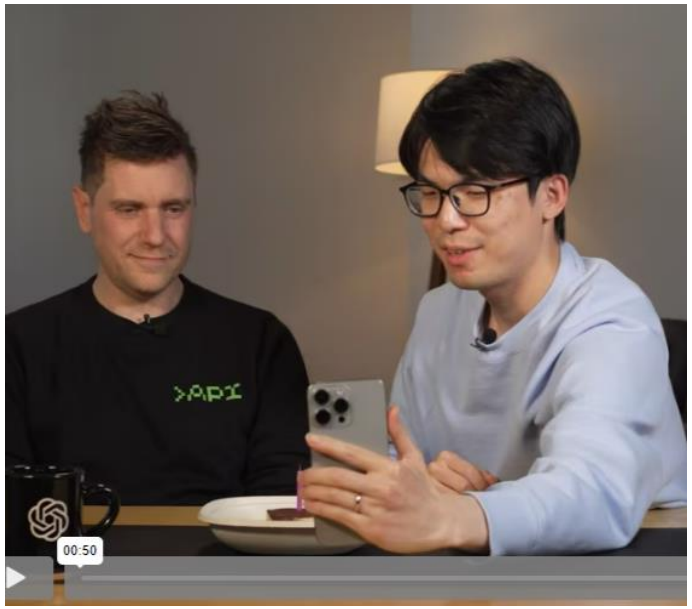
신뢰할 수 있는 데이터, 가치있는 데이터

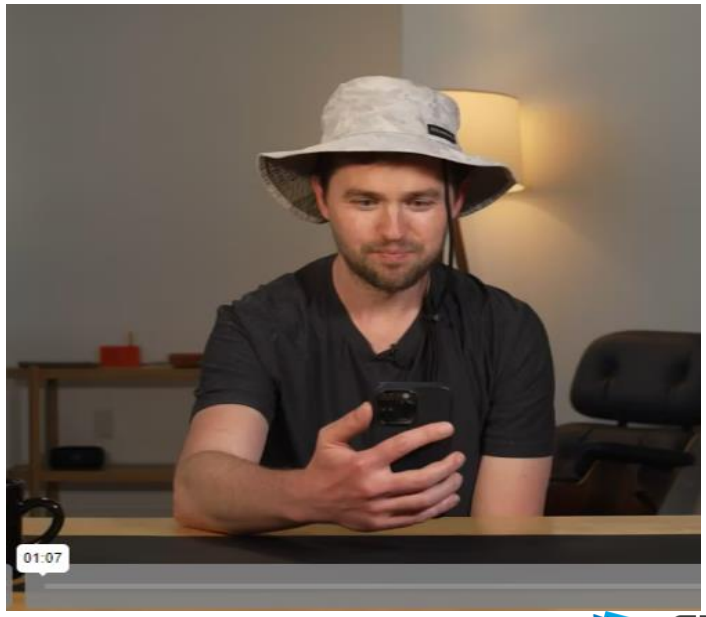
AI 시대

May 13, 2024

Hello GPT-4o

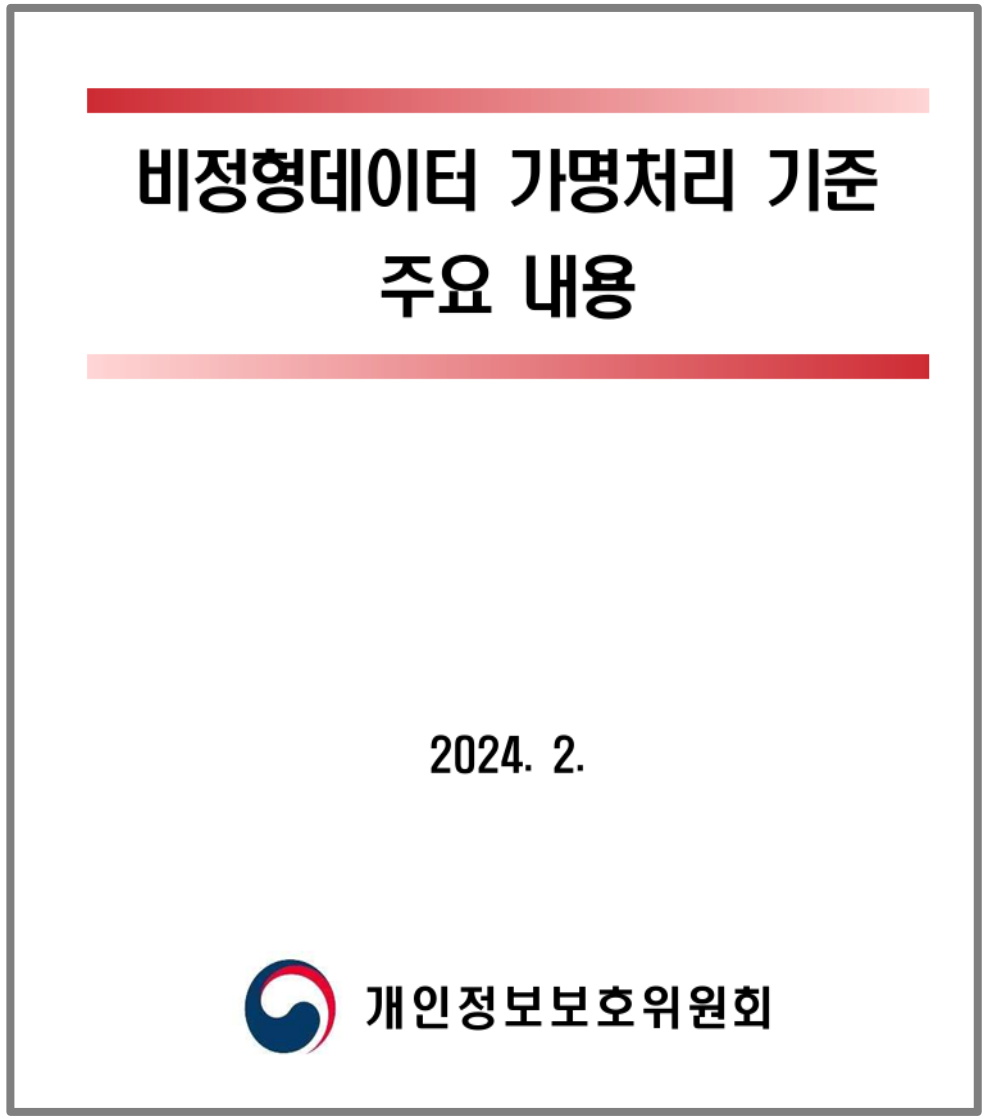
We're announcing GPT-4o, our new flagship model that can reason across audio, vision, and text in real time.





* 출처 : openai.com

비정형 데이터 가명처리 기준



AX를 위한 대응

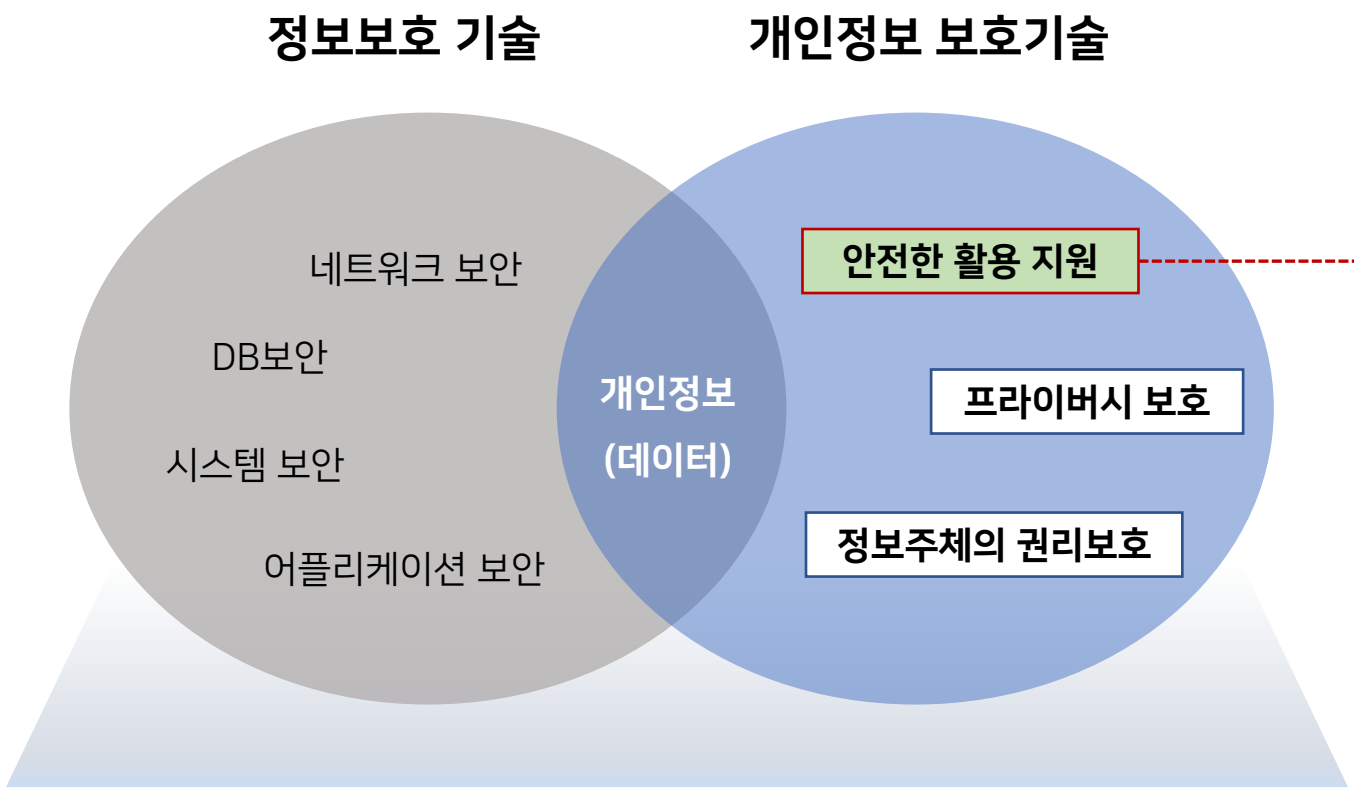
■ 데이터 3법 (2020.1)

- 데이터 활용을 위한 가명정보 개념 도입
- 가명처리(비식별) 조치 후 동의없이 사용가능
- 통계작성 및 과학적 연구, 공익적 기록보존 목적

■ 패러다임의 변화

- AI시대 대응을 위해
- 비정형 데이터를 안전하게 활용

개인정보 가명처리 기술의 필요성



보호대상	ICT자원 중심	개인정보 주체 중심
기술개발	시스템 보호	개인정보의 안전한 활용

가명처리(비식별) 솔루션

- 가명 · 익명처리
- 개인정보 분산저장
- 개인정보 재식별 차단

개인정보 보호 및 활용에 특화된
개인정보 보호강화 기술 필요

* 출처 : 2023 KAIST자료 재구성

SAT

비정형 데이터 가명처리 예시

1. 이미지 + 이미지 內 포함된 텍스트

항목	가명처리 前	가명처리 後
흉부CT 이미지		
	1. 흉부촬영 부분 2. 환자관련 정보 2.1 환자이름 2.2 생년월일 2.3 환자성별 2.4 환자번호	1. 그대로 유지 2. 환자관련 정보 2.1 가명처리로 대체 2.2 연도정보로 변경 2.3 성별값 “S”로 대체 2.4 일련번호로 대체

