증례기록지〉
▪ 차트번호, 환자이름은 환자 개인식별정보로, 삭제 또는 대체 필요
▪ 환자 성별, 검사일자, 환자 나이 정보는 다른 정보와 결합되어 개인을 식별할 수 있는 가능성이 존재</td></tr>
<tr><td>특이정보
유무</td><td>▪ Facial CT 영상·이미지에 특이한 안면골 골절 사항 등이 존재할 수는 있으나, 데이터셋에 포함된 환자와 관련 없는 제3자에 의해 연구될 경우 식별가능성이 거의 없음</td></tr>
<tr><td>재식별시
영향도</td><td>▪ 안면골 CT 촬영정보는 재식별시 영향도는 낮은 편</td></tr>
<tr><td rowspan="3">처리 환경
위험성</td><td>이용 및 제공
형태</td><td>▪ 한국대학교병원 영상의학과 강진단 교수 연구팀 및 ㈜퓨처비전데이터가 함께 분석할 예정으로 자체 활용과 제공(위탁)의 형태를 모두 가짐
▪ 다만 사업주체가 컨소시움형태로 데이터의 제3자 제공은 해당하지 않으며, 공동의 목적 달성을 위해 한국대학교병원 내 폐쇄공간 내에서만 데이터가 활용되므로 비교적 관리가 용이한 편
▪ 클라우드 분석환경에서의 데이터 접근 권한은 연구 컨소시움 참여자 6명에게만 부여되며, 데이터 수정 권한은 강진단 교수에게만 부여</td></tr>
<tr><td>처리 장소</td><td>▪ 한국대학교병원에서 제공하는 클라우드 기반의 폐쇄연구분석환경이 갖춰진 분석실
▪ 한국대학교병원은 개인정보(가명정보)처리시스템에 대한 ISMS-P인증 취득</td></tr>
<tr><td>다른 정보
결합 가능성</td><td>▪ 다른 정보와의 연계 분석이나 결합은 예정되어 있지 않음
▪ 폐쇄환경분석실 관리자 승인하에 제한된 데이터·프로그램(프로그램 패키지, 라이브러리, 코드설명서 등)만 반입 가능
▪ 분석대상 가명정보와 결합가능성 있는 데이터는 반입 제한</td></tr>
</table>

최종 검토의견	▪ 해당 연구는 그 자체로는 개인식별 위험성이 낮고, 재식별시 영향도가 미미한 Facial CT 촬영데이터를 다루는 연구로 전반적인 위험성이 낮은 편임 ▪ Facial CT 영상·이미지를 대용량으로 활용하므로 3차원 재건 시 복원 위험이 일부 있을 수는 있으나, 타 데이터의 반입 및 결합이 불가능한 클라우드 기반의 폐쇄연구 분석환경에서 연구될 뿐만 아니라, 데이터셋에 포함된 환자와 전혀 관련이 없는 타부서 연구자, 제3자((주)퓨처비전데이터)에 의해 연구되므로 환자에 대한 재식별·추정 가능성은 상당히 낮은 편임 - 다만, 연구에는 '안면부 부분'만 활용되므로 필요없는 '후두부 부분'은 가장자리 마스킹처리하여 활용하는 것이 바람직 ▪ Facial CT DICOM 헤더정보는 연구에 필요 없는 정보('환자번호', '환자이름')이거나, 증례기록지를 통해 확인이 가능한 정보('환자 성별', '생년월일')이므로 삭제 ▪ 증례기록지의 '차트번호'는 환자구분에 필요하므로 단순일련번호로 변환하여 사용하고, 검사일자, 환자 나이 등은 연구 목적이 달성한 수준으로 범주화하여 활용 ▪ 연구목적 달성에 반드시 그대로 사용이 필요한 '환자 성별'과 '진단코드'는 다른 정보 항목들이 충분히 가명처리된 경우, 그대로 활용해도 식별 위험이 크지 않음

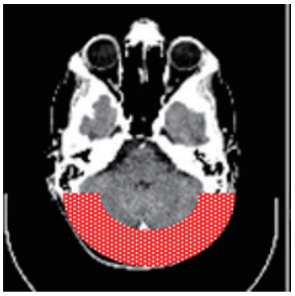
④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

항목별 가명처리 계획

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	Facial CT 데이터	① 안면부	☑ 그대로 사용	별도 처리하지 않음
		② 후두부(뇌뒷부분)	☑ 마스킹	연구에 필요 없으므로, 마스킹 처리하여 안전하게 활용
1-1	DICOM 헤더정보	환자 번호	☑ 삭제(포맷변환)	연구에 필요 없으므로 삭제 - DICOM 포맷을 TIFF 포맷으로 변경하여 저장 (헤더 삭제)
1-2		환자 이름	☑ 삭제(포맷변환)	
1-3		환자 성별	☑ 삭제(포맷변환)	증례기록지를 통해 확인 가능한 정보이므로 삭제 - DICOM 포맷을 TIFF 포맷으로 변경하여 저장 (헤더 삭제)
1-4		환자 생년월일	☑ 삭제(포맷변환)	
2-1	증례기록지	① 차트번호	☑ 대체	환자 구분을 위해 단순일련번호로 대체
2-2		② 검사일자	☑ 범주화	연 단위로 범주화
2-3		③ 환자 성별	☑ 그대로 사용	별도 처리하지 않음
2-4		④ 환자 나이	☑ 범주화(10세 단위)	10세 단위 범주화 적용, 90세 이상은 90대로 상단코딩 적용
2-5		⑤ 진단명(진단코드)	☑ 그대로 사용	별도 처리하지 않음

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

주요 비정형데이터 가명처리 수행 결과 (예시)

연번	항목명	가명처리 전	가명처리 후
1	Facial CT 데이터		

가명처리 결과 자체검증

비정형데이터 가명처리 기술의 적절성·신뢰성 관련 근거(예시)

연번	대상 항목	처리 기술명	처리 기술의 적절성·신뢰성 관련 근거 또는 배경	비고
1	Facial CT 데이터	표면 가장자리 마스킹	▪ 한국대학교병원은 B사에서 구매한 이미지 가명처리 솔루션을 활용하여 후두부 영역만 부분 마스킹 처리하였음 - B사 이미지 마스킹 솔루션은 설정한 객체를 인식하여 설정한 영역에 대해 부분적으로 마스킹 처리 가능 * B사의 솔루션 내부 테스트 결과, 객체 인식률 92%, 처리 정확도 98%로 측정 ※ 객체 인식률, 처리 정확도(오류율) 증빙자료 별첨 - B사 마스킹 솔루션이 후두부 부위를 제대로 인식하지 못하는 경우, 마스킹 처리에 오류가 있는 경우 등이 존재할 수 있으므로 솔루션 적용 후 처리결과에 대해 추가적인 자체 전수검사 수행	

비정형데이터 가명처리 결과에 대한 자체 검증 결과서

<table>
<tr><td rowspan="9">검증 대상
데이터 명세</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">한국대학교병원에서 수집한 Facial CT 데이터(DICOM 포맷을 TIFF 포맷으로 변환)에 대해
연구에 필요하지 않은 후두부(뇌 뒷부분) 영역을 B사의 '이미지 마스킹 솔루션'을 사용하여 마스킹 처리</td></tr>
<tr><td>데이터 유형</td><td>영상·이미지</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td>DICOM</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td>TIFF</td></tr>
<tr><td>데이터 규모</td><td>80,000건</td></tr>
<tr><td>데이터 크기(용량)</td><td>97TB</td></tr>
<tr><td>대상 데이터 항목명</td><td>Ficial CT 데이터 〈연번 1〉</td></tr>
<tr><td>가명처리 적용 기술</td><td>- 한국대학교병원에서 구매한 B사의 '이미지 마스킹 솔루션'을 활용하여 각
환자의 후두부 영역만 마스킹 처리</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 2월 20일 ~ 2023년 2월 27일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">회의실 (원격으로 분석실의 가상컴퓨터 접속)</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">(검증방법)
- 검증은 한국대학교병원 내부 가관리 개인정보보호부장의 주도하에 의료정보관리실장, 개인정보보호부 직원 4인이 함께 진행
- 실장 및 직원이 데이터를 나누어 환자관련정보 가명처리 정상 수행 여부를 육안으로 검수하고, 특이사항이 발생한 표본만 선별하여 전체인원이 추가 합동검수·처리 수행

(검증 시 확인사항)
1. Facial CT 후두부 영역이 다음과 같이 처리되었는지 전수 확인
① 후두부 영역 전체가 확인 불가능하도록 마스킹처리 되었는지?
② 안면부 영역까지 마스킹처리되지는 않았는지?
2. 위 사항 중 하나라도 만족하지 않았을 경우 수작업으로 영역을 지정한 뒤, 솔루션을 활용하여 마스킹 처리 수행
3. 처리가 제대로 되지 않는 표본은 최종 연구 표본에서 삭제처리</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">확인 결과 이상 없음
(검수 결과, 4,235장이 제대로 처리되지 않았으나, 수작업을 통해 최종 가명처리 완료)</td></tr>
<tr><td rowspan="6">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>한국대학교 개인정보보호부장</td><td>가관리</td><td></td></tr>
<tr><td>한국대학교 의료정보관리실장</td><td>나정보</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>다보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>라보호</td><td></td></tr>
<tr><td>한국대학교 개인정보보호부</td><td>마보호</td><td></td></tr>
</table>

시나리오 ❹ : 자율주행차 주행 시 비정상 상황인지 AI 개발 사례

교통 분야 (이미지, 영상)

한국영상연구원은 국가 R&D 사업의 일환으로 자율주행차를 통해 한국대학교 교내 도로 주행영상을 촬영하여 보유하고 있다(근거법률: 「개인정보 보호법」 제25조의2 제2호).한국영상연구원은 자율주행차의 주행환경에 대한 비정상 상황인지 AI 기술개발을 수행하고자 하는 ㈜한국자율테크에 해당 영상을 제공하고자 한다.

☑ 데이터의 이용 목적

▶ 자율주행자동차의 주행환경에 대한 비정상 상황 인지 AI 기술 연구

▶ 과학적 연구 목적으로 가명처리 후 AI 기술개발을 위한 인공지능 신경망 학습용 데이터로 활용

☑ 데이터 특징

▶ (이미지·영상 데이터) '22년 1월~2월 동안 자율주행차가 한국대학교 교내 도로 주행 상황을 촬영한 영상

- 차량 탑재 카메라로부터 MP4데이터를 취득하고 이를 PNG파일로 생성(327GB, 5만장)
- 이미지·영상 내 객체는 정적 객체(신호등, 도로 등)와 동적 객체(사람, 차량, 현수막 등)으로 구성
- 이미지·영상과 관련된 메타데이터는 별도로 생성·저장되지 않음

☑ 데이터의 이용 환경

▶ (한국영상연구원 내부 분석실 이용) USB보안 기능이 설치된 PC에서 관리하며, 이용자는 외부망이 차단된 PC에서 이용 및 활용

* 물리적 망분리, 가상 단말기(PC) 보안 준수

▶ (자료 반입) 자료 반입(오프라인으로 전달)시 한국영상연구원 관리자에게 요청(관리자가 자료 확인 후 반입), 사전 특정된 연구원 외 데이터 접근 불가

▶ (자료 반출) 분석결과 반출 시 한국영상연구원 관리자에게 요청(관리자 자료 확인 후 제공)

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	IMAGEINSTI_001_002	접수일	2023년 01월 03일

신청자	조직/부서명	한국영상연구원 / 자율주행연구실		
	담당자 직위	책임	담당자 성명	이테크
	전화번호	02-123-4567	이메일주소	lee_korea@imageinti.re.kr

처리목적		☐ 통계작성 ☑ 과학적 연구 ☐ 공익적 기록보존
		자율주행차의 비정상 주행 상황인지 AI 개발
활용 형태		☐ 내부이용 ☑ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체		☐ 동일 개인정보처리자 ☑ 제3자(제공받는 자) 제공
처리 환경		☑ 내부 ☑ 외부
처리 장소		☑ 폐쇄 환경 ☐ 제한없음
반복 제공여부		☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☐ 온라인 ☑ 오프라인
제공 받는 자		(내부) ☐ 동일부서 ☐ 타부서 (외부) ☑ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	도로 주행 영상 데이터
	데이터 내역	한국대학교 교내에서 주간과 야간, 새벽 시간대에 취득한 도로 주행 영상 데이터
	가명정보 이용기간	2023년 02월 01일 ~ 2023년 12월 31일(11개월)

위와 같이 가명정보 이용·제공을 신청합니다.

