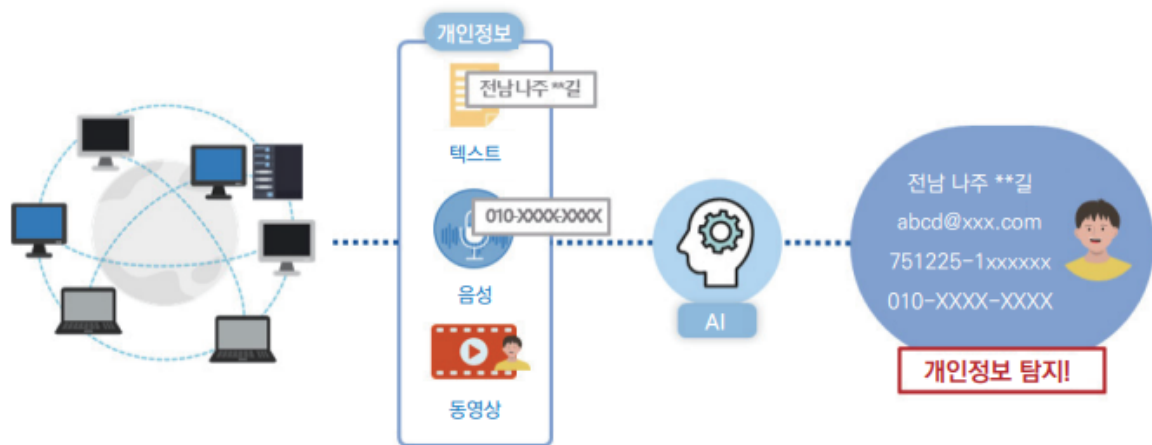


# 비정형 데이터(텍스트·영상·음성 등)에서 개인정보 탐지 기술



구분	2022	2023	2024	2025	2026
비정형 데이터에서 개인정보 탐지 기술	대화형 텍스트 데이터에서 개인정보 탐지 기술				
	영상 데이터에서 개인정보 탐지 및 동일객체 연결 분석 기술				
		음성 데이터에서 개인 식별 및 개인정보 탐지 기술			
				혼합형 비정형 데이터에서 개인정보 통합 탐지 기술(4년)	

## 기술 세부내용


- 다양한 유형의 데이터에 대한 **개인정보 탐지기술** 개발
  - 대화형 텍스트, 음성 및 영상 데이터 등 비정형 데이터에 대한 개인정보 식별을 위해 필요한 기술(**텍스트 분석·가공, 자연어 처리** 등) 개발
- 온라인 서비스 내 콘텐츠에 대한 개인정보 검색 및 수집을 위한 **AI기반** 콘텐츠 분석 및 비정형 개인정보 분석 기술 개발
- 개발된 기술에 대한 검증 및 성능측정을 위한 표준 데이터셋과 검증 방법론을 개발

## 표준화 추진

- 영상, 이미지, 음성, 텍스트 데이터 등 데이터 유형별로 개인정보의 효과적인 탐지를 위한 알고리즘 및 분석 방법 표준화 추진

개인정보위, 무단 정보수집 탐지 기술 등 R&D 청사진 공개

개인정보보호위원회가 2022년 사전 동의 없이 정보주체 온라인 활동 기록을 수집하는 추적사이트 탐지 기술 개발에 들어간다. 대화형 텍스트나 영상 및 음성 등 비정형 데이터에서 개인정보 존재여부를 탐지하는

 [https://it.chosun.com/site/data/html\\_dir/2021/11/10/2021111001576.html](https://it.chosun.com/site/data/html_dir/2021/11/10/2021111001576.html)

2. 유·노출 최소화	① 다크웹 개인정보 거래 추적 및 차단 기술	다크웹 접속 및 개인정보 검색, 불법거래 추적 기술
	④ 비정형 데이터 개인정보 탐지	대화형 데이터 개인정보 탐지 기술, 영상데이터 개인 식별 기술 ('22년 예산 반영)
3 안전한 활용	⑤ 개인정보 파편화 및 결합 기술	이미지 기반 신분증 파편화 및 결합 활용 기술
	⑥ 비정형 데이터에서 선택적 개인정보 파기	대화형 영상데이터에서 선택적 개인정보 삭제 기술
3 안전한 활용	⑦ 차세대 가명·익명 처리 및 결합 기술	실시간, 트랜잭션 데이터의 가명·익명 처리 기술('22년 예산 반영)
	⑧ 가명·익명 정보 안전성 평가	가명·익명정보 재식별 공격·검증 모델 기술
3 안전한 활용	⑨ 개인정보 변조 및 재현데이터 생성	음성 개인정보 변조 기술, 재현 데이터 자동생성 AI 모델 기술
	⑩ 프라이버시 보존형 개인 맞춤 서비스	영자식증명 기반 개인증명 기술, 차분 프라이버시 기반 AI 큐레이션 기술
3 안전한 활용	⑪ 마이데이터 처리 및 관리 기술	신뢰 전송 기술, 활용 지원 기술 및 통합 관리 플랫폼