2. 영상 + 이미지

항목	가명처리 前	가명처리 後
사람 얼굴		
차량 번호판		
처리 기술	이미지 필터링(Blurring) 기술	

3. 음성 : STT 처리 된 텍스트(json)

항목	가명처리 前	가명처리 後
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_CUSTOMER]’ 이었습니다. ➔ 네 감사합니다. 이상 상담원 ‘김신뢰 ' 였습니다.
생년 월일	가입해주신 고객님 혹시 1987년 7월 22일생 맞으실까요? <메타데이터의 범주화 된 연령(30~40대) 내 랜덤값 치환>	가입해주신 고객님 혹시 ‘[BIRTH]’ 생 맞으실까요? ➔ 가입해주신 고객님 혹시 ‘1991년 6월 8일 ' 생 맞으실까요?
주소 정보	고객님 서울시 신뢰동 신뢰아파트 백일동 구백삼호로 배송해 드리겠습니다. <메타데이터의 주거지역(서울) 내 가상주소로 치환>	고객님 ‘[ADDR]’ 로 배송해 드리겠습니다. ➔ 고객님 ‘서울시 신뢰동 신뢰아파트 1동 1호 ' 로 배송해 드리겠습니다.

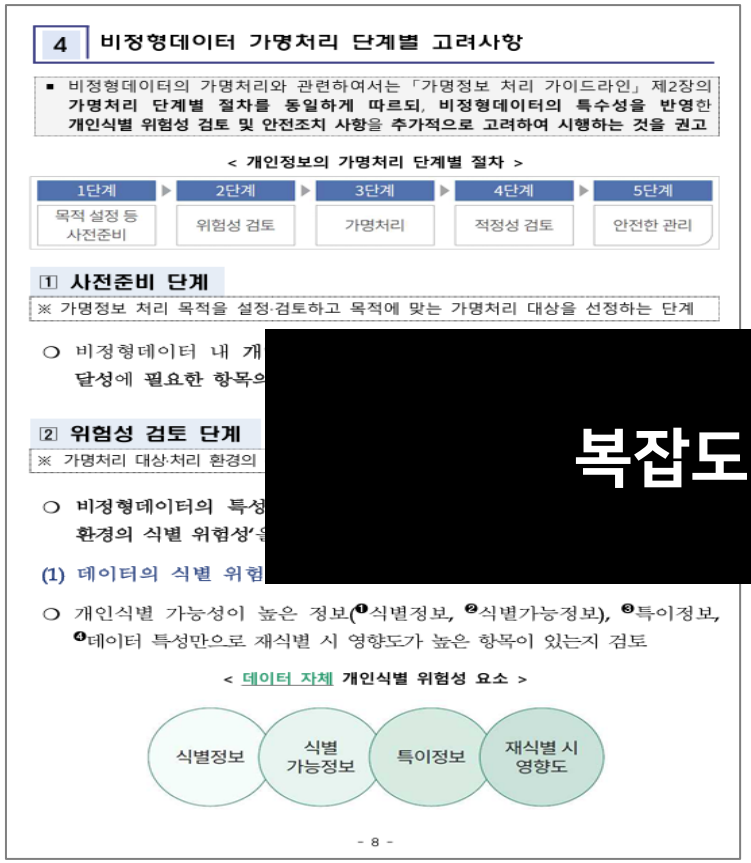
* 출처 : 2024 개인정보보호위원회 자료로 재구성

SAT

문서 파일에 대한 가명처리 방안은?

4. 문서를 포함함 텍스트

- 자연어처리(NLP) 기술을 이용하여 텍스트 가명처리
- 정규표현식 및 주석달기(Annotation) 기술 등을 이용하여 마스킹 및 대체 처리



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 개
달성에 필요한 항목의

2 위험성 검토 단계

※ 가명처리 대상·처리 환경의

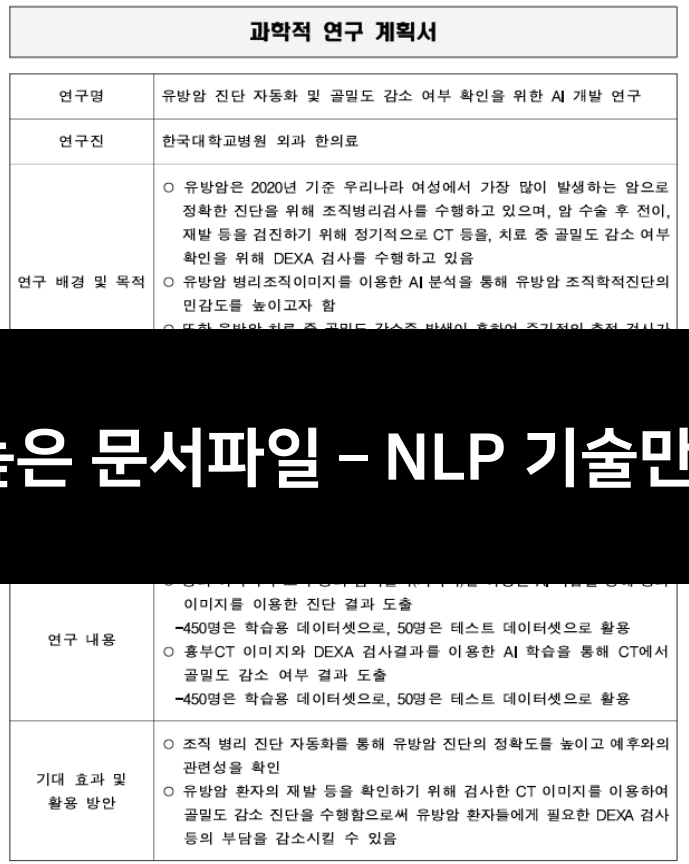
○ 비정형데이터의 특성
환경의 식별 위험성'을

(1) 데이터의 식별 위험

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

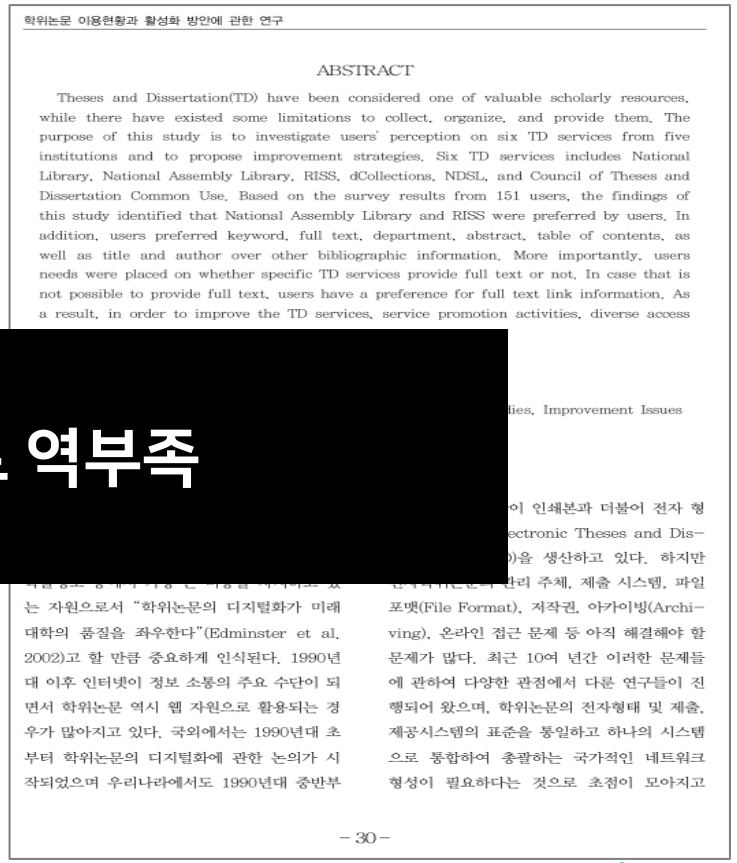
< 데이터 자체 개인식별 위험성 요소 >

- 8 -



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
연구 내용	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
기대 효과 및 활용 방안	○ 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



학위논문 이용현황과 활성화 방안에 관한 연구

ABSTRACT

Theses and Dissertation(TD) have been considered one of valuable scholarly resources, while there have existed some limitations to collect, organize, and provide them. The purpose of this study is to investigate users' perception on six TD services from five institutions and to propose improvement strategies. Six TD services includes National Library, National Assembly Library, RISS, dCollections, NDSL, and Council of Theses and Dissertation Common Use. Based on the survey results from 151 users, the findings of this study identified that National Assembly Library and RISS were preferred by users. In addition, users preferred keyword, full text, department, abstract, table of contents, as well as title and author over other bibliographic information. More importantly, users needs were placed on whether specific TD services provide full text or not. In case that is not possible to provide full text, users have a preference for full text link information. As a result, in order to improve the TD services, service promotion activities, diverse access

lies, Improvement Issues

이 인쇄본과 더불어 전자 형
ectronic Theses and Dis-
)을 생산하고 있다. 하지만
관리 주체, 제출 시스템, 파일
포맷(File Format), 저작권, 아카이빙(Archi-
ving), 온라인 접근 문제 등 아직 해결해야 할
문제가 많다. 최근 10여 년간 이러한 문제들
에 관하여 다양한 관점에서 다룬 연구들이 진
행되어 왔으며, 학위논문의 전자형태 및 제출,
제공시스템의 표준을 통일하고 하나의 시스템
으로 통합하여 총괄하는 국가적인 네트워크
형성이 필요하다는 것으로 초점이 모아지고

는 자원으로서 "학위논문의 디지털화가 미래 대학의 품질을 좌우한다"(Edminster et al. 2002)고 할 만큼 중요하게 인식된다. 1990년대 이후 인터넷이 정보 소통의 주요 수단이 되면서 학위논문 역시 웹 자원으로 활용되는 경우가 많아지고 있다. 국외에서는 1990년대 초부터 학위논문의 디지털화에 관한 논의가 시작되었으며 우리나라에서도 1990년대 중반부

– 30 –

복잡도가 높은 문서파일 – NLP 기술만으로 역부족

SAT

문서에 있는 ‘개인정보 가명처리’

어떻게 해야 하는가?