2023년 01월 03일

신청인(부서장명) 김 부 장 (서명 또는 인)

첨부서류	1. 연구계획서 2. 개인정보 유형 분류표 3. 가명정보에 대한 안전조치 이행 확약서 4. 보안 서약서 등

② 보호법에서 정한 목적 중 가명정보 처리 목적을 명확히 설정하였는지 검토

과학적 연구계획서

연구명	자율주행차의 주행환경에 대한 비정상 상황 인지 기술개발	
연구진	소속	㈜한국자율테크
	연구책임자	이 테 크
연구 배경 및 목적	▪ ㈜한국자율테크에서는 2023년 1월부터 산업부 R&D 사업의 일환으로 클라우드 기반 자율주행차 인지 기술 개발을 진행하고 있음 ▪ 상기 과제에서 본 기관은 자율주행차의 주행 경로상에서 발생할 수 있는 비정상 상황을 인지하여 자율주행차가 원활한 주행을 할 수 있도록 상황 인지 SW 개발을 추진하고 있음 ▪ 본 데이터를 자율주행차의 주행 경로상에서 발생할 수 있는 비정상적인 상황을 인지할 수 있는 인공지능 SW 개발의 학습데이터로 활용하고자 함	
예상 연구 기간	2022년 02월 01일 ~ 2023년 12월 31일 (11개월)	
연구 대상자 수	2022년 1월부터 2월까지 한국대학교 교내 캠퍼스에서 새벽, 주간, 야간에 수집한 자율주행 영상(20시간) · 이미지(50,000장) 데이터	
연구 방법	▪ 데이터의 학습에 필요한 대상 객체를 가공(라벨링)하여 활용 ▪ 상황 인지를 위한 딥러닝 인공지능 알고리즘의 학습용 데이터를 활용하며, 일부 데이터는 테스트용으로 활용함	
연구내용	▪ 자율주행차 주행환경 비정상 상황인지 기술개발 ① 주행 경로 동적, 정적 객체인지 - 자율주행차의 OEDR와 OEDR에 명시된 상황을 통해 예외 상황을 정의 - 예외 상황을 구성하는 객체(사람, 자동차, 쓰레기 더미 등)를 분류하고, 가명처리 데이터에서 학습용으로 활용할 수 있는 객체들을 구분함 - 상기 분류된 객체들을 가명처리 데이터에서 라벨링하고, 인공지능 학습데이터로 활용 ② 비정상 상항인지 기술 개발 - 상기에서 인지한 객체 정보를 활용하고, 객체 간의 관계와 시간, 장소, 날씨 등의 연관 관계를 통해 OEDR에 명시된 예외 상황 인지	
기대효과 및 활용방안	▪ 자율주행차의 상용화를 위한 인지 학습데이터로 활용함으로써 향후, 자율주행 모빌리티 산업의 활성화에 기여할 수 있음 ▪ 가명처리 데이터는 비정상 상황인지와 더불어 자율주행차의 동적 객체 인지에 활용될 수 있음	

③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

개인정보 유형 분류표

항목명	데이터 유형	데이터 규모
자율주행차 촬영 영상 및 이미지	영상(MP4), 이미지(PNG)	영상: 20시간 분량 이미지 : 50,000장(327GB)

연번	항목명	설명	예시
1	사람 전체형상	사람의 신체·형상 전체	
2	사람 얼굴	사람의 정면과 측면 얼굴	 ※ 이해를 돕기 위해 해당 부분을 확대하여 표시하였음
3	차량 전체형상	차량의 형상 전체	
4	차량 번호판	차량의 번호판 영역	 ※ 이해를 돕기 위해 해당 부분을 확대하여 표시하였음

활용데이터 요구 수준표

※ 가명정보 처리 목적 달성에 필요한 데이터 항목과 가명처리 요구수준을 검토

연번	항목명	요구 수준	비고
1	사람 전체형상	▪ 연구목적 달성을 위해 사람의 전체형상은 최대한 보존되어야 함 ▪ 해당 영역 분석을 통해 사람인지 아닌지 여부를 판단할 수 있어야 하며, 사람이 움직이는 방향·거리·속도 정보를 추출할 수 있어야 함	비정형 데이터
2	사람 얼굴	▪ 비정상적인 주행 상황 인지와 직접적인 관련이 없는 정보로, 일부 마스킹 가능 ▪ 다만, 얼굴 영역이 지나치게 과도하게 가명처리되어 AI가 사람인지 아닌지 여부를 판단할 수 없어질 경우 학습데이터로서의 가치가 무의미해질 수 있음	
3	차량 전체형상	▪ 차량의 차종(승용, SUV, 버스 등)을 구분할 수 있어야 하며, 차량의 움직이는 방향·거리·속도 정보를 추출할 수 있어야 함	
4	차량 번호판	▪ 비정상적인 주행 상황 인지와 관련 없는 정보로, 삭제되어도 무방함	

식별 위험성 검토 결과보고서

※ 식별 위험성 검토 점검표를 기반으로 식별 위험성 검토 결과보고서 작성

가명정보 활용목적	자율주행차 비정상 주행 상황 인지 AI 개발	
가명처리 대상 데이터 항목	한국대학교 교내 주행 촬영영상 (새벽, 주간, 야간) 〈 연번 1~4 〉	
데이터 위험성	식별성 유무	▪ 영상 내 사람 얼굴, 차량번호판 정보는 개인식별 위험성이 존재
	특이정보 유무	▪ (사람) 개인의 특이한 신체적·외형적 특징으로 인한 개인 식별 가능성 존재 ▪ (차량) 차량의 경우 특이한 차종(희귀 슈퍼카), 특이한 색상 등으로 인한 개인식별 가능성 존재
	재식별시 영향도	▪ 교내의 일반적인 주행상황에 대해 촬영된 영상으로, 재식별에 대한 영향도가 높지 않을 것으로 판단됨
처리 환경 위험성	이용 및 제공 형태	▪ 한국영상연구원에서 보안 USB(암호 처리)를 통해 가명처리된 영상 데이터를 제공
	처리 장소	▪ 한국영상연구원 연구실의 외부망이 차단된 PC에서 연구를 수행하며, 철저한 접근권한 통제 및 자료 반입·반출 수행
	다른 정보 결합 가능성	▪ 관리자 승인하에 제한된 데이터·프로그램(프로그램 패키지, 라이브러리, 코드설명서 등)만 반입 가능 ▪ 분석대상 가명정보와 결합가능성 있는 데이터는 반입 제한
최종 검토의견	▪ 영상 내 사람 얼굴, 차량번호판만 마스킹 처리하여 활용 ▪ 한정된 공간(교내)의 일반적인 주행상황을 촬영한 정보로 재식별시 영향도가 크지 않으며, 다른 정보와 결합이 불가능하도록 처리환경을 통제하고 있는 점 등을 고려하였을 때 사람얼굴, 차량번호판 블러링 처리 외에 특이정보 삭제 등 별도의 추가 가명처리는 불필요할 것으로 판단됨	

④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

항목별 가명처리 계획

연번	항목명	처리방법	세부방법 및 처리수준
1	사람 전체형상	☑ 그대로 사용	▪ 연구목적 달성을 위해 반드시 필요하며, 처리환경의 안전성을 고려하였을 때 개인식별 위험이 높지 않으므로 그대로 사용
2	사람 얼굴	☑ 마스킹	▪ 얼굴 영역을 사람과 컴퓨터가 식별 불가능한 수준으로 마스킹 처리
3	차량 전체형상	☑ 그대로 사용	▪ 연구목적 달성을 위해 반드시 필요하며, 처리환경의 안전성을 고려하였을 때 개인식별 위험이 높지 않으므로 그대로 사용
4	차량 번호판	☑ 마스킹	▪ 차량 번호판 영역을 사람과 컴퓨터가 식별 불가능한 수준으로 마스킹 처리

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

주요 비정형데이터 가명처리 수행 결과 (예시)

연번	항목명	가명처리 전	가명처리 후
2	사람 얼굴		
4	차량 번호판		

※ 이해를 돕기 위해 해당 부분을 확대하여 표시하였음

가명처리 결과 자체검증

비정형데이터 가명처리 기술의 적절성·신뢰성 관련 근거(예시)

<table>
<tr><th>연번</th><th>항목명</th><th>처리 기술명</th><th>처리 기술의 적절성·신뢰성 관련 근거 또는 배경</th></tr>
<tr><td rowspan="2">2

4</td><td rowspan="2">사람 얼굴

차량번호판</td><td rowspan="2">블러링(부분)</td><td>▪ 한국영상연구원은 자체보유하고 있는 ‘임베디드 장치 기반 영상데이터 개인정보 비식별화 시스템’을 활용하여 사람얼굴 영역과 차량 번호판 영역에 대해 마스킹을 수행
※ 해당 시스템은 아래와 같이 10가지 객체 종류 인식 가능
<table>
<tr><th>구분</th><th>연번</th><th>종류</th></tr>
<tr><td rowspan="4">사람얼굴</td><td>1</td><td>앞면(마스크 미착용)</td></tr>
<tr><td>2</td><td>앞면(마스크 착용)</td></tr>
<tr><td>3</td><td>측면(마스크 미착용)</td></tr>
<tr><td>4</td><td>측면(마스크 착용)</td></tr>
<tr><td rowspan="6">차량번호판</td><td>5</td><td>최신 일반(흰색)</td></tr>
<tr><td>6</td><td>이전 일반(흰색) 1:2사이즈</td></tr>
<tr><td>7</td><td>이전 일반(녹색) 1:2사이즈</td></tr>
<tr><td>8</td><td>특수 – 영업(노란색)</td></tr>
<tr><td>9</td><td>특수 – 이륜차(흰색)</td></tr>
<tr><td>10</td><td>특수 – 친환경차(파란색)</td></tr>
</table>
- 현재 복원기술의 발전 수준 및 데이터 처리 환경(타 정보·복원기술 활용 불가) 등을 고려하여 사람과 컴퓨터가 식별 및 복원이 불가능하도록 마스킹 처리
* 마스킹 방법 및 수준의 적정성은 외부전문가를 과반 이상 포함한 적적성 평가 위원회를 구성하여 검토 완료
※ 적정성 검토위원회 결과보고서 및 회의록 별첨
- 600개 파일을 통한 테스트 결과, 시스템의 객체 인식률은 99.75%, 마스킹 정확도 99.87%로 측정
※ 객체 인식률, 처리 정확도(오류율) 관련 한국기계전기전자시험연구원의 공인인증시험 성적서 별첨
- 사람 얼굴 및 차량 번호판을 인식하지 못하는 경우나 마스킹 처리가 제대로 되지 않는 경우가 존재하므로, 시스템 적용 후 처리결과에 대해 추가적인 자체 전수검사 수행</td></tr>
</table>

비정형데이터 가명처리 결과에 대한 자체 검증 결과서

<table>
<tr><td rowspan="9">검증 대상
데이터 명세</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">과학적연구 목적으로 2022년 1월부터 2월까지 한국대학교 교내에서 촬영한 자율주행 영상데이터 중
사람얼굴 영역과 차량 번호판 영역 마스킹 처리</td></tr>
<tr><td>데이터 유형</td><td>영상·이미지</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td>MP4, PNG</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td>MP4, PNG</td></tr>
<tr><td>데이터 규모</td><td>영상: 20시간 분량
이미지 : 50,000장</td></tr>
<tr><td>데이터 크기(용량)</td><td>327GB</td></tr>
<tr><td>대상 데이터 항목명</td><td>도로주행 영상데이터</td></tr>
<tr><td>가명처리 적용 기술</td><td>‘임베디드 장치 기반 영상데이터 개인정보 비식별화 시스템’을 활용하여
영상의 사람 얼굴과 차량번호판을 인식하고
마스킹 기법을 적용하여 가명처리</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 01월 10일 ~ 2023년 01월 20일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">한국영상연구원 내(외부 망 차단) 보안이 적용된 PC(보안 USB)</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">1. 전수조사 및 조사 보고서 작성
- 총 5명의 검수자가 10일간 SW 적용 후 육안 전수조사
⇒ 객체가 인식되지 않아 마스킹 처리가 안된 경우 검토
⇒ 마스킹 처리가 제대로 되지 않아 얼굴·차량번호판이 식별 가능한 경우 검토
- 전수조사 이후 마스킹 처리되지 않은 파일은 마스킹 SW로 재처리하였으며,
재처리되지 않는 파일은 수작업을 통해 마스킹 처리함
- 보고서 작성
2. 조사 내용 검토
- 조사 보고서에 대해 책임자가 검토 후 조사 완료</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">항목별 가명처리 계획에 맞게 처리되었음을 확인함</td></tr>
<tr><td rowspan="6">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>자율주행연구실 책임</td><td>가관리</td><td></td></tr>
<tr><td>자율주행연구실 선임</td><td>나정보</td><td></td></tr>
<tr><td>자율주행연구실 선임</td><td>다보호</td><td></td></tr>
<tr><td>자율주행연구실 주임</td><td>라보호</td><td></td></tr>
<tr><td>자율주행연구실 주임</td><td>마보호</td><td></td></tr>
</table>

시나리오 ❺ : 고속도로 다인승전용차로 단속 AI 개발 사례

교통 분야 (이미지)

A지자체의 교통정보센터는 교통단속 및 교통정보의 수집·분석 등을 위하여 고속도로에 CCTV를 설치하고 고속도로 통행차량 영상·이미지를 촬영하여 보관하고 있다. (근거법률: 「개인정보 보호법」 제25조 제1항 제4호, 제5호). A지자체는 AI솔루션 개발 전문기업 COREA-AI사에 과학적 연구 수행을 위한 데이터 제공을 요청받아 해당 고속도로 통행차량 영상·이미지를 제공하려 한다.

☑ 데이터의 이용 목적

▶ 재차인원 검지기반 다인승전용차로 단속 AI 개발 (학습데이터로 활용)

▶ (이미지 데이터) 고속도로 통행 차량을 촬영한 이미지 (120,000장)

- 재차인원 검지만을 위해 별도 제작된 카메라를 활용해 촬영된 사진으로, 차량번호판이 포함된 차량 앞부분은 촬영되지 않음
- 탑승된 인원의 얼굴이 함께 촬영된 경우가 많음

* 얼굴이 선명하게 촬영된 경우, 흐릿하게 촬영된 경우, 그림자·장애물 등으로 절반만 보이는 경우 등 다양

- 각 이미지마다 메타데이터 존재(촬영일시, 촬영장소)

☑ 데이터의 이용 환경

▶ (망분리 폐쇄형 개발 환경 활용) COREA-AI 개발본부에서 구축한 Private 클라우드 기반의 망분리된 폐쇄형 AI 개발 환경에서 데이터를 이용 및 활용

- 취급자에게만 접근권한 부여, 접근통제 솔루션을 통해 비인가자의 접근 통제

▶ (자료 반입) COREA-AI 관리자의 검토 및 승인 절차를 수립하여 운영(관리자가 자료 확인 후 반입)

▶ (자료 반출) 분석결과 반출 시 COREA-AI 관리자에게 요청(관리자 자료 확인 후 제공)

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	SG-20230110004	접수일	2023년 5월 8일

신청자	조직/부서명	COREA-AI 개발본부 인공지능개발팀		
	담당자 직위	수석	담당자 성명	박채주
	전화번호	03-1234-5678	이메일주소	chaejoo@corea-ai.kr

처리목적		☐ 통계작성 ☑ 과학적 연구 ☐ 공익적 기록보존 재차인원 검지기반 다인승전용차로 단속에 활용될 AI 기술개발
활용 형태		☐ 내부이용 ☑ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체		☐ 동일 개인정보처리자 ☑ 제3자(제공받는 자) 제공
처리 환경		☐ 내부 ☑ 외부
처리 장소		☑ 폐쇄 환경 ☐ 제한없음
반복 제공여부		☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☐ 온라인 ☑ 오프라인
제공 받는 자		(내부) ☐ 동일부서 ☐ 타부서 (외부) ☑ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	고속도로 통행 차량 사진
	데이터 내역	2022년 4월부터 2023년 4월까지 1년간 A지자체 OO고속도로 통행 차량에 대한 스냅사진 약 12만장
	가명정보 이용기간	2023년 6월 1일 ~ 2025년 6월 1일(2년)

위와 같이 가명정보 이용·제공을 신청합니다.

2023년 5월 8일

신청인(부서장명) 박채주 (서명 또는 인)

첨부서류	1. 연구계획서 2. 개인정보 유형 분류표 3. 가명정보에 대한 안전조치 이행 확약서 4. 보안 서약서 등

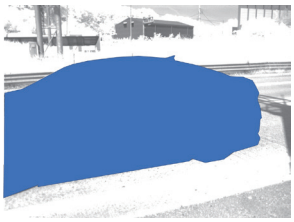
② 보호법에서 정한 목적 중 가명정보 처리 목적을 명확히 설정하였는지 검토

과학적 연구 계획서

연구명	재차인원 검지기반 다인승전용차로 단속 AI 개발
연구진	COREA-AI 개발본부 인공지능개발팀 (책임: 박채주 수석)
연구 배경 및 목적	▪ 고속도로의 수용 효율을 증대하기 위하여 다인승전용차로 제도 시행 중(1995년 2월~) -9인승이상 승용자동차 및 11,12인 승합자동차의 경우, 6인 이상이 실제 승차하였을 때만 다인승전용차로 이용 가능 ▪ 그러나 다인승전용차로 이용자의 상당 수가 불법으로 이용 중인 것으로 파악됨 ▪ 불법이용을 단속하고 교통흐름을 개선하기 위해 차량통행사진을 기반으로 재차인원을 파악하고 불법 이용자를 단속할 수 있는 시스템 개발 필요
예상 연구 기간	2023년 6월 1일 ~ 2025년 6월 1일(2년)
연구 대상자 수	▪ 2022년 4월부터 2023년 4월까지, A지자체 OO고속도로 통행 차량을 촬영한 이미지(12만장)
연구 방법	▪ A지자체가 고속도로 통행 차량 촬영 사진을 가명처리하여 COREA-AI에 제공 ▪ COREA-AI 개발본부 내 담당부서에서 라벨링 진행 ▪ 라벨링 결과를 받아서 AI 학습 수행 ▪ 수행 결과를 바탕으로 인원계수 알고리즘 개발
연구 내용	▪ 다인승전용차로 시행효과 극대화를 위해 재차인원 검지기술 개발 및 실증 수행 ▪ 다인승 전용차로 단속기능을 구현하기 위해 1) 영상촬영 기술 개발 2) 학습데이터 수집, 라벨링 및 학습 3) 차량 내 인원계수 알고리즘 개발 ▪ 다인승 전용차로 단속시스템 관련 AI 모듈 개발
기대 효과 및 활용 방안	▪ 고속도로 버스전용차로 운영 효율화 및 단속에 따른 불필요한 지체를 감소 ▪ 재차인원 검지기반 기술을 통해 재해재난 시 조난 대상자 파악 등 사회적 현안 해결에 활용 가능

③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

개인정보 유형 분류표

연번	항목명	구분	설명	예시(해당 부분 강조)	비고
1	고속도로 통행차량 촬영 데이터	① 차량 외형	특수차, 차종 등을 판단하기 위해 차량 외형이 촬영된 부분		이미지 (비정형데이터)
		② 차량 탑승자	차량 내에 탑승한 사람이 촬영된 부분		
		③ 주변 환경	차량, 탑승자 영역 외 도로 주변환경이 촬영된 부분		
1-1	촬영 일시		사진이 촬영된 날짜 및 시각	2023.1.2.14:00	각 사진에 대한 메타데이터
1-2	촬영 장소		사진이 촬영된 주소 정보	OO시 OO동 OO로 123-4	

활용데이터 요구 수준표

※ 가명정보 처리 목적 달성에 필요한 데이터 항목과 가명처리 요구수준을 검토

연번	항목명	구분	요구 수준	비고
1	고속도로 통행차량 촬영 데이터	① 차량 외형	차종에 따라 불법성 여부 판단, 재차인원 파악을 위해 감지해야 하는 영역 등이 달라지므로 연구에 필요한 정보 다만, 회사로고가 크게 적혀있는 경우, 소방차·구급차 등 특수차량 등 일반적이지 않은 소수 차량의 경우는 표본에서 제외하여도 연구목적 달성에 문제 없음	이미지 (비정형데이터)
		② 차량 탑승자	재차인원 파악, 사람인지 아닌지 여부 등만 인식하면 되므로, 블러링처리하여 활용해도 상관 없음	
		③ 주변 환경	차량부분과 주변환경을 구분하기 위해 필요하며, 개인식별 가능성이 없다면 그대로 활용 필요	
1-1	촬영 일시		분석에 필요 없으므로 삭제·대체 가능	각 사진에 대한 메타데이터
1-2	촬영 장소		분석에 필요 없으므로 삭제·대체 가능	

식별 위험성 검토 결과보고서

※ 식별 위험성 검토 점검표를 기반으로 식별 위험성 검토 결과보고서 작성

<table>
<tr><td>가명정보
활용목적</td><td colspan="2">재차인원 검지기반 다인승전용차로 단속 AI 개발</td></tr>
<tr><td>가명처리 대상
데이터 항목</td><td colspan="2">▪ 고속도로 통행 차량을 촬영사진(이미지)〈연번 1〉
▪ 고속도로 통행 차량 촬영사진의 메타데이터(촬영일시, 촬영장소)〈연번 1-1~1-2〉</td></tr>
<tr><td rowspan="3">데이터
위험성</td><td>식별성 유무</td><td>〈고속도로 통행차량 촬영사진〉
▪ (차량 외형, 주변환경 부분) 이미지 그 자체만으로는 개인식별 위험성이 거의 없으나, 촬영일시·장소 등 메타데이터와 결합될 시 식별성이 생길 수 있음
▪ (차량 탑승자) 얼굴이 선명하게 촬영된 경우, 동승자 파악이 가능한 경우 등 개인식별 가능성이 높으며, 촬영일시·장소 등 메타데이터와 결합될 시 이동동선까지 파악 가능
- 다만, 개발하려는 AI는 특정인을 분간하지 않으며 사람인지 아닌지 여부, 재차 여부를 파악하기 위한 목적으로만 활용되므로 특정 개인식별 위험성은 낮은 편
〈메타데이터〉
▪ (촬영 일시, 촬영 장소) 고속도로 통행차량 촬영사진과 결합될 시 개인식별 위험이 존재함</td></tr>
<tr><td>특이정보 유무</td><td>▪ ‘차량 외형’ 촬영부분의 경우 회사로고가 크게 적혀있는 경우, 소방차·구급차 등 특수차량의 경우 등 특이한 차량외형으로 인한 개인식별 위험이 존재</td></tr>
<tr><td>재식별시
영향도</td><td>▪ 차량통행 촬영사진은 재식별시 영향도는 낮은 편</td></tr>
<tr><td rowspan="3">처리 환경
위험성</td><td>이용 및 제공 형태</td><td>▪ 외부 기업(제3자)에 제공</td></tr>
<tr><td>처리 장소</td><td>▪ 개발본부 내 Private 클라우드를 구축하여 망분리된 폐쇄형 개발환경에서 처리
▪ 가명정보를 제공받는 기업은 개인정보(가명정보)처리시스템에 대한 ISMS-P인증을 취득하고 있음</td></tr>
<tr><td>다른 정보 결합 가능성</td><td>▪ 폐쇄환경분석실 관리자 승인하에 제한된 데이터·프로그램(프로그램 패키지, 라이브러리, 코드설명서 등)만 반입 가능
▪ 분석대상 가명정보와 결합가능성 있는 데이터는 반입 제한</td></tr>
<tr><td>최종 검토의견</td><td colspan="2">▪ 해당 연구는 차량 내 재차 인원을 파악하기 위한 목적으로 고속도로 통행차량 사진을 활용하므로, 활용 목적 및 처리 환경을 고려하였을 때 전반적인 개인식별 위험성은 낮은 편임
▪ 차량탑승자 촬영부분은 개인을 특정하여 활용할 필요가 없고, 탑승 여부만 확인하면 되므로 블러링 처리하여 활용하는 것이 적절
▪ 차량 외형의 경우 특이정보로 판단될 수 있는 차종 외형을 가명처리하여 활용할 필요가 있으며 연구 목적 달성에 심각한 영향을 주지 않는다면, 삭제 필요
▪ 메타데이터(차량일시, 차량장소)는 연구목적과 관계 없으므로 삭제하여 활용함으로써 데이터 간 연계·결합으로 인한 개인식별 위험성을 제거할 필요가 있음
- 이미지의 메타데이터는 연구목적 달성에 필요하지 않으며, 이미지와 연계되어 개인식별 위험성이 높아지므로 전체 삭제하고 이미지 정보만 활용 필요</td></tr>
</table>

④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

항목별 가명처리 계획

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	고속도로 통행차량 촬영 데이터	① 차량 외형	특이정보 파일 표본 제외	▪ 차량 외형이 일반적이지 않고, 특이하여 탑승자의 개인식별 위험성이 비교적 높은 경우(회사로고가 크게 적혀있는 경우, 소방차·구급차 등 특수차량 등)해당 파일을 표본에서 제외
		② 차량 탑승자	블러링 처리	▪ 탑승자의 위치 및 범위를 파악하고 얼굴·상체 등 차량 탑승자 영역 전체를 블러링 처리 (블러링 수준 5단계 적용)
		③ 주변 환경	그대로 사용	▪ 연구목적 달성에 필요한 정보이며 차량 외형(특이한 경우 연구표본에서 삭제), 차량 탑승자(블러링 처리), 메타데이터(삭제)를 가명처리하므로 타 정보와의 결합을 통한 개인식별 가능성이 거의 없어 별도 처리하지 않고 그대로 사용
1-1	촬영 일시		삭제	▪ 연구에 필요 없으므로 삭제
1-2	촬영 장소		삭제	▪ 연구에 필요 없으므로 삭제

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

주요 비정형데이터 가명처리 수행 결과 (예시)

<table>
<tr><th>연번</th><th>항목명</th><th>세부 항목</th><th>가명처리 전</th><th>가명처리 후</th></tr>
<tr><td rowspan="4">1</td><td rowspan="4">고속도로
통행차량
촬영 데이터</td><td rowspan="3">① 차량 외형</td><td>  </td><td rowspan="3">연구 표본에서 제외
(연구에 활용하지 않음)
※ 특수차량(응급차, 경찰차, 회사차 등), 특이한 스티커·로고·광고 등이 포함된 차량 등</td></tr>
<tr><td>  </td></tr>
<tr><td>  </td></tr>
<tr><td>② 차량 탑승자</td><td>  </td><td> 
(탑승자 위치·범위 파악)
▼
 
(1~10단계 중 5단계 수준 블러링 처리)</td></tr>
</table>

비정형데이터 가명처리 기술의 적절성·신뢰성 관련 근거(예시)

연번	항목명	세부 항목	처리 기술명	처리 기술의 적절성·신뢰성 관련 근거 또는 배경
1	고속도로 통행차량 촬영 데이터	② 차량 탑승자	블러링(부분)	▪ A지자체는 R&D사업을 통해 자체 개발한 '이미지 블러링 프로그램'을 활용하여 차량 탑승자 영역에 대한 블러링을 수행 ※ 해당 프로그램은 이미지 인식 알고리즘을 이용하여 차량 내 탑승자의 위치 및 범위를 판단할 수 있으며, 블러링 설정 값을 1단계에서 10단계로 구분하여 설정 가능 - 현재 복원기술의 발전 수준 및 데이터 처리 환경(타 정보·복원기술 활용 불가) 등을 고려하여 블러링 수준을 5단계로 설정 * 블러링 수준의 적정성은 외부전문가를 과반 이상 포함한 적적성 평가 위원회를 구성하여 검토 완료 ※ 적정성 검토위원회 결과보고서 및 회의록 별첨 - 자체 개발 이미지 블러링 프로그램은 내부 테스트 결과 객체 인식률 94%, 처리 정확도 99%로 측정 ※ 객체 인식률, 처리 정확도(오류율) 증빙자료 별첨 -탑승자를 인식하지 못하는 경우, 블러링 처리에 오류가 있는 경우 등이 존재하므로, 가명처리 솔루션 적용 후 처리결과에 대해 추가적인 자체 전수검사 수행