문서에서의 가명처리 방안

1 문서의 '구조' 이해

■ 다양한 문서포맷

- 한글, 오피스, PDF, 기타
- 다양한 버전

■ 수많은 문서 유형

- 문서 작성 Style : Lay-Out
- 문서 유형 : 보고서, 증명서 등

■ Object

- Font, Line, Table, Image
- 위치(좌표), Size, Color 등

2 문장의 '내용' 이해

■ 구문 분석

- 문장을 이루고 있는 구성 성분으로 분해 ➔ 관계 분석
  ➔ 문장의 구조 결정
- 문장화 처리
- 장 / 절 구분

■ 데이터 처리

- 형태소 분석
- 사전(Dictionary) 분석
- 기타

3 '결과물' 활용 방안

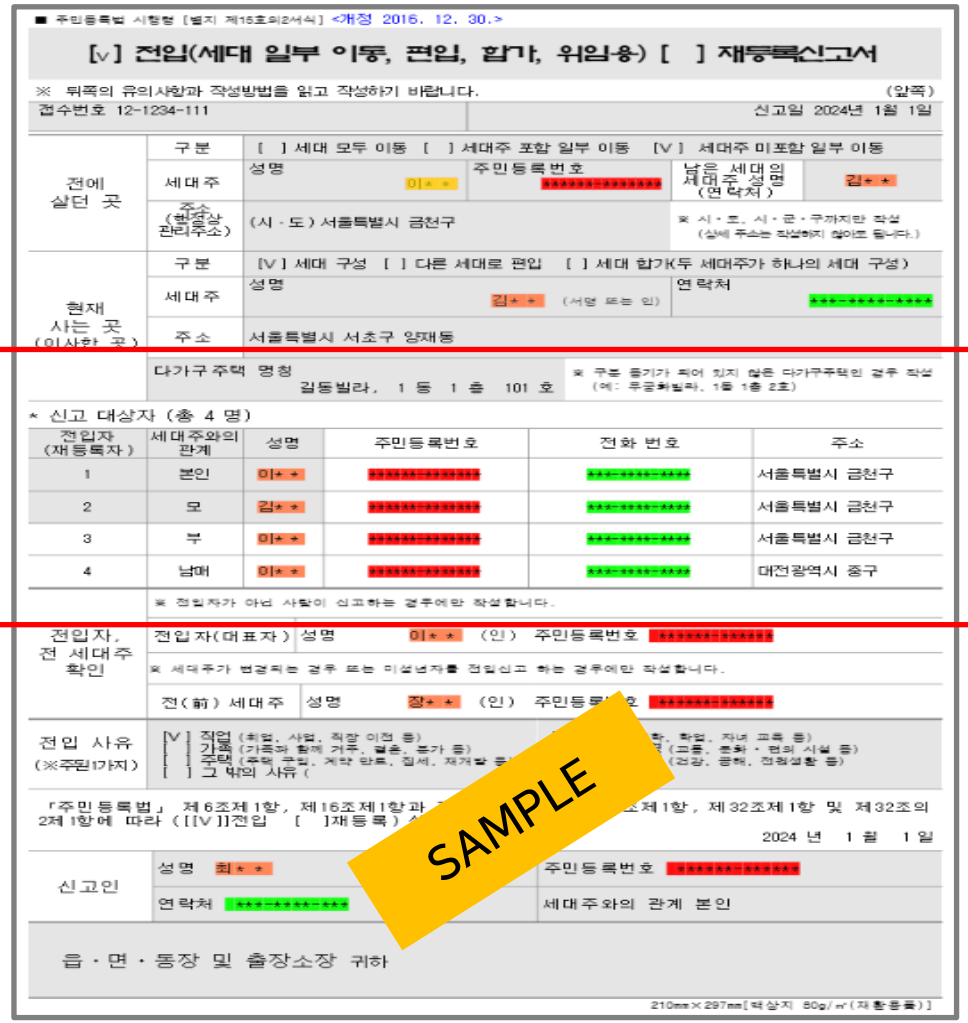
가명처리 된 결과를 어떻게 제공(output)할 것인가?

'반출할 수 있는 파일'

SAT

문서에서의 가명처리 예시

■ 대상 : 한글, MS오피스,PDF + 이미지



■ 주민등록법 시행령 [별지 제15호의2서식] <개정 2016. 12. 30.>

[V] 전입(세대 일부 이동, 편입, 합가, 위임용) [ ] 재등록신고서

※ 뒤쪽의 유의사항과 작성방법을 읽고 작성하기 바랍니다. (앞쪽)

접수번호 12-1234-111 신고일 2024년 1월 1일

	구분	[ ] 세대 모두 이동 [ ] 세대주 포함 일부 이동 [V] 세대주 미포함 일부 이동			
전에 살던 곳	세대주	성명 이* *	주민등록번호 **********	남은 세대의 세대주 성명 (연락처)	김* *
	주소 (현상살 관리주소)	(시·도) 서울특별시 금천구		※ 시·도, 시·군·구까지만 작성 (상세 주소는 작성하지 않아도 됩니다.)	
현재 사는 곳 (이사할 곳)	구분	[V] 세대 구성 [ ] 다른 세대로 편입 [ ] 세대 합가(두 세대주가 하나의 세대 구성)			
	세대주	성명 김* * (서명 또는 인)	연락처 **********		
	주소	서울특별시 서초구 양재동			
	다가구주택 명칭	길동빌라, 1 동 1 층 101 호		※ 구분 등기가 되어 있지 않은 다가구주택인 경우 작성 (예: 무궁화빌라, 1동 1층 2호)	

★ 신고 대상자 (총 4 명)

전입자 (재등록자)	세대주와의 관계	성명	주민등록번호	전화 번호	주소
1	본인	이* *	**********	**********	서울특별시 금천구
2	모	김* *	**********	**********	서울특별시 금천구
3	부	이* *	**********	**********	서울특별시 금천구
4	남매	이* *	**********	**********	대전광역시 중구

※ 전입자가 아닌 사람이 신고하는 경우에만 작성합니다.

전입자, 전 세대주 확인	전입자(대표자)	성명	이* * (인)	주민등록번호	**********
	※ 세대주가 변경되는 경우 또는 미성년자를 전입신고 하는 경우에만 작성합니다.				
	전(前) 세대주	성명	장* * (인)	주민등록번호	**********
전입 사유 (※주된1가지)	[V] 직업 (취업, 사업, 직장 이전 등) [ ] 가족 (가족과 함께 거주, 결혼, 분가 등) [ ] 주택 (주택 구입, 계약 만료, 집세, 재개발 등) [ ] 그 밖의 사유 (			확, 학업, 자녀 교육 등) (고용, 문화·법의 시설 등) (건강, 금융, 전원생활 등)	

「주민등록법」 제6조제1항, 제16조제1항과 [illegible] 조제1항, 제32조제1항 및 제32조의 2제1항에 따라 ([V]전입 [ ]재등록) [illegible]

2024 년 1 월 1 일

신고인	성명 최* *		주민등록번호 **********
	연락처 **********		세대주와의 관계 본인

읍·면·동장 및 출장소장 귀하

210mm×297mm[백상지 80g/㎡(재활용품)]

SAMPLE

1. META DATA 추출 및 구조화

{"page":{"pagenum":1,"sz":{"cx":59528,"cy":84188},"body":{"l":5669,"t":5669,"r":53859,"b":81354},"ct":[{"ct":[]},{"ot":"shape","pos":{"x":4252,"y":4252},"sz":{"cx":51024,"cy":78519},"type":"pageborder","ct":[{"sz":{"cx":48198,"cy":75681},"tr":[{"sz":{"cx":48198,"cy":1182},"tc":[{"cols":11,"pos":{"x":5952,"y":5952},"pos":{"x":6235,"y":6858},"rect":{"l":6235,"t":6171,"r":6979,"b":6971},"tid":0,"face":"돋움체","sz":800,"
{"l":6979,"t":6171,"r":7351,"b":6971},"tid":1,"adv":"372","v":" "},{"ot":"t","pos":{"x":7351,"y":6858},"
"v":"주민등록법"},{"ot":"t","pos":{"x":11071,"y":6858},"rect":{"l":11071,"t":6171,"r":11443,"b":6971},"ti
"t":6171,"r":13675,"b":6971},"tid":4,"adv":"744,744,744","v":"시행령"},{"ot":"t","pos":{"x":13675,"y":685
["},{"ot":"t","pos":{"x":14419,"y":6858},"rect":{"l":14419,"t":6171,"r":15907,"b":6971},"tid":6,"adv":"7
"t":6171,"r":16279,"b":6971},"tid":7,"adv":"372","v":" "},{"ot":"t","pos":{"x":16279,"y":6858},"rect":{"
"pos":{"x":17023,"y":6858},"rect":{"l":17023,"t":6171,"r":17767,"b":6971},"tid":9,"adv":"372,372","v":"1
"b":6971},"tid":10,"adv":"744,744","v":"호의"},{"ot":"t","pos":{"x":19255,"y":6858},"rect":{"l":19255,"t
{"x":19627,"y":6858},"rect":{"l":19627,"t":6171,"r":21115,"b":6971},"tid":12,"adv":"744,744","v":"서식"}
"b":6971},"tid":13,"adv":"372,372","v":"] "},{"ot":"t","pos":{"x":21859,"y":6858},"rect":{"l":21859,"t":
{"ot":"t","pos":{"x":22309,"y":6858},"rect":{"l":22309,"t":6085,"r":24109,"b":6985},"tid":15,"adv":"900,
"r":30859,"b":6985},"tid":16,"adv":"450,450,450,450,450,450,450,450,450,450,450,450,450,450,450","v":" 2
"cy":1182},"ct":[{"ot":"p","align":"right","ct":[]}]}]},{"sz":{"cx":48198,"cy":3966},"tc":{"cols":15,"po
7134 L 53952 7134 L 53952 11063 L 5952 11063 Z","fillPr":{"type":"clr","clrs":[{"clr":"ffe5e5e5", "pos":
"rect":{"l":9076,"t":8395,"r":9546,"b":9995},"tid":17,"face":"휴먼옛체","sz":1600,"clr":"ff000000","adv":
"r":10363,"b":9825},"tid":18,"face":"HY신명조","sz":1048,"adv":"817","v":"V"},{"ot":"t","pos":{"x":10363,
"face":"휴먼옛체","sz":1600,"adv":"470,699","v":"] "},{"ot":"t","pos":{"x":11532,"y":9677},"rect":{"l":11
"pos":{"x":14330,"y":9677},"rect":{"l":14330,"t":8395,"r":14844,"b":9995},"tid":21,"adv":"514","v":"(
"b":9995},"tid":22,"adv":"1399,1399","v":"세대"},{"ot":"t","pos":{"x":17642,"y":9677},"rect":{"l":17642,
{"x":18341,"y":9677},"rect":{"l":18341,"t":8395,"r":21139,"b":9995},"tid":24,"adv":"1399,1399","v":"일부
"b":9995},"tid":25,"adv":"699","v":" "},{"ot":"t","pos":{"x":21838,"y":9677},"rect":{"l":21838,"t":8395,