비정형데이터 가명처리 결과에 대한 자체 검증 결과서

<table>
<tr><td rowspan="10">검증 대상
데이터 명세</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">A지자체에서 수집한 고속도로 통행차량 촬영데이터 중 차량 탑승자 영역을 자체개발한 '이미지 블러링 프로그램'을 활용하여 블러링 처리, 차량특이치는 삭제</td></tr>
<tr><td>데이터 유형</td><td>이미지</td></tr>
<tr><td>원본 데이터 형식
(파일 포맷)</td><td>JEPG</td></tr>
<tr><td>처리 결과 데이터 형식
(파일 포맷)</td><td>JEPG</td></tr>
<tr><td>데이터 규모</td><td>120,000장</td></tr>
<tr><td>데이터 크기(용량)</td><td>18.31GB</td></tr>
<tr><td>대상 데이터 항목명</td><td>고속도로 통행차량 촬영데이터 〈연번 1〉</td></tr>
<tr><td>가명처리 적용 기술</td><td>- A지자체에서 자체개발한 '이미지 블러링 프로그램'을 활용하여 각 이미지 내 차량탑승자 영역을 블러링 처리(5단계 수준)
- 차량특이치는 직접 검수하여 삭제</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 5월 12일 ~ 2023년 5월 30일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">COREA-AI 개발본부 AI품질팀 사무실</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">(검증방법)
- 개인정보보호팀 담당 과장 주도하에 총 2명의 검수자가 Workday 10일간 전수조사 (1인당 1일 평균 약 6,000개 검수)
- 전수조사 이후 파일대상 진단 툴 조사 및 보고서 작성
(처리된 이미지의 영역정보를 기준으로 원본 이미지와 가명처리된 이미지의 영역을 비교하여 진단)

(검증 시 확인사항)
① 차량의 색상, 특수차량, 회사로고 등 특이한 차량인가?
⇒ 해당 시 표본 삭제 (연구 대상에서 제외)
② 탑승자를 식별할 수 없도록 블러링 처리가 되었는가?
⇒ 제대로 처리되지 않았을 시 표본 삭제 (연구 대상에서 제외)</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">차량외형의 특이치가 존재하여 표본을 삭제한 경우 : 1,871장
탑승자 블러링처리가 제대로 안되어 표본을 삭제한 경우 : 142장
나머지 이미지는 정상 처리되어 연구 활용 가능</td></tr>
<tr><td rowspan="5">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>개인정보보호 팀장</td><td>박지원</td><td></td></tr>
<tr><td>개인정보보호팀 과장</td><td>최우선</td><td></td></tr>
<tr><td>AI 품질팀 인턴</td><td>김용두</td><td></td></tr>
<tr><td>AI 품질팀 인턴</td><td>이용미</td><td></td></tr>
</table>

시나리오 ❻ : 한국어 대화가 가능한 AI 챗봇 개발

대화·검색 분야 (텍스트)

인공지능 챗봇 전문기업 ㈜한국데이터테크는 그간 채팅앱을 통해 수집한 고객 간의 다양한 대화 텍스트 데이터를 가명처리하여 한국어 대화가 가능한 AI 챗봇 개발에 활용하고자 한다.

☑ 데이터의 이용 목적

▶ 한국어 대화가 가능한 AI 챗봇 언어모델 개발

☑ 데이터 특징

▶ 채팅앱을 통해 수집된 고객들 간 일상 대화 데이터(텍스트)

- 고객 1,500명이 대화한 대화 데이터셋 총 20,000개 파일
- 대화 당사자 간의 일상 대화 내용이 포함되어 있으며 데이터에 정해진 형식과 법칙성이 없고 대화 주제가 다양함

* 격식을 갖추지 않은 일상의 구어체 대화가 대부분이며 한국어의 문법, 어법, 철자 등을 철저하게 준수하기보다는 편리함에 초점을 둔 대화 데이터임

- 대화 당사자의 이름, 연락처, 계좌번호, 사생활의 영역까지 개인식별위험이 있는 다양한 정보가 포함되어 있음
- 대화파일마다 대화에 참여한 사용자 계정정보가 메타데이터로 포함

☑ 데이터의 이용 환경

▶ (폐쇄형 내부 개발환경 활용) ㈜한국데이터테크 내부 공간에 마련한 철저한 폐쇄망 환경에서만 데이터 활용

- 취급자에게만 접근권한 부여, 접근통제 솔루션으로 비인가자 접근 통제

▶ (자료 반입) 내부 보안팀의 검토 및 승인 절차에 따라 운영(관리자가 자료 확인 후 반입)

▶ (자료 반출) 분석결과 반출 시 관리자에게 요청(관리자 자료 확인 후 제공)

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	SG-20230110083	접수일	2023년 3월 2일

신청자	조직/부서명	㈜한국데이터테크 AI 데이터 구축팀		
	담당자 직위	책임	담당자 성명	이신뢰
	전화번호	03-1234-5678	이메일주소	security@kor_datatech.kr

처리목적		☐ 통계작성 ☑ 과학적 연구 ☐ 공익적 기록보존 한국어 대화가 가능한 AI 챗봇 언어모델 개발
활용 형태		☑ 내부이용 ☐ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체		☑ 동일 개인정보처리자 ☐ 제3자(제공받는 자) 제공
처리 환경		☑ 내부 ☐ 외부
처리 장소		☑ 폐쇄 환경 ☐ 제한없음
반복 제공여부		☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☑ 온라인 ☐ 오프라인
제공 받는 자		(내부) ☑ 동일부서 ☐ 타부서 (외부) ☐ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	채팅앱 내 고객 간 일상대화 데이터
	데이터 내역	2018~2020년동안 자사 채팅앱 내에서 1,500명의 고객들 간 이루어진 대화 데이터셋 20,000개 파일
	가명정보 이용기간	2023년 4월 1일 ~ 2026년 4월 1일(3년)

위와 같이 가명정보 이용·제공을 신청합니다.

2023년 3월 2일

신청인(부서장명) 이 신 뢰 (서명 또는 인)

첨부서류	1. 연구계획서 2. 개인정보 유형 분류표

② 보호법에서 정한 목적 중 가명정보 처리 목적을 명확히 설정하였는지 검토

과학적 연구 계획서

연구명	한국어 대화가 가능한 AI 챗봇 언어모델 개발
연구진	㈜한국데이터테크 AI 데이터 구축팀 (연구책임: 이신뢰 책임)
연구 배경 및 목적	▪ ㈜한국데이터테크는 사람과 한국어로 자유롭게 일상적인 대화를 할 수 있는 AI 챗봇 서비스를 제공하고자 하며, 이를 위해 한국어 대화의 맥락과 행간을 깊이 있게 이해하고 실제 사람과 대화하는 것 같은 대화 경험을 제공하기 위한 한국어 AI 언어모델을 개발하고자 함
예상 연구 기간	2023년 4월 1일 ~ 2026년 4월 1일 (3년)
연구 대상자 수	2018~2020년동안 자사 채팅앱 내에서 1,500명의 고객들 간 이루어진 20,000개의 대화 텍스트 데이터셋 파일
연구 방법	▪ 대화 텍스트 데이터셋을 가명처리하여 학습용 데이터베이스를 구축하고, 이를 컴퓨터가 이해할 수 있는 벡터 형태로 변환한 뒤, 대화 문맥 내 단어들 간 등장 확률을 학습
연구 내용	▪ 학습용 데이터베이스를 통해 상대방의 발화를 이해하고 적절한 답변을 선택하기 위한 모델, 문맥을 파악하고 부적절한 답변을 걸러내는 모델을 개발 ▪ AI 챗봇의 대답은 외부에 공개되기 때문에 가명정보를 통해 학습된 대화가 그대로 발화되지 않도록, 답변에 활용되는 데이터베이스는 별도로 구축하여 가명정보 노출 가능성을 근본적으로 차단 - 답변 데이터베이스는 ㈜한국데이터테크 자체 생성AI 모델이 만들어낸 문장과 내부 AI 데이터구축팀이 직접 작성한 문장으로 구성
기대 효과 및 활용 방안	▪ 쉽고 자연스럽게 대화할 수 있는 AI 챗봇을 개발하여 사회 소외계층, 디지털 소외계층도 함께 누릴 수 있는 챗봇 생태계 조성 ▪ 한국어 대화 맥락 파악 고도화를 통한 맞춤형 상담·검색 기능 확대

③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

개인정보 유형 분류표

연번	항목명	구분	설명	비고
1	채팅앱 내 고객 간 일상대화 데이터	① 이름	대화에서 언급된 대화 당사자 또는 제3자의 이름	텍스트 (비정형 데이터)
		② 생년월일	대화에서 언급된 대화 당사자 또는 제3자의 생년월일	
		③ 주민등록번호	대화에서 언급된 대화 당사자 또는 제3자의 주민등록번호	
		④ 연락처	대화에서 언급된 대화 당사자 또는 제3자의 연락처 정보	
		⑤ 카드번호	대화에서 언급된 대화 당사자 또는 제3자의 카드번호	
		⑥ 계좌번호	대화에서 언급된 대화 당사자 또는 제3자의 계좌번호	
		⑦ 차량번호	대화에서 언급된 대화 당사자 또는 제3자의 차량번호	
		⑧ 학번	대화에서 언급된 대화 당사자 또는 제3자의 학번	
		⑨ 여권번호	대화에서 언급된 대화 당사자 또는 제3자의 여권번호	
		⑩ 운전면허번호	대화에서 언급된 대화 당사자 또는 제3자의 운전면허번호	
		⑪ 외국인등록번호	대화에서 언급된 대화 당사자 또는 제3자의 외국인등록번호	
		⑫ 건강보험증 번호	대화에서 언급된 대화 당사자 또는 제3자의 건강보험증 번호	
		⑬ 상세 주소	대화에서 언급된 대화 당사자 또는 제3자의 상세 주소 정보	
		⑭ 아이디	대화에서 언급된 대화 당사자 또는 제3자의 아이디	
		⑮ 비밀번호	대화에서 언급된 대화 당사자 또는 제3자의 비밀번호	
		⑯ 이메일	대화에서 언급된 대화 당사자 또는 제3자의 이메일 주소	
		⑰ URL	대화에서 언급된 대화 당사자 또는 제3자와 관련 있을 수 있는 URL 주소	
		⑱ 나이	대화에서 언급된 대화 당사자 또는 제3자의 나이 정보	
1-1	대화 사용자 계정정보		대화 사용자에 대해 고유하게 부여되는 사용자 계정정보 해당 대화가 어떤 사용자와 연결되는지 파악하기 위해 사용	메타 데이터 (정형 데이터)

(참고) 데이터 처리목적 및 활용 방식을 고려한 가명처리 대상 선정

㈜한국데이터테크가 개발하고자 하는 한국어 대화 AI 챗봇은 가명정보를 통해 학습하여 만들어진 '학습용 데이터베이스'와 실제 AI 챗봇의 답변 대화 구성에 활용되는 '답변 데이터베이스'가 완전히 분리되어 운영되므로 가명처리된 텍스트데이터 내의 문장이 AI 챗봇의 답변에 그대로 노출될 확률이 없음

즉, 가명처리한 텍스트 데이터는 AI 언어모델을 학습하는데만 활용되고 실제 AI챗봇 답변을 통해 발화되지 않으므로, 이름, 생년월일, 연락처 등 개인식별성이 높은 항목들만 철저히 필터링하여 삭제·대체한다면 이용자 프라이버시가 침해될 확률을 크게 낮출 수 있음

활용데이터 요구 수준표

연번	항목명	구분	요구 수준	비고
1	채팅앱 내 고객 간 일상대화 데이터	① 이름	대화상대 구분 등 연구목적에 필요하여 삭제하면 안 되며, 임의의 다른 이름으로 치환하여 활용 필요	텍스트 (비정형 데이터)
		② 생년월일	대화상대 연령에 맞는 대화문 생성 등 연구목적에 필요하여 삭제하면 안 되며, 임의의 다른 값으로 치환하여 활용 필요	
		③ 주민등록번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		④ 연락처	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑤ 카드번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑥ 계좌번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑦ 차량번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑧ 학번	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑨ 여권번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑩ 운전면허번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑪ 외국인등록번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑫ 건강보험증 번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑬ 상세 주소	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑭ 아이디	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑮ 비밀번호	연구목적에 필요 없어 삭제 가능(해당 정보 포함 전체 문장 삭제)	
		⑯ 이메일	이메일이 포함된 대화패턴 인식이 필요하여 토큰으로 대체 필요	
		⑰ URL	URL이 포함된 대화패턴 인식이 필요하여 토큰으로 대체 필요	
		⑱ 나이	발화자 나이에 따른 대화문 생성 등 연구목적에 필요하여 삭제하면 안 되며, 비슷한 나이대로 범주화하여 활용 필요	
1-1	대화 사용자 계정정보		대화 사용자와 일상대화 데이터의 연결은 불필요하지만 대화 데이터 발화자의 구분은 필요하므로 랜덤ID로 대체 필요	메타 데이터 (정형 데이터)

식별 위험성 검토 결과보고서

※ 식별 위험성 검토 점검표를 기반으로 식별 위험성 검토 결과보고서 작성

가명정보 활용목적	한국어 대화가 가능한 AI 챗봇 언어모델 개발	
가명처리 대상 데이터 항목	▪ 2018~2020년동안 자사 채팅앱 내에서 1,500명의 고객들 간 이루어진 20,000개의 대화 텍스트 데이터셋 파일〈연번 1〉 ▪ 대화 텍스트 데이터셋의 메타정보(대화 사용자 계정정보)〈연번 1-1〉	
데이터 위험성	식별성 유무	▪ 이름, 주민등록번호, 연락처, 카드번호, 계좌번호, 운전면허번호, 여권번호, 외국인등록번호, 건강보험증번호, 상세 주소 등은 개인식별 위험성이 상당히 높음 ▪ 아이디, 생년월일, 학번, 나이 등도 다른 대화내용 및 메타데이터와 연계될 시 개인식별 위험성이 존재 ▪ 개인정보로 선정한 항목 외에 일반대화내용도 상황에 따라 다른 대화내용과 연계될 시 개인식별 위험성이 생길 수는 있으나, AI 언어모델 학습에만 활용되고 AI챗봇의 답변을 통해 발화되지 않는다면 개인식별 위험성이 미미함 ▪ 또한, 메타데이터(대화 사용자 계정정보)를 삭제하여 특정 개인과의 연결성을 제거한 형태로 활용되면 개인식별 위험성이 더욱 낮아질 것임
	특이정보 유무	▪ 일반대화내용에 포함된 대화 맥락 및 특수 상황에 대한 언급이 존재할 시 개인식별 위험이 높아질 수 있음
	재식별시 영향도	▪ 일상대화의 내용이 다양하게 담겨있고 상황에 따라 사생활의 내밀한 내용이 포함될 수 있어 재식별 시 영향도가 높은 편임
처리 환경 위험성	이용 및 제공 형태	▪ 동일 개인정보처리자, 동일 부서 내 활용
	처리 장소	▪ 회사 내부의 철저한 폐쇄망 개발환경 활용
	다른 정보 결합 가능성	▪ 관리자 승인하에 데이터(프로그램 패키지, 라이브러리, 코드설명서 등) 반입이 가능함 ▪ 기존 가명정보와 결합가능성 있는 데이터는 반입이 제한됨
최종 검토의견	▪ 내 개인식별성이 있는 항목들을 빠짐없이 선별하여 삭제 또는 치환하여야 함 - 특히 고유식별정보인 주민등록번호, 운전면허번호, 여권번호, 외국인등록번호는 반드시 삭제하여야 함 - 나이정보는 개인식별 위험이 있으나, 발화자 나이에 따른 대화문 생성에 필요한 정보이므로 5세 단위로 범주화하여 활용하는 것이 적절 ▪ 메타데이터인 대화 사용자 계정정보는 파기하고 랜덤 ID로 대체하여 활용함으로써 특정 개인과의 연결성을 제거하여 활용하여야 함 ▪ 가명처리된 텍스트데이터가 활용되는 학습데이터베이스와 AI 챗봇을 통해 발화되는 답변데이터베이스가 분리되어 있어, 가명정보가 AI 챗봇의 답변으로 그대로 발화되지 않기에 가명정보의 직접적인 노출로 인한 개인식별 위험은 발생하지 않을 것으로 판단됨	

④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

항목별 가명처리 계획

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	채팅앱 내 고객 간 일상대화 데이터	① 이름	치환	▪ 인공지능 모델(Personal Name Recognition) 기반으로 문장 내에서 이름이라고 판단되는 단어가 발견되면, 해당 이름을 관련 통계에 따른 한국인 이름 분포에 따라 랜덤하게 생성한 임의의 이름으로 치환
		② 생년월일	치환	▪ 생년월일 패턴이 발견되면 해당 정보를 랜덤하게 생성한 임의의 생년월일로 치환함
		③ 주민등록번호	삭제	▪ 주민등록번호 패턴(000000-0000000)을 포함하는 문장 전체를 삭제
		④ 연락처	삭제	▪ 전화번호 패턴(00 또는 000-000-0000 등)을 포함하는 문장 전체를 삭제 ▪ 숫자를 한글로 표현(예: 일이삼, 영, 공 등)한 경우까지 고려하여 삭제하며, 한글 숫자와 아라비아 숫자가 병용된 패턴 역시 문장 삭제
		⑤ 카드번호	삭제	▪ 카드번호 패턴을 포함하는 문장 전체를 삭제
		⑥ 계좌번호	삭제	▪ 계좌번호 패턴을 포함하는 문장 전체를 삭제
		⑦ 차량번호	삭제	▪ 차량번호 패턴을 포함하는 문장 전체를 삭제
		⑧ 학번	삭제	▪ 학번 등의 패턴을 포함하는 문장 전체를 삭제
		⑨ 여권번호	삭제	▪ 여권번호 패턴을 포함하는 문장 전체를 삭제
		⑩ 운전면허번호	삭제	▪ 운전면허번호 패턴을 포함하는 문장 전체를 삭제
		⑪ 외국인등록번호	삭제	▪ 외국인등록번호 패턴을 포함하는 문장 전체를 삭제
		⑫ 건강보험증 번호	삭제	▪ 건강보험증번호 패턴을 포함하는 문장 전체를 삭제
		⑬ 상세 주소	삭제	▪ 지역명 주소로 구성된 읍·면·동·리 상세주소 및 도로명 주소로 구성된 길 이름이 포함된 문장 전체를 삭제 ▪ 아파트 동호수 등의 패턴이 포함된 문장 전체를 삭제 ▪ 길안내, 네비게이션 앱 메시지 삭제
		⑭ 아이디	삭제	▪ 숫자, 영문, 특수문자 등으로 구성된 7자리 이상의 단어가 포함되면 해당 문장 전체를 삭제 ▪ '아이디', 'ID' 등의 문구가 포함되면 해당 문장 전체를 삭제
		⑮ 비밀번호	삭제	▪ 숫자, 영문, 특수문자 등으로 구성된 7자리 이상의 단어가 포함되면 해당 문장 전체를 삭제 ▪ '비밀번호' '패스워드' 등의 문구가 포함되면 해당 문장 전체를 삭제 ▪ 아파트 비밀번호 등에 자주 쓰이는 특수문자(#, *) 등으로 시작하거나 끝나는 패턴이 있으면 해당 문장을 삭제
		⑯ 이메일	치환	▪ '@' 기호가 포함되어 있는 이메일 패턴을 토큰 [MAIL]로 치환
		⑰ URL	치환	▪ http:// 및 .com 등 URL 패턴 부분을 토큰 [URL]로 치환
		⑱ 나이	범주화	▪ 5세 단위 범주화 (16~20세, 21~25세, 26세~30세, 31세~35세 등)
1-1	대화 사용자 계정정보		대체	▪ 사용자 계정정보를 파기하고 랜덤한 ID로 대체

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

비정형데이터 가명처리 기술의 적절성·신뢰성 관련 근거(예시)

연번	항목명	세부 항목	처리 기술명	처리 기술의 적절성·신뢰성 입증 관련 근거 또는 배경
1	채팅앱 내 고객 간 일상대화 데이터	① 이름 ② 생년월일 ⑯ 이메일 ⑰ URL	치환	▪ '이름', '생년월일', '이메일', 'URL' 항목은 자체 필터링 솔루션을 활용하여 임의의 값 또는 토큰으로 대체 - 자체 필터링 솔루션은 자연어 처리 기준으로 검출 정확성은 89%, 처리 정확도는 97%로 측정 ※ 검출 정확도, 처리 정확도(오류율) 증빙자료 별첨 - 검출 정확도가 100%가 아니므로, 솔루션 적용 후 처리결과에 대해 추가적인 자체 전수검사를 수행
		③ 주민등록번호 ④ 연락처 ⑤ 카드번호 ⑥ 계좌번호 ⑦ 차량번호 ⑧ 학번 ⑨ 여권번호 ⑩ 운전면허번호 ⑪ 외국인등록번호 ⑫ 건강보험증 번호 ⑬ 상세 주소 ⑭ 아이디 ⑮ 비밀번호	삭제	▪ 자체 필터링 솔루션을 활용하여 등록된 패턴을 기준으로 해당 항목을 검출하고, 검출된 항목이 포함된 문장 전체를 삭제 처리 - 자체 필터링 솔루션은 자연어 처리 기준으로 검출 정확성은 89%, 처리 정확도는 97%로 측정 ※ 검출 정확도, 처리 정확도(오류율) 증빙자료 별첨 - 검출 정확도가 100%가 아니므로, 솔루션 적용 후 처리결과에 대해 추가적인 자체 전수검사를 수행
		⑱ 나이	범주화	▪ 자체 필터링 솔루션을 활용하여 나이값을 검출하고, 5세 단위로 범주화하여 토큰으로 대체 - 자체 필터링 솔루션은 자연어 처리 기준으로 검출 정확성은 89%, 처리 정확도는 97%로 측정 ※ 검출 정확도, 처리 정확도(오류율) 증빙자료 별첨 - 검출 정확도가 100%가 아니므로, 솔루션 적용 후 처리결과에 대해 추가적인 자체 전수검사를 수행

비정형데이터 가명처리 결과에 대한 자체 검증 결과서

<table>
<tr><td rowspan="8">검증 대상
데이터 명세</td><td colspan="3">개요</td></tr>
<tr><td colspan="3">㈜한국데이터테크는 채팅앱을 통해 수집된 고객들 간 일상대화 데이터 중
개인식별 위험성이 높은 항목들을 자체 필터링 솔루션을 활용하여 가명처리(치환·삭제·범주화)하였음</td></tr>
<tr><td>데이터 유형</td><td colspan="2">텍스트</td></tr>
<tr><td>원본 데이터 형식(파일 포맷)</td><td colspan="2">JSON</td></tr>
<tr><td>처리 결과 데이터 형식(파일 포맷)</td><td colspan="2">JSON</td></tr>
<tr><td>데이터 규모 및 크기</td><td colspan="2">고객 1,500명이 대화한 일상대화 데이터 20,000개 파일</td></tr>
<tr><td>대상 데이터 항목명</td><td colspan="2">일상대화 데이터 텍스트 파일 〈연번 1〉</td></tr>
<tr><td>가명처리 적용 기술</td><td colspan="2">- ㈜한국데이터테크에서 자체 개발한 필터링 솔루션을 활용하여 주요 개인식별 위험 항목을 검출한 뒤 치환, 삭제, 범주화 처리
- 일반대화정보 내 개인식별 가능성이 있는 단어, 문장은 직접 검수하여 삭제 또는 대체</td></tr>
<tr><td>자체 검증 기간</td><td colspan="3">2023년 3월 15일 ~ 2023년 3월 25일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="3">㈜한국데이터테크 데이터센터</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="3">(검증방법)
- 개인정보보호팀 담당 팀장 주도하에 총 4명의 검수자가 Workday 10일간 전수조사행

(검증 시 확인사항)
① 치환되어야 할 항목(이름, 생년월일, 이메일, URL)들이 정해진 텍스트나 정해진 범주 내 랜덤값으로 치환되었는가?
⇒ 치환되지 않았을 시 수작업으로 변환 실시
② 삭제되어야 할 문장(주민번호, 연락처 등 포함)들이 삭제되었는가?
⇒ 삭제되지 않았을 시 수작업으로 삭제 실시
③ 나이정보가 범주화되어 토큰으로 대체 되었는가?
⇒ 대체되지 않았을 시 수작업으로 범주화 및 토큰화 실시
④ 가명처리되어야 할 항목들이 인식·검출되지 않아 제대로 처리가 되지 않은 문장이 존재하는가?
⇒ 해당 문장 발견시 수작업으로 추가 가명처리 실시
⑤ 가명처리 항목 외에도 일반대화정보에서 개인식별 가능성이 있는 단어나 문장이 존재하는가?
⇒ 확인 시 팀장 주도 최종회의에서 삭제·대체여부 결정하여 추가 가명처리 실시</td></tr>
<tr><td>자체 검증 결과</td><td colspan="3">20,000개 파일 중 4,765개 파일에 대해 오류 발견되어 추가 가명처리 실시 완료</td></tr>
<tr><td rowspan="6">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>개인정보보호 팀장</td><td>가관리</td><td></td></tr>
<tr><td>개인정보보호팀 직원</td><td>나정보</td><td></td></tr>
<tr><td>AI 데이터 구축팀 직원</td><td>다보호</td><td></td></tr>
<tr><td>AI 데이터 구축팀 직원</td><td>라보호</td><td></td></tr>
<tr><td>AI 데이터 구축팀 직원</td><td>마보호</td><td></td></tr>
</table>

시나리오 ❼ : 콜센터 직원 실습용 가상상담 시나리오 생성 AI 개발

대화·교육 분야 (음성, 텍스트)

인터넷통신 기업 ㈜코리아인터넷은 인터넷·통신 상품 관련 고객상담 시 분쟁방지 및 서비스 품질 향상 목적으로 통화내용이 녹음됨을 고객에게 공지하고 상담 음성정보를 수집·보관하고 있다. ㈜코리아인터넷은 해당 음성정보를 가명처리한 뒤, 자사 콜센터 상담직원들에게 상황별·업무별 상담실습 교육을 진행하기 위한 목적으로 가상상담 시나리오를 생성하는 AI를 개발하고자 한다.