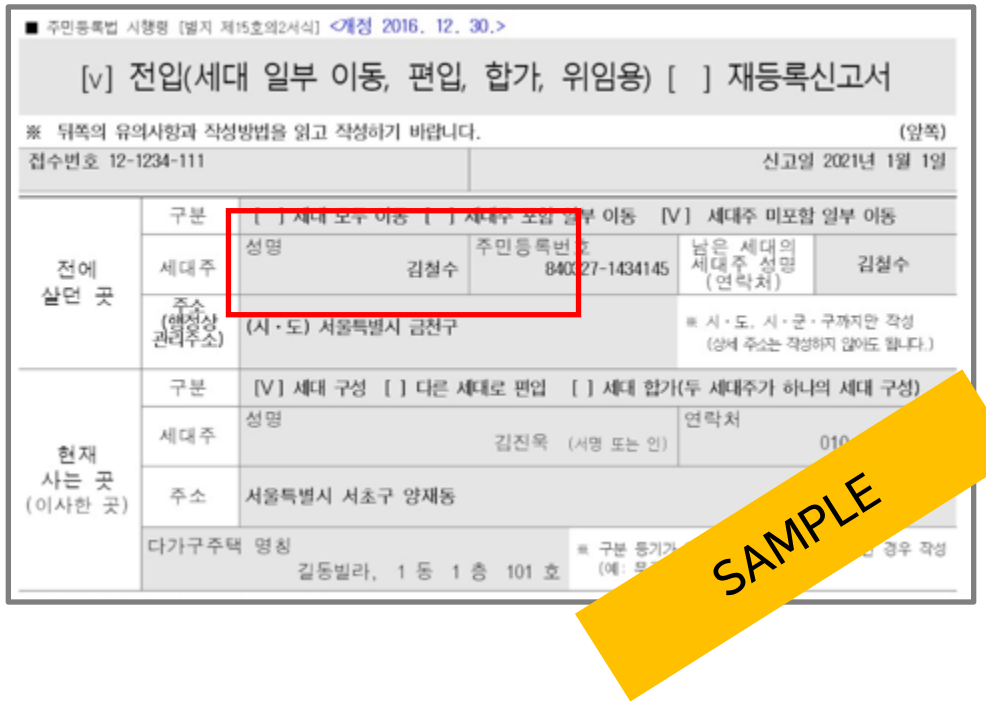
Meta Data 추출 : 문서의 내용을 구성하는 주요정보

Text	Object	Lay-Out
- txt & 좌표 - font & size & color 등	- Line & 좌표 - Table & 속성 - Image 등	- 문서 유형 - 문서 형식

SAT

문서에서의 가명처리 예시

■ 대상 : 한글, MS오피스,PDF + 이미지



■ 주민등록법 시행령 [별지 제15호의2서식] <개정 2016. 12. 30.>

[v] 전입(세대 일부 이동, 편입, 합가, 위임용) [ ] 재등록신고서

※ 뒤쪽의 유의사항과 작성방법을 읽고 작성하기 바랍니다. (앞쪽)

접수번호 12-1234-111 신고일 2021년 1월 1일

전에 살던 곳	구분	[ ] 세대 모두 이동 [ ] 세대주 포함 일부 이동 [V] 세대주 미포함 일부 이동			
	세대주	성명 김철수	주민등록번호 840327-1434145	남은 세대의 세대주 성명 (연락처)	김철수
	주소 (행정상 통상적 관리주소)	(시·도) 서울특별시 금천구		※ 시·도, 시·군·구까지만 작성 (상세 주소는 작성하지 않아도 됩니다.)	
현재 사는 곳 (이사한 곳)	구분	[V] 세대 구성 [ ] 다른 세대로 편입 [ ] 세대 합가(두 세대주가 하나의 세대 구성)			
	세대주	성명 김진욱 (서명 또는 인)		연락처 010	
	주소	서울특별시 서초구 양재동			
	다가구주택 명칭	길동빌라, 1 동 1 층 101 호		※ 구분 등기가 (예: 무	경우 작성

SAMPLE

2. 개인정보 검출


```
"ot" : "p",
"align" : "right",
"ct" : [
    {
        "ot": "t",
        "pos": {
            "x": 25796,
            "y": 19265
        },
        "rect": {
            "l": 25796,
            "t": 18492,
            "r": 28496,
            "b": 19392
        },
        "tid": 99,
        "sz": 900,
        "adv": "900,900,900",
        "v": "김철수"
    }
]
```

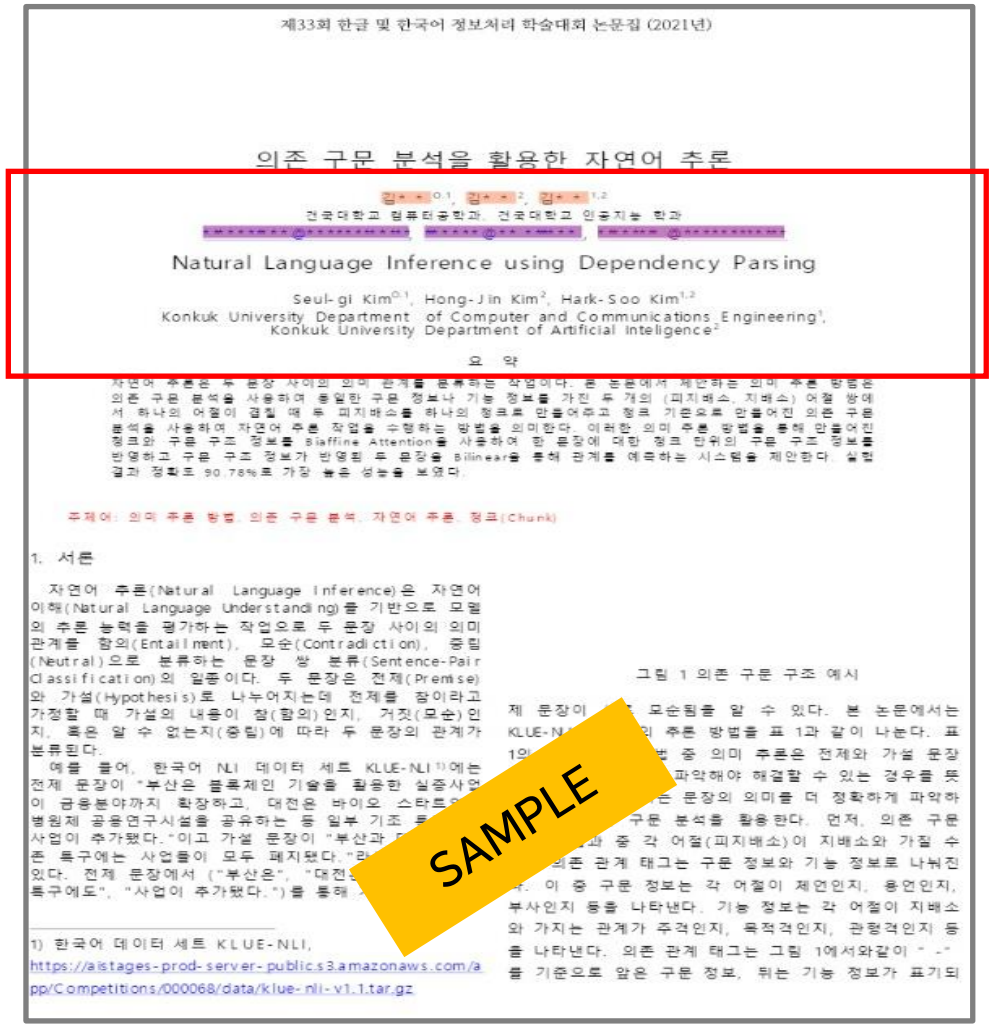

3. 개인정보 가명처리


```
"ot" : "p",
"align" : "right",
"ct" : [
    {
        "ot": "t",
        "pos": {…
        },
        "rect": {…
        },
        "tid": 99,
        "sz": 900,
        "adv": "900,900,900",
        "v": "김철수",
        "pim": {
            "use": "y",
            "config_use": "y",
            "bgcolor": "#FF7F50",
            "type": "highlight_per",
            "vid": "김**",
            "deidtype": ["휴리스틱1"],
            "findlist": ["김철수"],
            "word_idxs": [99],
            "indexOf": 0,
            "convertedword": ""
        }
    }
]
```


SAT

문서에서의 가명처리 예시

■ 논문 : 다단 구조 형식



제33회 한글 및 한국어 정보처리 학술대회 논문집 (2021년)

의존 구문 분석을 활용한 자연어 추론

김*기[O.1], 김*진[2], 김*수[1.2]
건국대학교 컴퓨터공학과, 건국대학교 인공지능 학과
[redacted]

Natural Language Inference using Dependency Parsing

Seul-gi Kim[O.1], Hong-Jin Kim[2], Hark-Soo Kim[1.2]
Konkuk University Department of Computer and Communications Engineering[1],
Konkuk University Department of Artificial Inteligence[2]

요 약

[redacted]

주제어: 의미 추론 방법, 의존 구문 분석, 자연어 추론, 청크(Chunk)

1. 서론

자연어 추론(Natural Language Inference)은 자연어 이해(Natural Language Understanding)를 기반으로 모델의 추론 능력을 평가하는 작업으로 두 문장 사이의 의미 관계를 함의(Entailment), 모순(Contradiction), 중립(Neutral)으로 분류하는 문장 쌍 분류(Sentence-Pair Classification)의 일종이다. 두 문장은 전제(Premise)와 가설(Hypothesis)로 나누어지는데 전제를 참이라고 가정할 때 가설의 내용이 참(함의)인지, 거짓(모순)인지, 혹은 알 수 없는지(중립)에 따라 두 문장의 관계가 분류된다.