☑ 데이터의 이용 목적

▶ 콜센터 직원 상담 실습교육용 가상상담 시나리오 생성 AI 모듈 개발

☑ 데이터 특징

▶ 상담사와 고객 간 질의-응답으로 이루어진 음성 데이터(WAV 포맷)

- 상담 목적(❶단순 문의, ❷A/S, ❸결제, ❹교환, ❺설치, ❻반품, ❼기타), 성별, 연령, 주거 지역별로 샘플링한 상담음성 녹취파일 총 10,000개 파일
- 각 파일 당 평균 3분 분량으로 총 500시간 (60GB)
- 고객 및 상담사의 음성이 녹음되어 있으며 대화 내용에 고객의 이름, 생년월일, 주소, 연락처 등 개인식별성 있는 정보가 포함

☑ 데이터의 이용 환경

▶ (폐쇄형 내부 개발환경 활용) ㈜한국데이터테크 내부 공간에 마련한 철저한 폐쇄망 환경에서만 데이터 활용

- 취급자에게만 접근권한 부여, 접근통제 솔루션으로 비인가자 접근 통제

▶ (자료 반입) 내부 보안팀의 검토 및 승인 절차에 따라 운영(관리자가 자료 확인 후 반입)

▶ (자료 반출) 분석결과 반출 시 관리자에게 요청(관리자 자료 확인 후 제공)

① 사전준비 단계에서 필요서류가 법/제도 목적에 적합하게 작성되었는지 검토

가명정보 이용·제공 신청서

접수번호	SG-20230110085	접수일	2023년 1월 10일

신청자	조직/부서명	㈜코리아인터넷 빅데이터본부 음성AI팀		
	담당자 직위	책임	담당자 성명	박안전
	전화번호	03-1234-5678	이메일주소	park_security@kor_internet.kr

처리목적		☐ 통계작성 ☑ 과학적 연구 ☐ 공익적 기록보존
		콜센터 직원 실습용 가상상담 시나리오 생성 AI 모듈 개발
활용 형태		☑ 내부이용 ☐ 제3자 제공 ☐ 결합전문기관을 통한 결합
이용 주체		☑ 동일 개인정보처리자 ☐ 제3자(제공받는 자) 제공
처리 환경		☑ 내부 ☐ 외부
처리 장소		☑ 폐쇄 환경 ☐ 제한없음
반복 제공여부		☑ 1회 제공 ☐ 시계열 분석 등을 위한 반복 제공(회 예정)
제공 방법		☑ 온라인 ☐ 오프라인
제공 받는 자		(내부) ☑ 동일부서 ☐ 타부서 (외부) ☐ 제3자, 결합전문기관 제공
신청 명세	가명처리 대상 명칭	콜센터 녹취 파일(WAV)
	데이터 내역	2022년 녹취통합 시스템에 저장되어 파일 중 학습용으로 샘플링(상담목적, 성별, 연령, 주거지역 고려)한 10,000개 파일
	가명정보 이용기간	2023년 3월 16일 ~ 2024년 3월 16일(1년)

위와 같이 가명정보 이용·제공을 신청합니다.

2023년 1월 10일

신청인(부서장명) 박 안 전 (서명 또는 인)

첨부서류	1. 연구계획서 2. 개인정보 유형 분류표

② 보호법에서 정한 목적 중 가명정보 처리 목적을 명확히 설정하였는지 검토

과학적 연구 계획서

연구명	콜센터 직원 실습교육용 가상상담 시나리오 생성 AI 모듈 개발
연구진	㈜코리아인터넷 빅데이터본부 음성AI팀 (연구책임: 박안전 책임)
연구 배경 및 목적	▪ ㈜코리아인터넷에서는 자사 콜센터 직원을 대상으로 상담 업무 능력 향상을 위한 교육에 활용할 교육용 음성생성 AI 기술개발을 진행하고 있음 ▪ 다양한 상담 목적, 고객 유형에 따라 단순 상담부터 난이도 높은 상담까지 문제없이 대응할 수 있도록 AI를 통해 상황별 고객상담 음성을 생성하여 상담직원 교육에 활용하고자 함
예상 연구 기간	2023년 3월 15일 ~ 2024년 3월 16일 (1년)
연구 대상자 수	다양한 발화자(성별, 연령, 지역 등)의 특징을 고려하여 샘플링한 10,000건의 녹취 파일
연구 방법	언어모델 개발과 언어모델 학습을 통한 상담 데이터 생성 모델 개발
연구 내용	▪ 상담음성을 텍스트로 변환(STT)한 뒤, 개인식별성이 있는 단어, 문장 등을 제거 또는 치환하여 개인식별 위험성을 제거 ▪ 가명처리된 텍스트 데이터를 기반으로 고객 질문 의도 분류 모델 및 상담사 답변 추천 모델 개발 - 고객 질문 입력 이후 고객 질문 의도를 분류하는 합성곱 신경망 기반 분류 모델 - 모델이 분류한 결과를 실제 고객 질문의도 범주와 비교하여 오차를 계산하고, 오차 역전파(backpropagation) 기법을 통해 모델 학습 진행 - 성능평가 지표를 통해 학습된 모델을 평가 ▪ 개발된 AI 모델에 의해 새롭게 생성된 콜센터 상담 시나리오(텍스트) 생성 ▪ 해당 상담 시나리오를 음성파일로 재변환(TTS)하여 교육실습자료로 활용
기대 효과 및 활용 방안	▪ 다양한 분야의 상담 데이터와 상담 사례별 데이터를 통해 상담 지원 대상 교육 효과 증대 ▪ 상담 만족도 향상을 위한 상담업무 효율 및 비즈니스 효과 증대

③ 가명처리 단계에서 데이터의 자체 식별 위험성, 처리 환경의 식별 위험성 등 판단 항목을 누락 없이 검토하였는지 개인식별 위험성 체크리스트 및 결과보고서 기반으로 검토

연구목적에 따른 연구방법 설정 및 데이터 전처리 (예시)

- (연구목적 검토) 가상상담 시나리오 생성 AI 개발에는 상담 목적, 고객 특성(성별, 연령, 지역)에 따른 요청·문의 내용 및 질의-응답에 따른 대화 흐름이 중요하며 실제 고객 및 상담사의 음성 자체는 필요하지 않음
- (데이터 전처리) 음성변환(STT, Speech To Text) 엔진을 통해 상담음성 녹음파일의 음성을 모두 텍스트로 변환하고, 변환된 텍스트에서 제대로 변환되지 않은 단어·문장, 비문 등을 전문 작업자가 보완·수정하여 텍스트 데이터셋을 구축

※ 음성데이터를 텍스트데이터셋으로 변환하여 활용할 시 상담음성 데이터 원본에 포함된 고객 및 상담사의 실제 음색, 억양, 발음 등을 통한 개인식별 위험성을 제거할 수 있음
⇒ (음성→텍스트)로 변환된 데이터를 가명처리하여 AI 학습에 활용하고자 함

데이터 전처리(STT) 완료된 텍스트 데이터셋 (예시)

발화자	대화 텍스트
(고객)	아 저희가 지금 인터넷TV 그 사용하고 있는데 속도가 좀 너무 안 나와서요
(상담원)	아 설치 시보다 인터넷 속도가 너무 느리시다는 말씀이신 거죠
(고객)	네
(상담원)	이용 많이 불편하실 텐데 너무 죄송하구요 우선 고객 확인 먼저 하겠습니다. 홍길동 고객님 본인 맞으실까요?
(고객)	맞습니다
(상담원)	가입해주신 고객님 혹시 1987년 7월 22일생 맞으실까요?
(고객)	네 맞습니다
(상담원)	확인 감사드리구요. 인터넷 속도저하 부분이 와이파이나 인터넷 모두 동일한 현상인가요?
(고객)	와이파이가 조금 많이 느려요.. 그래서 TV가 자꾸 끊기거든요
···(중략)···	
(상담원)	인터넷 가입하신지 12년 5개월되셨는데, 가족 세 분과 가족할인 요금제로 사용하고 계시거든요. 다음 달 8일부터 신규 상품으로 변경하시면서 배우자분이랑 아드님도 가족결합할인으로 추가하시겠다는거 맞으시죠? 상품먼저 배송해드릴게요. 주소가 어떻게 되시죠?
(고객)	서울시 신뢰동 신뢰아파트 백일동 구백삼호입니다.
(상담원)	네 고객님 서울시 신뢰동 신뢰아파트 백일동 구백삼호로 배송해드리겠습니다
(고객)	찾기가 좀 어려우실텐데 신뢰역 3번출구 앞에서 오른쪽으로 쭉 올라오시면 돼요

발화자	대화 텍스트
(상담원)	네 기사분께 전달드려놓겠습니다. 기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 팔천오백이니까 모르는 연락처라도 통화 잘 부탁드리겠습니다
(고객)	네 알겠습니다. 혹시 안 받으면 저희 회사쪽으로 연락주시면 됩니다. 공이 일이삼사 오육칠팔입니다.
(상담원)	네 혹시 또 다른 문의사항 있으실까요?
(고객)	아니요 없어요.
(상담원)	하하하 네 이상 상담원 이선정이었습니다

개인정보 유형 분류표

연번	항목명	구분	설명	비고
1	고객상담 음성파일	고객-상담사 간 상담녹취음성	상담사와 고객 간 질의-응답으로 이루어진 음성 데이터	음성 (비정형데이터) ⇒STT 변환후 삭제
(변환: STT) ↓		※ 음성변환(STT, Speech To Text) 기술을 통해 텍스트로 변환하여 활용 각 음성파일에 고객 관련 메타데이터(상담목적, 성별, 연령, 주거지역) 포함		
2	고객상담 텍스트 파일	① 이름	고객 및 상담사의 이름	텍스트 (비정형데이터)
		② 생년월일	고객의 생년월일과 관련된 정보	
		③ 주소정보	고객의 상세 주소와 관련된 정보	
		④ 거주정보	상세 주소는 아니나 고객이 거주하는 공간과 관련된 정보	
		⑤ 일반 전화번호	고객의 집·회사 연락처	
		⑥ 휴대폰번호	고객의 휴대폰 연락처	
		⑦ 일반대화정보	고객과 상담사 간 질의-응답 내 대화의 맥락을 통해 고객, 상담사, 또는 다른 특정인을 식별할 위험이 있는 대화정보	
2-1	상담 목적		고객의 상담 목적 (1.단순문의, 2.A/S, 3.결제, 4.교환, 5.설치, 6.반품, 7.기타)	각 상담파일의 메타데이터 (정형데이터)
2-2	성별		고객의 성별 (1. 남성, 2. 여성)	
2-3	연령		고객의 연령 (1세단위, 17세~74세까지 분포)	
2-4	주거지역		고객의 주거지역 (1.서울, 2.부산, 3.대구, 4.인천, 5.광주, 6.대전, 7.울산, 8.세종, 9.경기, 10.강원, 11.충북, 12.충남, 13.전북, 14.전남, 15.경북, 16.경남, 17.제주)	

개인정보 유형별 예시

연번	항목명	구분	예시
2	고객상담 텍스트 파일	① 이름	“홍길동 고객님 본인 맞으실까요?” “네 감사합니다. 이상 상담원 이선정이었습니다”
		② 생년월일	“가입해주신 고객님 혹시 1987년 7월 22일생 맞으실까요?”
		③ 주소정보	“고객님 서울시 신뢰동 신뢰아파트 백일동 구백삼호로 배송해드리겠습니다”
		④ 거주정보	“찾기가 좀 어려우실텐데 신뢰역 3번출구 앞에서 오른쪽으로 쭉 올라오시면 돼요”
		⑤ 일반 전화번호	“저희 회사쪽으로 연락주시면 됩니다. 공이 일이삼사 오육칠팔입니다”
		⑥ 휴대폰번호	“기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 팔천오백이니까”
		⑦ 일반대화정보	“인터넷 가입하신지 12년 5개월되셨는데, 가족 세 분과 가족할인 요금제로 사용하고 계시거든요”, “다음 달 8일부터 신규 상품으로 변경하시면서 배우자분이랑 아드님도 가족결합할인으로 추가하시겠다는거 맞으시죠?”
2-1	상담 목적		‘5’ (설치 목적)
2-2	성별		‘2’ (여성)
2-3	연령		‘25’ (25세)
2-4	주거지역		‘1’ (서울)

활용데이터 요구 수준표

※ 가명정보 처리 목적 달성에 필요한 데이터 항목과 가명처리 요구수준을 검토

연번	항목명	구분	요구 수준
2	고객상담 텍스트 파일	① 이름	연구목적 달성에 필요 없으므로, 다른 이름으로 대체 가능
		② 생년월일	연구목적 달성에 필요 없으므로, 다른 값으로 대체 가능 (고객 연령대는 필요하나, 메타데이터를 통해 확인 가능)
		③ 주소정보	연구목적 달성에 필요 없으므로, 다른 값으로 대체 가능
		④ 거주정보	연구목적 달성에 필요 없으므로, 다른 값으로 대체 가능
		⑤ 일반 전화번호	연구목적 달성에 필요 없으므로, 다른 값으로 대체 가능
		⑥ 휴대폰번호	연구목적 달성에 필요 없으므로, 다른 값으로 대체 가능
		⑦ 일반대화정보	상담 목적 판단, 대화 맥락 인식 등에 필요 다른 항목들을 충분히 가명처리하였을 경우 그대로 활용하거나 훼손 최소화 필요
2-1	상담 목적		연구목적 달성을 위해 그대로 사용 필요
2-2	성별		연구목적 달성을 위해 그대로 사용 필요
2-3	연령		고객 연령대별 구분이 필요하므로, 10~20대, 30~40대, 50~60대, 70대 이상으로 범주화 가능
2-4	주거지역		연구목적 달성을 위해 그대로 사용 필요