예를 들어, 한국어 NLI 데이터 세트 KLUE-NLI[1]에는 전제 문장이 "부산은 블록체인 기술을 활용한 실증사업이 금융분야까지 확장하고, 대전은 바이오 스타트[redacted] 병원체 공용연구시설을 공유하는 등 일부 기조 [redacted] 사업이 추가됐다."이고 가설 문장이 "부산과 [redacted] 있다. 전제 문장에서 ("부산은", "대전은 [redacted] 특구에도", "사업이 추가됐다.")를 통해 [redacted]

그림 1 의존 구문 구조 예시

제 문장이 [redacted] 모순됨을 알 수 있다. 본 논문에서는 KLUE-N[redacted]의 추론 방법을 표 1과 같이 나눈다. 표 1의 [redacted] 중 의미 추론은 전제와 가설 문장 [redacted] 파악해야 해결할 수 있는 경우를 뜻[redacted] 는 문장의 의미를 더 정확하게 파악하[redacted] 구문 분석을 활용한다. 먼저, 의존 구문 [redacted] 과 중 각 어절(피지배소)이 지배소와 가질 수 [redacted] 의존 관계 태그는 구문 정보와 기능 정보로 나눠진 [redacted]. 이 중 구문 정보는 각 어절이 체언인지, 용언인지, 부사인지 등을 나타낸다. 기능 정보는 각 어절이 지배소와 가지는 관계가 주격인지, 목적격인지, 관형격인지 등을 나타낸다. 의존 관계 태그는 그림 1에서와같이 " -" 를 기준으로 앞은 구문 정보, 뒤는 기능 정보가 표기되

SAMPLE

1) 한국어 데이터 세트 KLUE-NLI, https://aistages-prod-server-public.s3.amazonaws.com/app/Competitions/000068/data/klue-nli-v1.1.tar.gz

■ 개인정보 검출

검출정보

- 김슬기
- 김홍진
- 김학수
- cludyju@konkuk.ac.kr
- jin34@gmail.com
- mlpdrkm@konkuk.ac.kr

■ 개인정보 가명처리

가명처리 결과

- 김**
- 김**
- 김**
- ******@******.**.**
- *****@*****.***
- *******@******.**.**

SAT

문서에서의 가명처리 예시

■ 이력서 : 표형식 + 이미지 객체

이 력 서

희 망 부 서	개발부
희 망 연 봉	6,500만원

성명	한 글	김**	주민등록번호	******-*******
	한 자	金一中	생 년 월 일	1990년 1월 1일
연락처	주 소	서울특별시 구로구 구로동 1 (우편번호:11111)		
	전화번호	**-***-****	핸드폰	***-****-****

학력		
고등학교	2009 년 2 월	길동 고등학교 졸업
대학	2009 년 3 월	길동 대학교 컴퓨터 공학 학과 입학
	2014 년 2 월	길동 대학교 컴퓨터 공학 학과 졸업
대학원	2014 년 3 월	길동 대학원 컴퓨터 공학 전공 입학
	2016 년 2 월	길동 대학원 컴퓨터 공학 전공 수료
기타	2012 년 4 월	컴퓨터 언어 특수 강의 2달 과정 수료
	2017 년 7 월	데이터 가공 특별 수업 1달 과정 수료

성적	4.4	자격증	① 정보처리기능사	외국어능력점수	영 어	TOEIC : 950
			② 코딩지도사		제2외국어	① HSK 4급
			③ 컴퓨터 활용능력 1급			② JLT 5급

경력	근무기간	근무처	직위	주요업무	급여	사직이유
	2015 년 2 월 ~ 2017 년 6 월	개발팀	사원	서버관리	2,800만원	개인사유
	2017 년 9 월 ~ 2018 년 5 월	유지보수	사원	개발	2,900만원	개인사유
	2018 년 5 월 ~ 2020 년 12 월	개인정보	대리	개발	3,400만원	개인사유

과외활동	활동 기간	단체명
	2010 년 4 월 ~ 2011 년 9 월	봉사활동
	2015 년 4 월 ~ 진행중	재능기부

신체	신장	시력	혈액	기타	취미	특기	종교
	175cm	좌:1.0 우:1.0			음악감상	기타연주	무교

SAMPLE

위의 [redacted]과 틀림이 없습니다.

2021 년 1 월 1 일

지 원 자 : 김** (인)

■ 개인정보 검출

검출정보

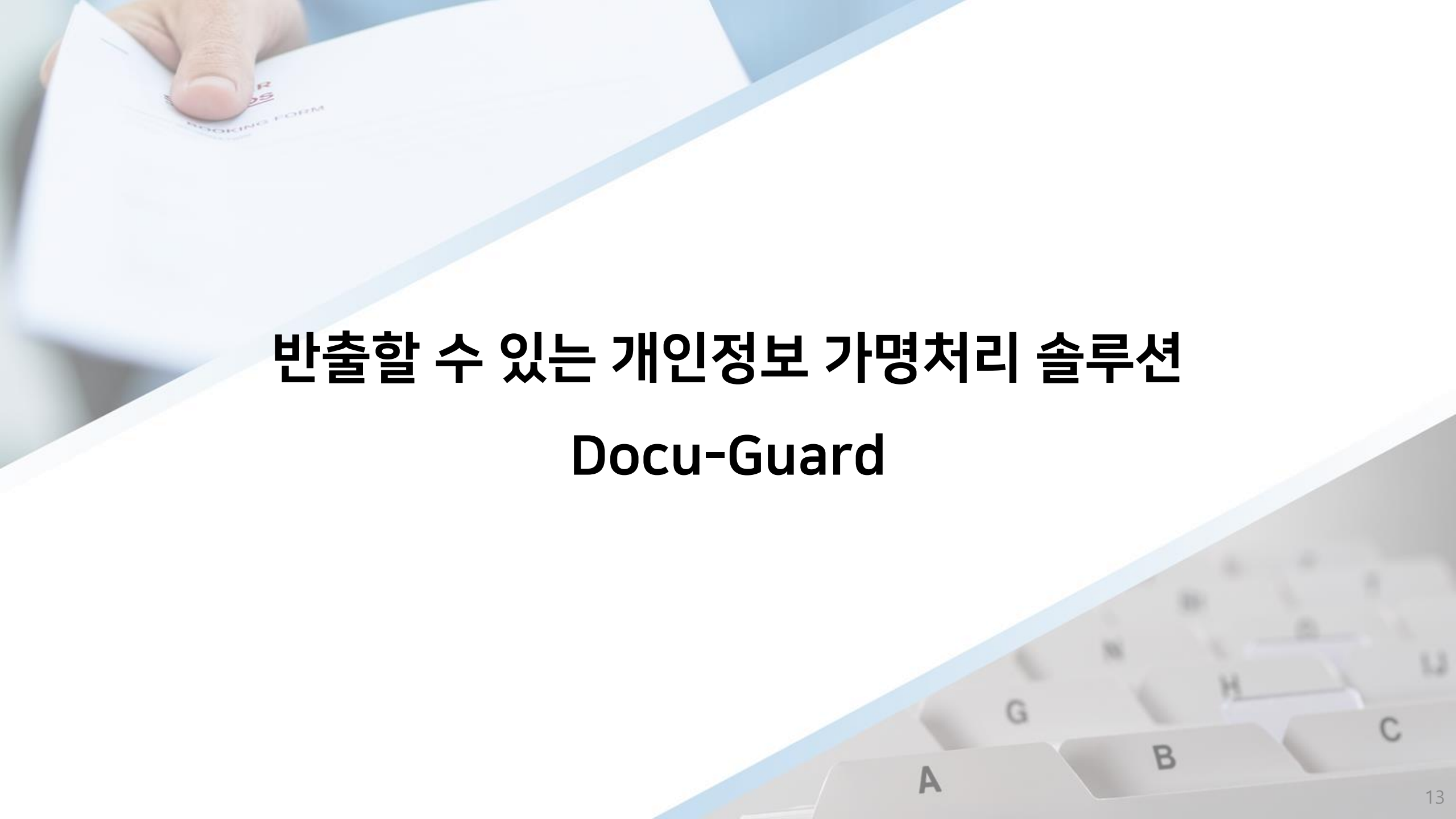
- 김일중
- 910101-1222222
- 02-111-2222
- 010-1111-2222
- 김일중