식별 위험성 검토 결과보고서

※ 식별 위험성 검토 점검표를 기반으로 식별 위험성 검토 결과보고서 작성

가명정보 활용목적	▪ 콜센터 직원 교육용 가상상담 시나리오 생성 AI 모듈 개발	
가명처리 대상 데이터 항목	▪ 고객상담텍스트 파일 내 개인식별 위험성이 있는 정보 일체 (이름, 생년월일, 주소정보, 거주정보, 일반 전화번호, 휴대폰번호, 일반대화정보 등)	
데이터 위험성	식별성 유무	▪ 이름, 주소, 휴대폰 번호는 개인식별 위험성이 상당히 높음 ▪ 거주정보, 일반 전화번호는 다른 대화내용 및 메타데이터와 연계될 시 개인식별 위험성이 존재 ▪ 일반대화내용도 상황에 따라 다른 대화내용 및 메타데이터와 연계될 시 낮은 확률로 개인식별 위험성이 생길 수 있음
	특이정보 유무	▪ 메타데이터 내 '연령'은 아주 낮거나 높을 시 특이치 존재 가능 ▪ 일반대화내용에 포함된 상품명, 상품변경날짜 및 내용, 가족관계 등으로 인해 특이정보가 존재할 수 있음
	재식별시 영향도	▪ 콜센터 일반상담 업무 및 업종 특성을 고려하면 관련 대화내용은 재식별 시 영향도가 미미함
처리 환경 위험성	이용 및 제공 형태	▪ 동일 개인정보처리자, 동일 부서 내 활용
	처리 장소	▪ 빅데이터본부 내 Private 클라우드를 구축하여 망분리된 폐쇄형 개발환경에서 처리 ▪ 개인정보(가명정보)처리시스템에 대한 ISMS-P인증 취득
	다른 정보 결합 가능성	▪ 관리자 승인하에 데이터(프로그램 패키지, 라이브러리, 코드설명서 등) 반입이 가능함 ▪ 기존 가명정보와 결합가능성 있는 데이터는 반입이 제한됨
최종 검토의견	▪ 해당 가명정보는 기관 내 연구자에게 제공될 뿐만 아니라, 클라우드 기반의 폐쇄연구 분석환경(외부 인터넷 이용, 다른 데이터의 반입 및 결합, 데이터 외부반출 등이 제한)에서만 접속하여 분석 가능하므로 식별 가능성이 낮은 편이며 안전한 처리 환경을 고려할 때 다음과 같은 조치가 필요함 - (고객상담 텍스트 파일) 실제 고객과 상담을 진행하고 관련 원본 음성데이터를 보유하고 있는 한 동일 부서에서 활용되므로, 특정 고객이 식별되지 않도록 개인식별정보(이름, 주소, 휴대폰 번호)뿐만 아니라 일반 대화정보까지 특정 고객이 식별될 여지가 없도록 보다 철저한 가명처리 필요 (1) 이름 : 랜덤 이름 또는 교육용으로 설정한 가상의 이름으로 치환 (2) 생년월일 : 범주화된 연령값 중에 랜덤한 값으로 치환 (3) 주소정보 : 가상의 주소값으로 치환 (4) 거주정보 : 가상의 거주정보로 치환하고, 대체가 어려운 경우 삭제 (5) 일반 전화번호 : 가상의 번호로 치환 (6) 휴대폰 번호 : 가상의 번호로 치환 (7) 일반대화정보 : 그대로 사용하되, 전수검사하여 대화 맥락이나 특이정보 등을 고려하여 특정 개인이 식별될 가능성이 있다고 판단되는 문장은 삭제 - (메타데이터) 상담목적, 성별, 주거지역(시 단위)은 개인식별 위험이 거의 없고 연구목적 달성에 필요하므로 그대로 활용하되, 연령정보는 다른 정보와 결합하여 개인식별 위험이 있으므로 범주화하여 활용	

④ 가명처리 단계에서 위험성 검토 결과를 반영하여 가명처리 방법 및 수준을 적정하게 정의하였는지 확인

항목별 가명처리 계획

연번	항목명	세부 항목	처리방법	세부방법 및 처리수준
1	고객상담 음성파일	고객-상담사 간 상담녹취음성	☑삭제	▪ 음성변환(STT)을 통해 텍스트로 변환한 후 원본 음성파일은 삭제
	(변환: STT) ↓		※ 음성변환(STT, Speech To Text) 기술을 통해 텍스트로 변환	
2	고객상담 텍스트 파일	① 이름	☑ 치환	▪ 고객 이름은 [NAME_CUSTOMER] 토큰 정보로 변환한 뒤 '김행복'으로 변경 ▪ 상담원 이름은 [NAME_CENTER] 토큰 정보로 변환한 뒤, '김신뢰'로 변경
		② 생년월일	☑ 치환	▪ 생년월일은 [BIRTH] 토큰 정보로 변환한 뒤, 메타데이터 중 연령대(10~20대, 30~40대, 50~60대, 70대 이상) 범주 내에서 랜덤한 값으로 치환
		③ 주소정보	☑ 치환	▪ 주소는 [ADDR] 토큰 정보로 변한한 뒤, 메타데이터 내 주거지역(시단위) 내 가상의 주소로 대체
		④ 거주정보	☑ 삭제 (또는 대체)	▪ 검수자(시나리오 작가)가 직접 가상의 문장으로 각색하고, 대체가 어려운 경우 해당 질의-응답 문장 삭제
		⑤ 일반 전화번호	☑ 치환	▪ 일반전화번호는 [PHONE] 토큰 정보로 변환한 뒤, 실재하지 않는 가상 전화번호로 대체
		⑥ 휴대폰번호	치환	▪ 휴대폰번호는 [CELLPHONE] 토큰 정보로 변환한 뒤, 실재하지 않는 가상 휴대폰번호로 대체
		⑦ 일반대화정보	☑ 그대로 사용 (또는 삭제)	▪ 연구 목적에 필요하므로 그대로 사용하는 것을 원칙으로 하되, 검수자가 전수 검사하여 대화 맥락이나 특이정보 등을 고려하여 특정 개인이 식별될 가능성이 있다고 판단되는 문장은 삭제 처리
	2-1	상담 목적	☑ 그대로 사용	▪ 연구 목적 달성을 위해 별도 처리하지 않음
	2-2	성별	☑ 그대로 사용	▪ 연구 목적 달성을 위해 별도 처리하지 않음
	2-3	연령	☑ 범주화 (20세 단위)	▪ 고객 연령대 별 분석이 필요하므로, 20세 단위 범주화 처리하고 70세 이상은 상단코딩 적용 (10~20대, 30~40대, 50~60대, 70대 이상)
	2-4	주거지역	☑ 그대로 사용	▪ 연구 목적 달성을 위해 별도 처리하지 않음

⑤ 계획한 가명처리 방법 및 수준에 따라 실제 가명처리를 수행하였는지 확인

순번	항목명	세부 항목	가명처리 전	가명처리 후
2	고객상담 텍스트 파일	① 이름	▪ 홍길동 고객님 본인 맞으실까요? 〈고객이름 '김행복'으로 치환〉	▪ '[NAME_CUSTOMER]' 고객님 본인 맞으실까요? ⇒ '김행복' 고객님 본인 맞으실까요?
			▪ 네 감사합니다. 이상 상담원 이선정이었습니다 〈상담원 이름 '김신뢰'로 치환〉	▪ 네 감사합니다. 이상 상담원 '[NAME_CENTER]' 이었습니다" ⇒ 네 감사합니다. 이상 상담원 '김신뢰'였습니다
		② 생년월일	▪ 가입해주신 고객님 혹시 1987년 7월 22일생 맞으실까요? 〈메타데이터의 범주화된 연령(30~40대) 내 랜덤값으로 치환〉	▪ 가입해주신 고객님 혹시 '[BIRTH]'생 맞으실까요? ⇒ 가입해주신 고객님 혹시 '1991년 6월 8일'생 맞으실까요?
		③ 주소정보	▪ 고객님 서울시 신뢰동 신뢰아파트 백일동 구백삼호로 배송해드리겠습니다 〈메타데이터의 주거지역(서울) 내 가상주소로 치환〉	▪ 고객님 '[ADDR]'로 배송해드리겠습니다 ⇒ 고객님 '서울시 신뢰동 신뢰아파트 1동 1호'로 배송해드리겠습니다.
		④ 거주정보	▪ 찾기가 좀 어려우실텐데 신뢰역 3번출구 앞에서 오른쪽으로 쭉 올라오시면 돼요	〈삭제〉 ⇒ 문맥상 다른 문장으로 대체가 어려우므로, 질의-응답문 세트 전체를 삭제 * 시나리오 작가가 가상의 문장으로 각색하는 것도 가능
		⑤ 일반 전화번호	▪ 저희 회사쪽으로 연락주시면 됩니다. 공이 일이삼사 오육칠팔입니다 〈일반전화번호 가상 전화번호로 치환〉	▪ 저희 회사쪽으로 연락주시면 됩니다. '[PHONE]'입니다 ⇒ 저희 회사쪽으로 연락주시면 됩니다. '03-111-1111'입니다
		⑥ 휴대폰 번호	▪ 기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 팔천오백이니까 〈휴대폰번호 가상 전화번호로 치환〉	▪ 기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 '[CELLPHONE]'이니까 ⇒ 기사분 내방전에 전화드리고 갈 건데 휴대폰 번호 가운데가 '일일일일'이니까
		⑦ 일반 대화정보	▪ 인터넷 가입하신지 12년 5개월되셨는데, 가족 세 분과 가족할인 요금제로 사용하고 계시거든요 ▪ 다음 달 8일부터 신규 상품으로 변경하시면서 배우자분이랑 아드님도 가족결합할인으로 추가하시겠다는거 맞으시죠?	〈그대로 사용〉 ▪ 인터넷 가입하신지 12년 5개월되셨는데, 가족 세 분과 가족할인 요금제로 사용하고 계시거든요 ▪ 다음 달 8일부터 신규 상품으로 변경하시면서 배우자분이랑 아드님도 가족결합할인으로 추가하시겠다는거 맞으시죠? * 단, 전수검사하여 대화 맥락이나 특이정보를 고려하여 개인식별 위험이 있는 문장은 삭제·변경 필요

비정형데이터 가명처리 기술의 적절성·신뢰성 관련 근거(예시)

연번	항목명	세부 항목	처리 기술명	처리 기술의 적절성·신뢰성 관련 근거 또는 배경
2	고객상담 텍스트 파일	① 이름 ② 생년월일 ③ 주소정보 ⑤ 일반전화번호 ⑥ 휴대폰번호	치환	▪ '이름', '생년월일', '주소정보', '일반전화번호', '휴대폰번호'는 'A사의 텍스트 비식별화 솔루션'을 활용하여 정해진 텍스트나 정해진 범주 내 랜덤텍스트로 대체 - A사의 텍스트 비식별화 솔루션은 자연어 처리 기준으로 텍스트 데이터셋에서 93%의 검출 정확성으로 측정되었으며, 검출된 텍스트에 대한 처리 정확도는 99%로 측정 ※ 검출 정확도, 처리 정확도(오류율) 증빙자료 별첨 - 검출 정확도가 100%가 아니므로, 솔루션 적용 후 처리결과에 대해 추가적인 자체 전수검사를 수행
		④ 거주정보	삭제 (또는 대체)	▪ 정형화된 패턴이 없어 솔루션으로 검출이 어려우므로, 자체 육안 전수검사를 통해 검출하여 삭제 처리 ⇒ 대화 맥락상 중요하고, 개인식별성이 없도록 변경이 가능한 경우 시나리오 작가가 각색 수행)
		⑦ 일반대화정보	삭제	▪ 대화맥락이나 특이정보 등을 통해 개인식별 위험이 생기는 문장에 대해서는 정형화된 패턴이 없어 솔루션으로 검출이 어려우므로, 자체 육안 전수검사를 통해 검출하여 삭제 처리 ⇒ 대화 맥락상 중요하고, 개인식별성이 미미하다고 판단 되는 경우에는 삭제하지 않고 그대로 활용)

비정형데이터 가명처리 결과에 대한 자체 검증 결과서

<table>
<tr><td rowspan="8">검증 대상
데이터 명세</td><td colspan="2">개요</td></tr>
<tr><td colspan="2">㈜코리아인터넷은 고객상담 음성데이터를 음성변환(STT)을 통해 텍스트데이터로 변환한 뒤,
개인식별 위험성이 있는 대화문장에 대해 대체·삭제처리</td></tr>
<tr><td>데이터 유형</td><td>텍스트</td></tr>
<tr><td>원본 데이터 형식(파일 포맷)</td><td>WAV ⇒ JSON</td></tr>
<tr><td>처리 결과 데이터 형식(파일 포맷)</td><td>JSON</td></tr>
<tr><td>데이터 규모 및 크기</td><td>녹취파일(WAV) 10,000개(각 3분 분량, 총 500시간, 60GB)를 텍스트로 변환한 데이터셋</td></tr>
<tr><td>대상 데이터 항목명</td><td>고객상담 텍스트 파일 〈연번 2〉</td></tr>
<tr><td>가명처리 적용 기술</td><td>- A사 '텍스트 비식별화 솔루션'을 활용하여 이름, 생년월일, 주소, 연락처 정보를 검출하여 토큰 정보로 변환한 뒤 정해진 텍스트나 정해진 범주 내 랜덤값으로 치환
- 일반대화정보 내 개인식별 가능성이 있는 단어, 문장은 직접 검수하여 삭제 또는 각색</td></tr>
<tr><td>자체 검증 기간</td><td colspan="2">2023년 2월 10일 ~ 2023년 2월 20일</td></tr>
<tr><td>자체 검증 장소</td><td colspan="2">㈜코리아인터넷 녹취통합시스템 분석PC</td></tr>
<tr><td>자체 검증 과정
및 방법</td><td colspan="2">(검증방법)
- 개인정보보호팀 담당 팀장 주도하에 총 2명의 검수자가 Workday 10일간 전수조사
- 각색이 필요한 문장을 선별하여 시나리오 작가 2인이 각색 수행

(검증 시 확인사항)
① 치환되어야 할 항목들이 정해진 텍스트나 정해진 범주 내 랜덤값으로 치환되었는가?
⇒ 치환되지 않았을 시 수작업으로 변환 실시
② 치환대상 외에 일반대화정보 내 개인식별 가능성이 있는 단어, 문장이 포함되어 있는가?
⇒ 발견 시 팀장과 논의 하에 삭제할지 가상의 내용으로 각색할지 결정
⇒ 각색 대상으로 선별된 문장은 2명의 시나리오 작가가 각색 수행</td></tr>
<tr><td>자체 검증 결과</td><td colspan="2">- 수작업으로 이름, 생년월일, 주소, 연락처 정보 추가 변환 : 6,582개 파일
- 가상의 내용으로 일부 문장 각색 : 231개 파일
> 10,000개 파일 전체 전수검사 완료 및 이상 없음 확인</td></tr>
<tr><td rowspan="6">자체 검증자</td><td>소속 및 직위</td><td>성명</td><td>서명(인)</td></tr>
<tr><td>개인정보보호 팀장</td><td>가관리</td><td></td></tr>
<tr><td>음성 AI팀 인턴</td><td>나정보</td><td></td></tr>
<tr><td>음성 AI팀 인턴</td><td>다보호</td><td></td></tr>
<tr><td>음성 AI팀 시나리오 작가</td><td>라작가</td><td></td></tr>
<tr><td>음성 AI팀 시나리오 작가</td><td>마작가</td><td></td></tr>
</table>

부록 4

자주 묻는 질문(FAQ)

Q 가명정보 자체결합(셀프결합) 개념

A 가명정보의 자체결합(셀프결합)이란 결합전문기관이 자신이 보유한 가명정보와 다른 개인정보처리자가 보유한 가명정보를 스스로 결합하여 활용까지 수행하고자 하는 결합 형태를 의미함

'가명정보의 결합 및 반출 등에 관한 고시'('24.1.30)에 따라 결합전문기관으로 지정된 공공 및 민간기관은 보유한 데이터를 다른 개인정보처리자가 보유한 가명정보와 스스로 결합하여 제3자에게 제공하거나 직접 활용할 수 있도록 규정

Q 개인정보 가명처리 및 분석 지원의 범위(ex. 모의결합 등)

A 결합전문기관의 업무 지원 사항에서의 가명처리 지원은 결합 전 결합대상정보의 가명처리를 지원하는 것을 의미함

특히 모의결합이란 결합신청자가 결합에 따른 효용성 및 유용성을 본 결합 전 사전에 판단할 수 있도록 일부 데이터를 미리 결합/분석할 수 있는 절차로 결합전문기관의 업무 지원사항임

결합전문기관의 업무 지원 사항에서의 분석지원은 반출 전 결합정보의 분석과 반출 후 반출정보의 분석이 모두 가능

Q 가명정보는 개인정보인지 여부

A 가명정보는 성명, 연락처 등 식별정보를 삭제하거나 대체하는 등의 방법으로 식별가능성을 낮춘 개인정보임. 이에 가명정보도 다른 개인정보에 준하는 안전조치를 하여야 함

Q 개인정보 중 민감정보나 고유식별번호도 가명처리하여 활용할 수 있는지 여부

A 주민등록번호를 제외한 다른 고유식별번호와 민감정보는 가명처리하여 활용할 수 있으며 주민등록번호는 법률, 대통령령 등의 구체적 근거가 있는 경우에 한하여 활용 가능함(법률 등에 활용에 대한 명확한 근거가 있는 경우)

Q 가명정보의 유상 판매가 허용되는지 여부

A 가명정보를 과학적 연구 등 법에서 허용하는 목적 범위로 제공하면서 대가를 받는 것은 가능하나, 법에서 정한 목적 범위를 벗어나 판매할 목적으로 가명처리하는 것은 허용되지 않음

Q 통계법에 따라 수집되는 개인정보에 대해 가명정보 특례 규정을 적용하여 가명처리 및 결합 등이 가능한지 여부

A 공공기관이 처리하는 개인정보 중 「통계법」에 따라 수집되는 개인정보는 「개인정보 보호법」의 제3장부터 제7장까지 적용하지 않으나(「개인정보 보호법」 제58조 제1항 제1호),

개인정보위 결정례에 따라 「개인정보 보호법」 제58조 제1항 제1호는 「통계법」에 따른 승인통계, 지정통계 작성을 위해 수집되는 개인정보에 한하므로 승인·지정 통계 작성 외 기타 정책 활용 목적 등의 통계작성은 보호법 제3장부터 제7장까지가 적용되어 가명정보 특례 규정에 따라 가명처리 및 결합을 할 수 있음

참고로 개인정보 보호법은 일반법으로 제6조에 의거, 다른 법률에 개인정보 보호에 관한 특별한 규정이 있는 경우 그 규정을 우선하여 적용됨

관련 법령

개인정보 보호법 제28조의2(가명정보의 처리 등) ① 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있다.

제28조의3(가명정보의 결합 제한) ① 제28조의2에도 불구하고 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 서로 다른 개인정보처리자 간의 가명정보의 결합은 보호위원회 또는 관계 중앙행정기관의 장이 지정하는 전문기관이 수행한다.

② 결합을 수행한 기관 외부로 결합된 정보를 반출하려는 개인정보처리자는 가명정보 또는 제58조의2에 해당하는 정보로 처리한 뒤 전문기관의 장의 승인을 받아야 한다.

② 개인정보처리자는 제1항에 따라 가명정보를 제3자에게 제공하는 경우에는 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함해서는 아니 된다.

제58조(적용의 일부 제외) ① 다음 각 호의 어느 하나에 해당하는 개인정보에 관하여는 제3장부터 제7장까지를 적용하지 아니한다.

1. 공공기관이 처리하는 개인정보 중 「통계법」에 따라 수집되는 개인정보
2. 국가안전보장과 관련된 정보 분석을 목적으로 수집 또는 제공 요청되는 개인정보
3. 공중위생 등 공공의 안전과 안녕을 위하여 긴급히 필요한 경우로서 일시적으로 처리되는 개인정보

4. 언론, 종교단체, 정당이 각각 취재·보도, 선교, 선거 입후보자 추천 등 고유 목적을 달성하기 위하여 수집·이용하는 개인정보

② 제25조제1항 각 호에 따라 공개된 장소에 영상정보처리기기를 설치·운영하여 처리되는 개인정보에 대하여는 제15조, 제22조, 제27조제1항·제2항, 제34조 및 제37조를 적용하지 아니한다.

③ 개인정보처리자가 동창회, 동호회 등 친목 도모를 위한 단체를 운영하기 위하여 개인정보를 처리하는 경우에는 제15조, 제30조 및 제31조를 적용하지 아니한다.

④ 개인정보처리자는 제1항 각 호에 따라 개인정보를 처리하는 경우에도 그 목적을 위하여 필요한 범위에서 최소한의 기간에 최소한의 개인정보만을 처리하여야 하며, 개인정보의 안전한 관리를 위하여 필요한 기술적·관리적 및 물리적 보호조치, 개인정보의 처리에 관한 고충처리, 그 밖에 개인정보의 적절한 처리를 위하여 필요한 조치를 마련하여야 한다.

Q '과학적 연구'가 순수 '과학(Science)'과 관련한 협의의 개념인 것인지, 아니면 인간사회의 여러 현상을 연구하는 '사회과학(Social Science)' 등도 포함하는 광의의 개념인 것인지 여부

A 보호법 제2조제8호에서 규정한 '과학적 연구'는 순수과학 뿐만 아니라 사회과학 등 도 포함하는 광의의 개념임

'직장 내 성희롱·괴롭힘' 등과 같은 사회의 여러 현상·사례를 체계적으로 분석하여 어떠한 결과를 도출하는 연구를 '사회과학'에 관한 연구로 보아 과학적 연구로 볼 수 있음

과학적 방법을 적용하는 연구와 관련하여 사례분석, 설문조사 등을 통해 어떠한 결과를 도출하는 것도 '과학적 방법'에 해당

사회과학은 사회의 여러 현상·사례를 체계적으로 분석하여 결과를 도출하는 연구를 말하며 '직장 내 성희롱·괴롭힘' 이라는 사례를 분석·연구하고자 하는 경우도 사회과학으로 해석

Q A부서에서 수집한 개인정보를 B부서에서 익명처리하여 수집 목적 외로 처리할 수 있는지 여부

A B부서가 A부서에서 수집한 개인정보에 대한 접근 권한이 있으며, 처리한 정보가 익명정보에 해당한다면 개인정보 보호법 제58조의2에 따라 개인정보 보호법의 적용을 받지 않음.

개인정보 보호법의 적용이 제외되는 범위는 제58조의2에 따른 정보를 처리하는 경우 뿐 아니라, 개인정보를 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없도록 처리하는 경우까지 포함. 따라서 개인정보처리자가 이러한 정보 생성을 위해 개인정보를 처리하는 경우에는 정보주체의 동의를 받을 필요가 없음

다만 해당 정보가 시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보에 해당하는지는 제공받는 자의 처리목적, 이용 또는 제공환경, 정보의 특성 등을 종합적으로 고려하여 판단하여야 함

Q 가명정보에서 추가정보를 삭제할 경우 익명정보로 볼 수 있는지 여부

A 추가정보가 삭제된 가명정보가 그 자체만으로 개인을 알아볼 수 없는 정보라고 하더라도 익명정보인지 여부는 시간·비용·기술 등을 합리적으로 고려하여 별도로 판단하여야 함

Q 영상 수집 목적 외 영상정보를 이용하고자 하는 경우 영상정보를 마스킹 했을 때 가명정보 및 익명정보 여부

A 사람이나 자동차 영상 등 개인을 식별할 수 있는 사진에서 그 일부 또는 전체를 마스킹 하였어도 주변 상황 및 환경 등으로 개인을 식별할 수 있으므로 그 정보가 가명정보인지 익명정보인지 일률적으로 판단하기 어려움

가명정보인지 또는 익명정보인지 여부는 시간·비용·기술 등을 합리적으로 고려하여 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보인지 검토하여 판단 필요

Q A사에서 다수의 데이터분석회사에 동일한 가명정보를 전송하여 이를 분석한 결과 값을 회신 받을 수 있는지 여부

A A사가 다수의 데이터분석회사에 가명정보 처리를 위탁하는 경우라면, 개인정보의 업무위탁에 대하여 규정한 개인정보 보호법 제26조에 따른 의무사항을 준수하여야 함

Q 가명정보 결합 신청 중 신청 내용의 변경이 없는 기관의 데이터를 파기 하지 않고 변경된 데이터와의 결합에 다시 활용 가능한지 여부

A 기 접수된 결합신청과 동일한 건이므로 다른 결합신청자가 제출한 자료를 활용하여 결합을 수행하는 것이 가능함

Q 결합 신청 시 결합전문기관 선택의 기준이 있는지

A 결합전문기관의 선택에 별도의 제한은 없으나 분야별 가이드라인*을 참고하여 선택할 수 있음

* 보건의료 데이터 활용 가이드라인, 교육분야 가명·익명정보 처리 가이드라인, 공공분야 가명정보 제공 실무안내서 등

Q 내부 관리계획 수립 관련 통합 혹은 별개 마련 여부

A 보호법은 개인정보처리자가 내부 관리계획 등을 작성할 때 가명정보에 관한 사항을 기존 개인정보에 관한 내용과 묶어서 작성할지 별개로 조항을 추가하여 작성할지에 대해서는 별도로 규정하고 있지 않음

Q 개인정보 가명처리 과정에서 일시적으로 가명정보와 추가정보가 같은 서버에 존재하게 되는 것이 보호법 제28조의4 제1항 및 동 법 시행령 제29조의5 제1항 제2호·제3호가 정하는 안전조치의무 위반인지 여부

A 가명정보와 추가정보를 분리·보관하고 각 접근권한을 분리하도록 한 것은 가명처리 이후 재식별을 방지하기 위한 것으로, 가명처리 과정에서 가명정보와 추가정보가 일시적으로 동일 서버에 존재하는 것은 보호법 제28조의4 및 동법 시행령 제29조의5의 위반에 해당하지 않음

Q 가명정보를 제공하였을 시, 가명정보 활용 과정에서 생긴 문제에 대해 제공자도 법적책임이 있는지 여부

A 개인정보를 보호법에서 정한 처리 목적에 따라 가명처리하고 관련 안전조치 등 법률에서 정한 사항을 모두 준수하여 가명정보를 제공한 경우,
가명정보를 제공받은 자가 가명정보 이용 과정에서 의도치 않게 특정 개인을 알아볼 수 있는 정보가 생성되었다는 사실만으로는 가명정보를 제공한 자에 대해 개인정보 보호법상 행정처분을 하지 아니함 (단, 제공받은 자는 위 생성된 정보의 처리를 즉시 중지하고, 지체없이 회수·파기하여야 함)
또한, 가명정보를 제공받은 자가 안전조치 미이행 등으로 가명정보를 유출하였거나 고의로 재식별 행위를 하는 경우, 해당 행위자만 제재함

가명정보 처리 가이드라인

2024년 2월 발행
발행처 : 개인정보보호위원회
지원기관 : 한국인터넷진흥원

본 가이드라인은 2024년 1월 기준으로 작성되었습니다.
항상 최신의 가이드라인은 개인정보보호위원회 홈페이지(www.pipc.go.kr) 또는 가명정보 지원 플랫폼(dataprivacy.go.kr)에서 확인하시기 바랍니다.

가명정보 처리 가이드라인

개인정보보호위원회
Personal Information Protection Commission