■ 개인정보 가명처리

가명처리 결과

- 김**
- ******-*******
- **-***-****
- ***-****-****
- 김**

SAT

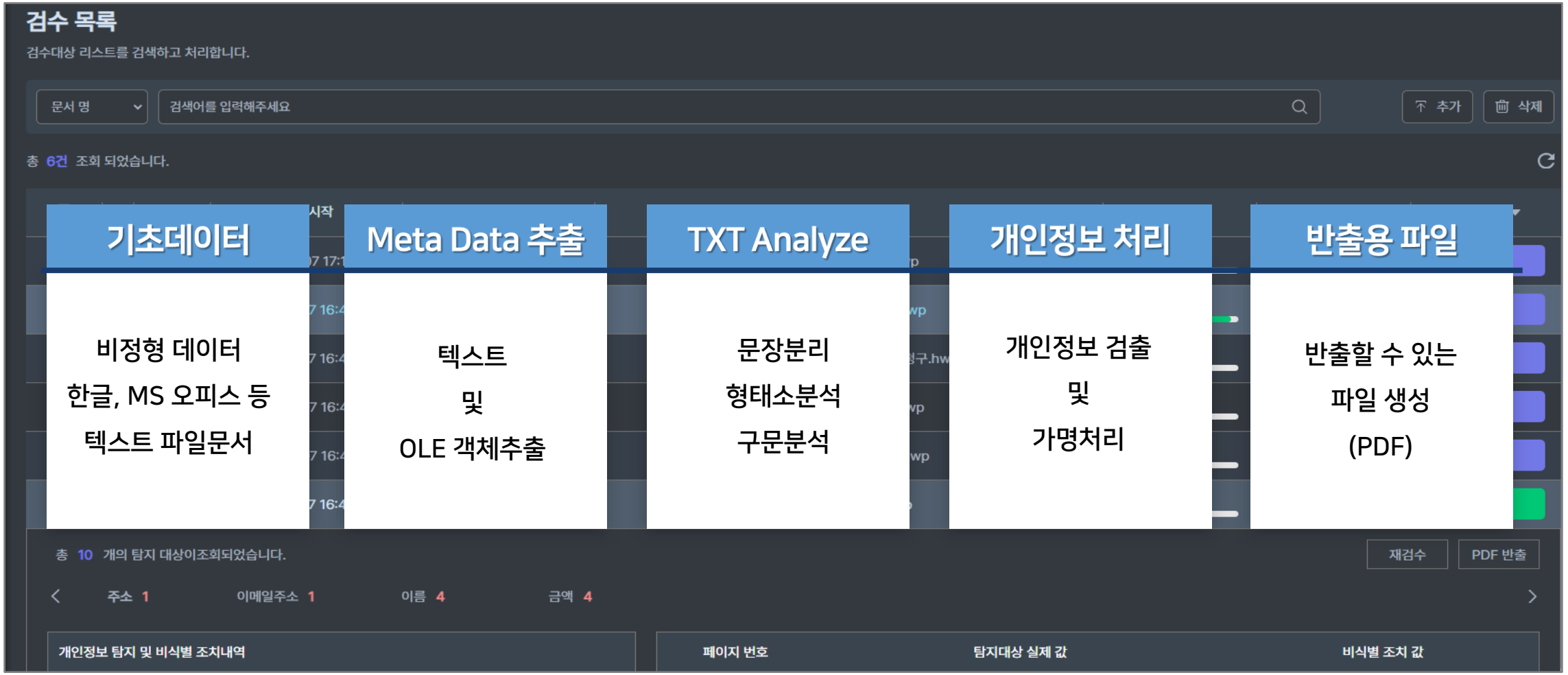


반출할 수 있는 개인정보 가명처리 솔루션

Docu-Guard

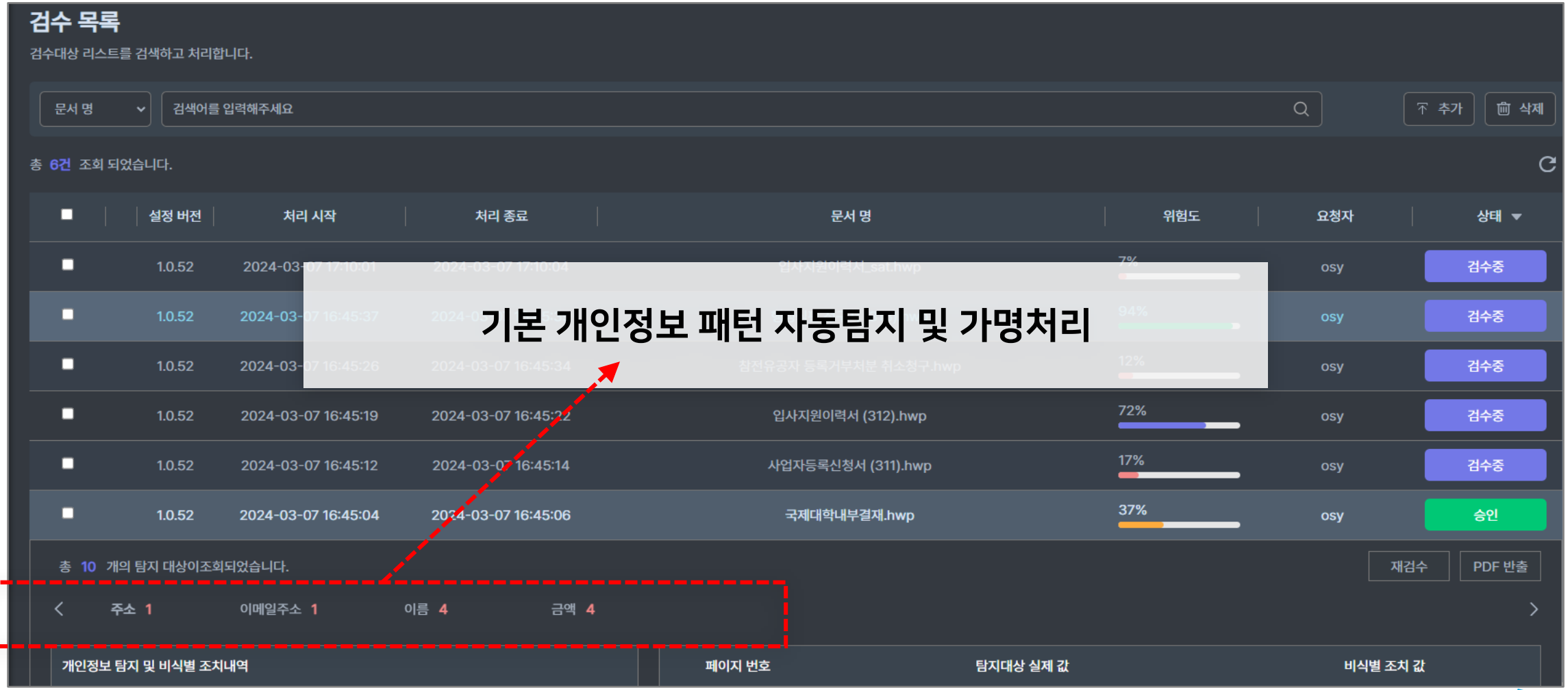
Docu-Guard Process

반출(재사용)할 수 있는 구조



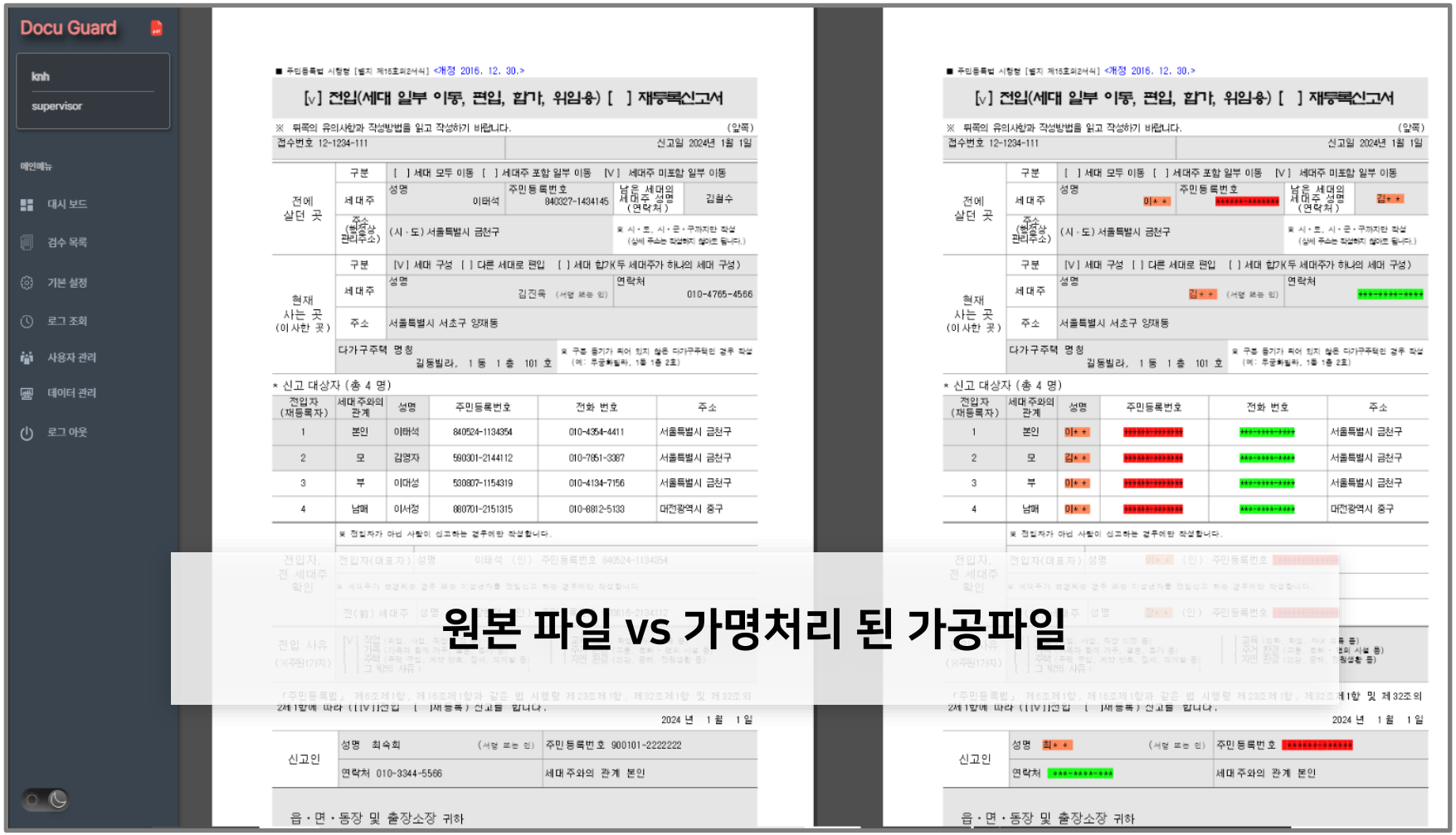
개인정보 자동 가명처리

개인정보 자동 가명처리



개인정보 자동 가명처리

쉽게 확인 할 수 있는 가명처리 정보

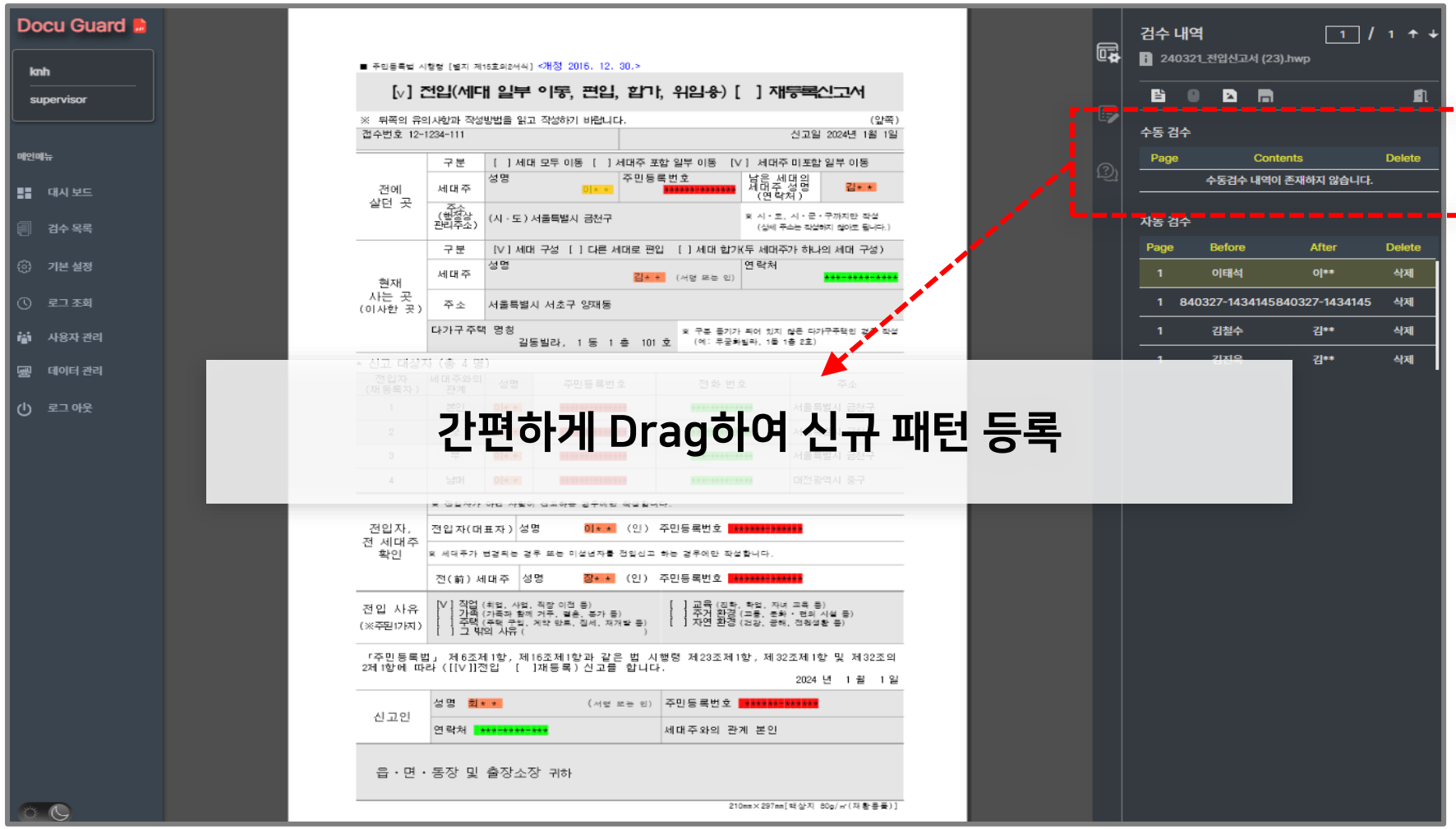


■ Dual Mode 적용

- HTML5 지원
- 한 화면에서 원본 vs 가공파일 비교
- 기본 Base Pattern 적용

개인정보 사용자 처리방식

사용자 정의에 의한 수동 가명처리 방식

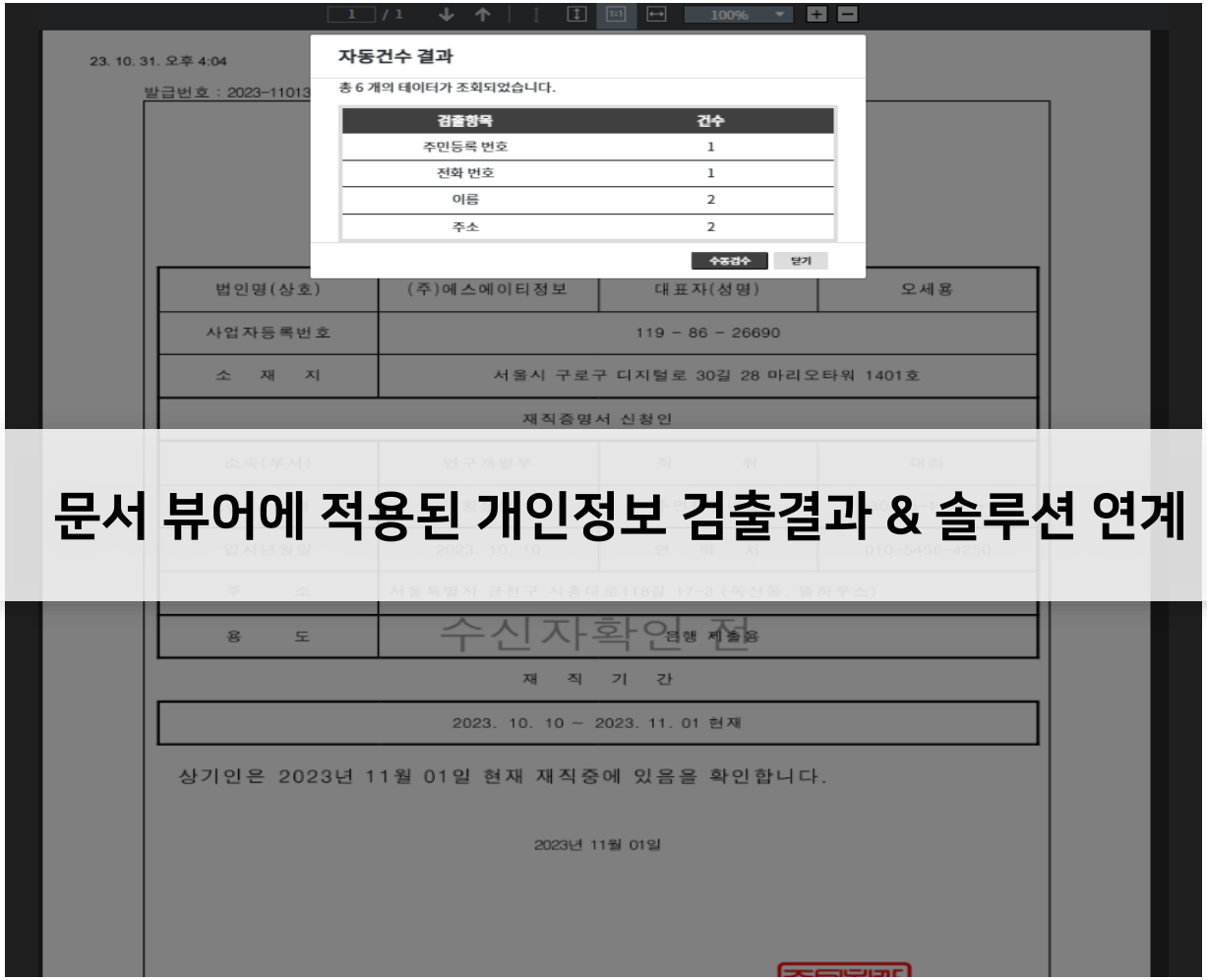


■ Drag 방식

- Base Pattern 이외의 텍스트 정보 등을 drag하여 손쉽게 가명처리
- 신규 등록 패턴에 대한 일괄적용 가능
- 문서 內 이미지 객체에 대해 인식 & 가명처리 가능

사용성 강화 – 他 솔루션 연계

문서 뷰어 및 타 개인정보보호 솔루션 연계를 통한 사용성 강화



문서 뷰어에 적용된 개인정보 검출결과 & 솔루션 연계

■ API 제공

1. Viewer 연계

- 기관에서 이용중인 문서 뷰어 연계 가능
- 파일 열람 시, 뷰어에서 개인정보 자동 검출 기능이용
- 자동 검출 된 개인정보 내역 확인 후,
- Docu-Guard 연계 API를 이용하여 사용 가능

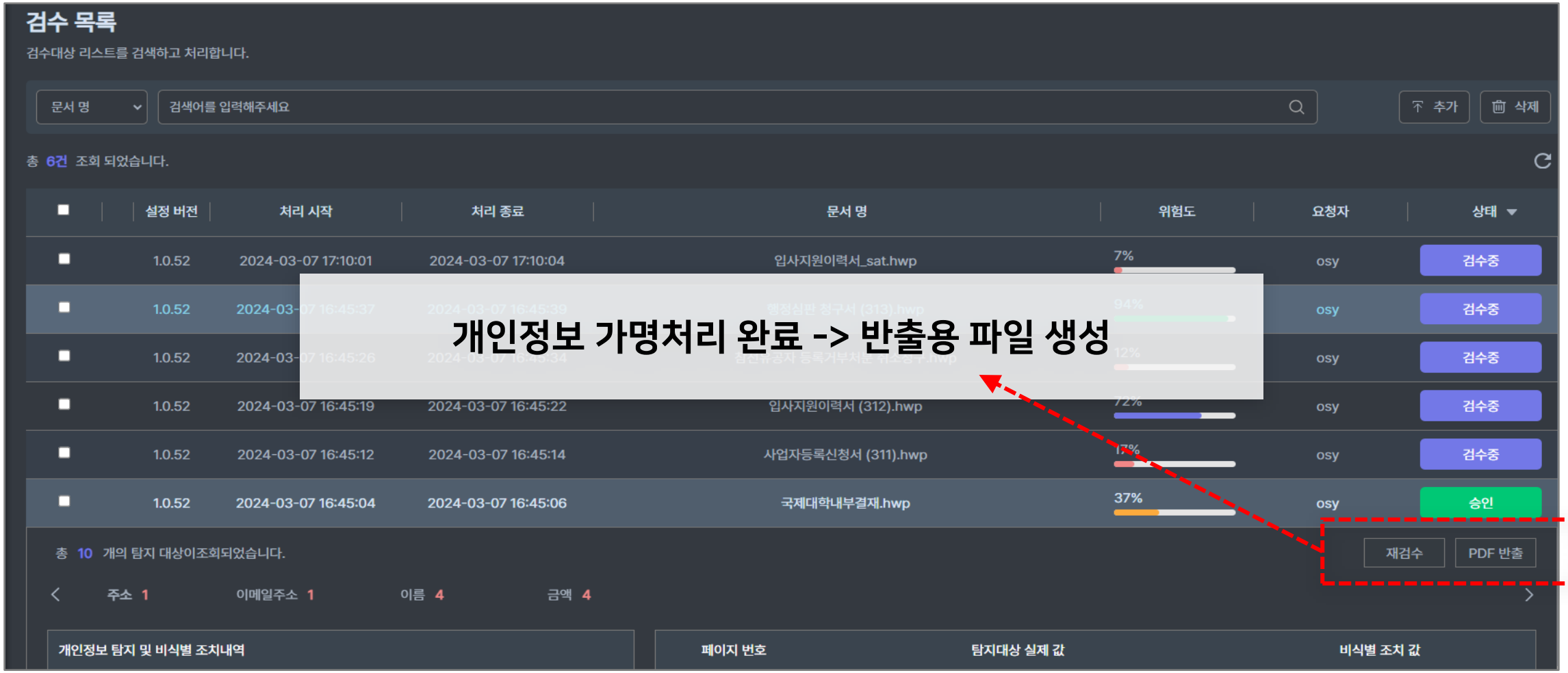
2. 개인정보 탐지 솔루션 연계

- 기관에서 이용중인 개인정보 탐지 솔루션 연계 가능
- 기 사용중인 솔루션을 통해 개인정보 탐지 결과를 제공받고,
- 이를 메타데이터와 연계하여 원본 파일에 가명처리

SAT

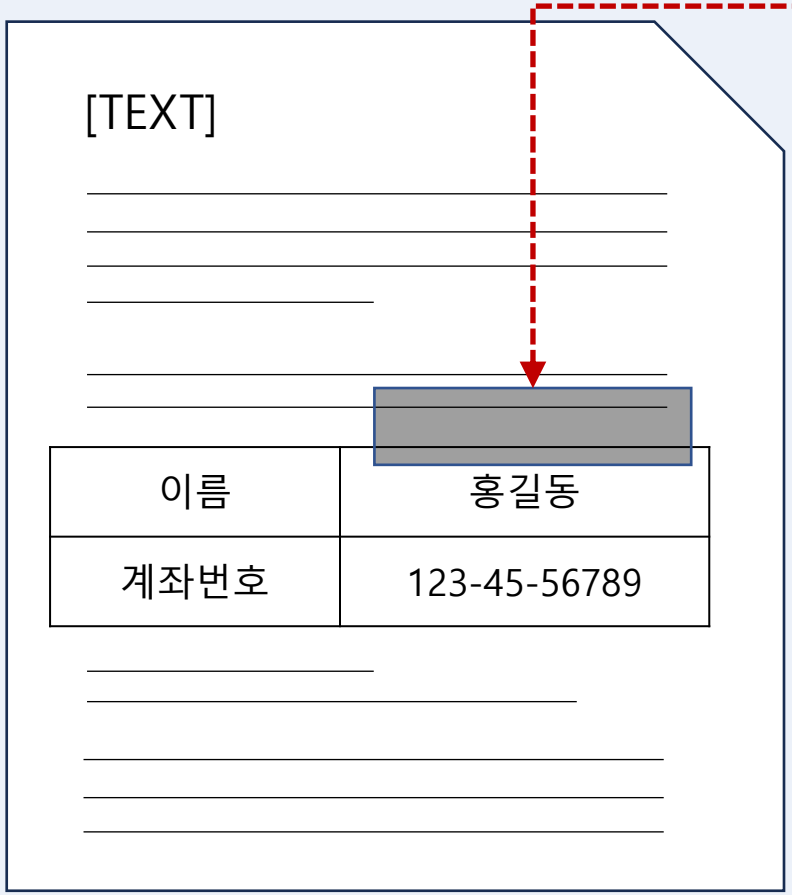
반출용 파일 생성

반출용 파일(PDF) 생성



가명처리 기술비교

■ 한글, OFFICE, PDF 등 텍스트 파일



1. Black Masking 방식

- 원본 ➔ 이미지 or PDF 파일로 변환
- [redacted] 을 좌표 위에 덧씌우는 방식
- 좌표가 어긋나는 현상 & 기본 패턴 외 개인정보 처리 난항

2. Text 추출 방식

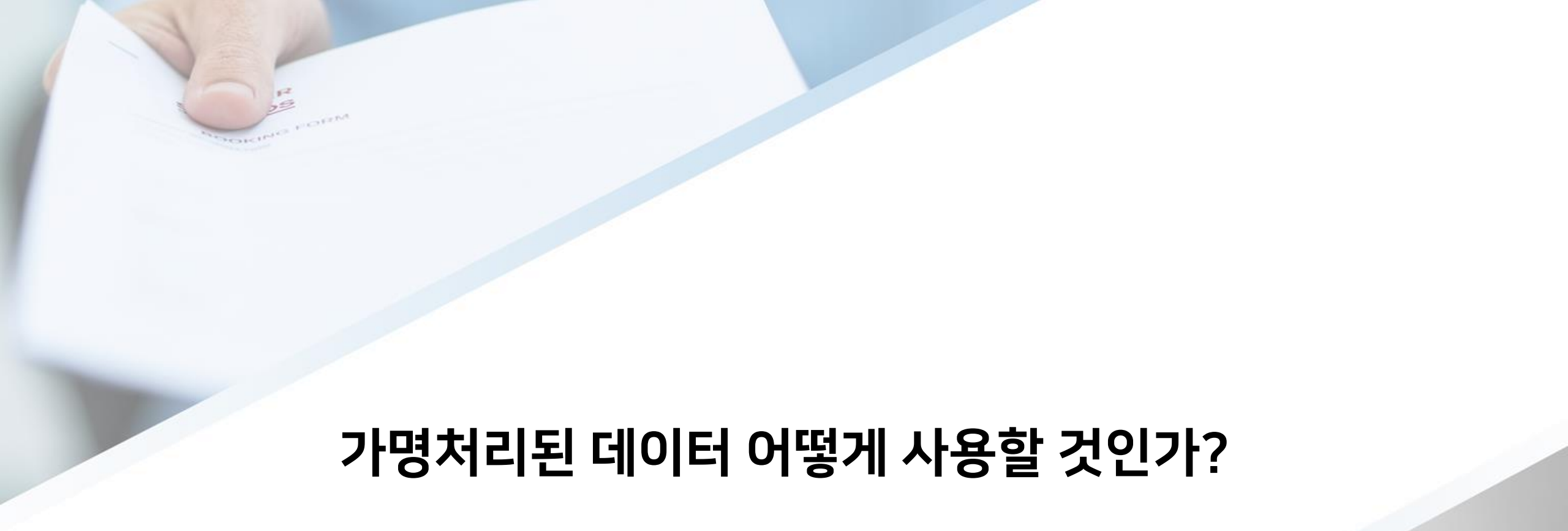
- 원본 ➔ 텍스트 파일을 추출
- 텍스트 파일 ➔ 개인정보 가명처리 ➔ 가명처리 된 별도의 파일생성
- 가명처리 정보를 원본에 재결합 하는 방법 난항

3. Meta Data 추출 방식

- 원본 ➔ Meta Data 추출
- Source Data ➔ 직접 가명처리 ➔ 가명처리 된 사본 생성
- 원본형식을 유지한 사본 생성

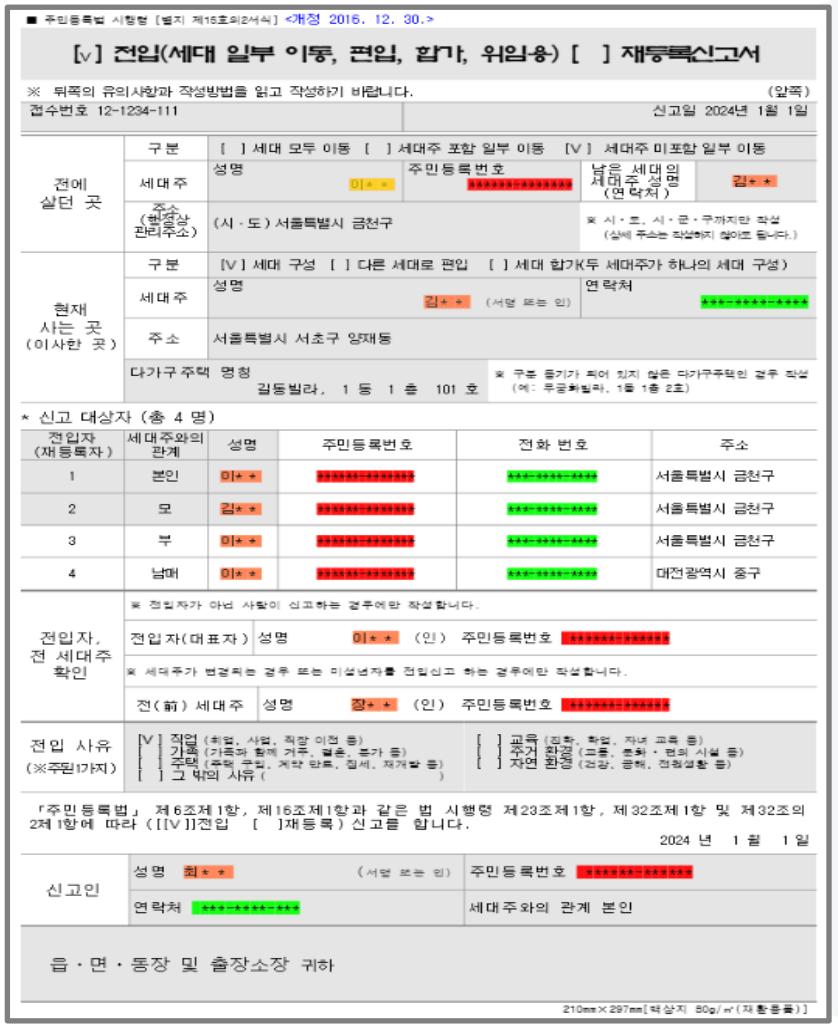
레퍼런스 예시


행정안전부
국가기록원
통합검색
검색어를 입력해주세요.
상세검색
마이페이지
Language
기록물 열람
참여 · 민원
정보공개
업무안내 · 자료
뉴스 · 소식
기록원소개
기록관리 이슈페이퍼
VOL. 40
기록관리 이슈페이퍼
공익적 기록보존
소장 기록물 현황
일반 기록물
5,583,892
342,325
853,286
358,043
정보열람 및 정보공개
찾아오시는 길
주요전화번호 안내
저작권정책
개인정보처리방침
고객만족도조사
검색도움말
FAQ
행정안전부
국가기록원



가명처리된 데이터 어떻게 사용할 것인가?

가명처리 된 반출용 파일 - 활용방안



■ 주민등록법 시행령 [별지 제15호의2서식] <개정 2016. 12. 30.>

[v] 전입(세대 일부 이동, 편입, 합가, 위임용) [ ] 재등록신고서

※ 뒤쪽의 유의사항과 작성방법을 읽고 작성하기 바랍니다. (앞쪽)

접수번호 12-1234-111 | 신고일 2024년 1월 1일

전에 살던 곳	구분	[ ] 세대 모두 이동 [ ] 세대주 포함 일부 이동 [V] 세대주 미포함 일부 이동			
	세대주	성명 이**	주민등록번호 **********	남은 세대의 세대주 성명 (연락처)	김**
	주소 (법정상 관리주소)	(시·도) 서울특별시 금천구		※ 시·도, 시·군·구까지만 작성 (상세 주소는 작성하지 않아도 됩니다.)	
현재 사는 곳 (이사한 곳)	구분	[V] 세대 구성 [ ] 다른 세대로 편입 [ ] 세대 합가(두 세대주가 하나의 세대 구성)			
	세대주	성명 김** (서명 또는 인)		연락처 ***-****-****	
	주소	서울특별시 서초구 양재동			
	다가구주택 명칭	길동빌라, 1 동 1 층 101 호		※ 구분 등기가 되어 있지 않은 다가구주택인 경우 작성 (예: 무궁화빌라, 1동 1층 2호)	

* 신고 대상자 (총 4 명)

전입자 (재등록자)	세대주와의 관계	성명	주민등록번호	전화 번호	주소
1	본인	이**	**********	***-****-****	서울특별시 금천구
2	모	김**	**********	***-****-****	서울특별시 금천구
3	부	이**	**********	***-****-****	서울특별시 금천구
4	남매	이**	**********	***-****-****	대전광역시 중구

전입자, 전 세대주 확인	※ 전입자가 아닌 사람이 신고하는 경우에만 작성합니다.			
	전입자(대표자)	성명 이** (인)	주민등록번호 **********	
	※ 세대주가 변경되는 경우 또는 미성년자를 전입신고 하는 경우에만 작성합니다.			
	전(前) 세대주	성명 장** (인)	주민등록번호 **********	
전입 사유 (※주된1가지)	[V] 직업 (취업, 사업, 직장 이전 등) [ ] 가족 (가족과 함께 거주, 결혼, 분가 등) [ ] 주택 (주택 구입, 계약 만료, 입세, 재개발 등) [ ] 그 밖의 사유 ( )		[ ] 교육 (진학, 학업, 자녀 교육 등) [ ] 주거 환경 (교통, 문화·편의 시설 등) [ ] 자연 환경 (건강, 공해, 전원생활 등)	

「주민등록법」 제6조제1항, 제16조제1항과 같은 법 시행령 제23조제1항, 제32조제1항 및 제32조의 2제1항에 따라 ([V]전입 [ ]재등록) 신고를 합니다.

2024 년 1 월 1 일

신고인	성명 최** (서명 또는 인)	주민등록번호 **********	
	연락처 ***-****-****	세대주와의 관계 본인	

읍·면·동장 및 출장소장 귀하

210mm×297mm[백상지 80g/㎡(재활용품)]

1 **정보를 '공유' 할 수 있는 데이터로 사용가능**

- 가명처리 된 파일 ➔ 외부와의 **'정보 공유'**

2 **정보를 '공개' 할 수 있는 데이터로 사용가능**

- 소극적 민원대응 ➔ **'적극적 민원대응'** 으로

3 **신뢰할 수 있는 '데이터' 로 사용가능**

- 가명처리 된 파일 ➔ **'AI를 활용할 수 있는 기초데이터'** 로 활용

SAT

가명처리 된 반출용 파일 - 활용방안

가명처리 데이터

AI를 위한 기초데이터

Machine Readable(정형) 데이터

Fine Tunning

SLM

4 AI 활용을 위한 '가치있는 데이터'로 사용가능

■ 주요 정보자산 활용으로 '데이터 부족' 문제 해결

- 비정형 데이터는 조직의 주요자산
- AI를 Business Domain으로 사용하기 위해 필수 데이터

■ 보다 '안전한 AI 생태계' 조성 기반

- 출처가 명확한 데이터 사용으로 Hallucination 최소화
- 개인정보 오남용 문제 최소화

Thank You