## 개보위가 어떤 기술 개발하는지에 대한 내용

### 개인정보위, 무단 정보수집 탐지 기술 등 R&D 청사진 공개

중분류(분야)	핵심 요소기술	주요 세부기술
1. 정보주체 권리보장	① 개인정보 동의 관리 기술	동의 관리 플랫폼 기술
	② 정보주체의 온라인 활동기록 통제	맞춤형 온라인 활동기록 통제 기술 및 실증('22년 예산 반영)
	③ 다크웹 개인정보 거래 추적 및 차단 기술	다크웹 접속 및 개인정보 검색, 불법거래 추적 기술
2. 유·노출 최소화	④ 비정형 데이터 개인정보 탐지	대화형 데이터 개인정보 탐지 기술, 영상데이터 개인 식별 기술 ('22년 예산 반영)
	⑤ 개인정보 파편화 및 결합 기술	이미지 기반 신분증 파편화 및 결합 활용 기술
	⑥ 비정형 데이터에서 선택적 개인정보 파기	대화형·영상데이터에서 선택적 개인정보 삭제 기술
3 안전한 활용	⑦ 차세대 가명·익명 처리 및 결합 기술	실시간, 트랜잭션 데이터의 가명·익명 처리 기술('22년 예산 반영)
	⑧ 가명·익명 정보 안전성 평가	가명·익명정보 재식별 공격·검증 모델 기술
	⑨ 개인정보 변조 및 재현데이터 생성	음성 개인정보 변조 기술, 재현 데이터 자동생성 AI 모델 기술
	⑩ 프라이버시 보존형 개인 맞춤 서비스	영자식증명 기반 개인증명 기술, 차분 프라이버시 기반 AI 큐레이션 기술
	⑪ 마이데이터 처리 및 관리 기술	신뢰 전송 기술, 활용 지원 기술 및 통합 관리 플랫폼

- 인공지능 챗봇·스피커 등에서 처리되는 텍스트 음성에서 개인정보를 탐지하고 삭제하는 비정형 데이터 개인정보 탐지·삭제 기술
- 개인을 알아볼 수 없도록 개인정보를 분해해 저장하고, 필요한 경우에만 결합해 활용할 수 있는 개인정보 파편화 및 결합 기술도 지원

## 인공지능 기술을 활용한 이미지 개인정보와 기술 유출 차단 방안

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/575078b6-bedd-4911-ac1d-cc091b40dc66/%EC%B5%9C%EB%B3%B5%ED%9D%AC\\_\(%EC%B5%9C%EC%A2%85\).pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/575078b6-bedd-4911-ac1d-cc091b40dc66/%EC%B5%9C%EB%B3%B5%ED%9D%AC_(%EC%B5%9C%EC%A2%85).pdf)

### OCR 활용(광학문자인식)

- 이미지에서 문자를 추출하는 소프트웨어
- ImageOCR 모듈을 이용하여 업무 시스템, 망연계시스템, 내부정보 유출차단 장비, 개인정보 차단장비, 문서보안솔루션, 파일서버 등과 연동
- 기업과 국가 제휴..

## 기타 조사

- 국가적 공조수사 체계 확립
- 암호화폐사와도

### 잊힐 권리 지원 대상 게시물 유형(안)

- 정보주체가 게시한 글을 제3자가 공유(링크, 복제 등)한 경우

### 유출 시 피해 확산방지 및 피해 복구 필요

- 피해 확산 방지: 다크웹 상 개인정보 노출 실시간 모니터링 후 즉각 처치
- 피해 복구: 디지털 장의사 확대 도입?

- 클라우드를 통해 국가 및 국내 기관 별 공조수사/연계수사 서버 플랫폼 구축

## 다크넷 범죄현상과 형사법적 대응방안

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/632cc855-6fa3-47c6-b468-d2ee7fdcd822/KCI\\_FI002327753.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/632cc855-6fa3-47c6-b468-d2ee7fdcd822/KCI_FI002327753.pdf)

- 현행법으로는 다크넷 범죄에 소극적으로 대응할 수밖에 없는 것이 현실이다.
- 수사기관이 다크넷 시장에서 범죄가 발생했다는 것을 인지한다해도 그 특유의 폐쇄성으로 인해 당사자에 대해 대인적 강제 처분을 집행하는 것이 현실적으로 불가능에 가까우며, 대물적 강제처분을 하기 위해서도 필요한 자료를 입수하기가 매우 어렵기 때문이다.
- 따라서 수사기관이 다크넷 시장이라는 온라인 공간에서 범죄혐의 입증이 가능하고 영장 발부를 위한 전제요건을 충족할 수 있는 수단이 필요하다.

### EU의 사이버범죄 방지조약

- 온라인 수색과 원격지 압수(제19조), 공식적 또는 동의에 의한 컴퓨터 데이터의 초국경적 접속(제32조), 통신데이터의 실시간 수집(제33조) 등 다크넷 범죄에 유효적절하게 적용할 수 있는 대응수단을 규정
- 우리나라도 본 조약에 가입하기 위해 꾸준한 노력을 기울이고 있으나 가입의 전제요건인 국내입법절차가 이행되지 않아서 아직 가입하지 못한 상태이다
  - 이에 대한 개괄적 연구검토로는 이경렬·하건우, “유럽평의회 사이버범죄조약의 가입·비준을 위한 국내 이행법률의 마련과 준비 비교”, 『비교형사법연구』, 제19권 제4호(2018), 501면 이하

### 원격지 압수

- 유체물에 대한 증거확보는 사람과 장소를 장악하면 통제할 수 있으나, 디지털 정보는 시간과 장소에 구애받지 않고 권한만 있다면 접속할 수 있어 수사기관의 통제가 큰 의미를 갖지 못한다는 점을 고려해야 한다.
- 따라서 행위자와 서버 등이 공간적으로 분리되어 있는 경우에도 압수를 진행할 수 있도록 강제처분의 관련성 요건을 완화하는 원격지 압수의 필요성이 제기된다

- 온라인 수색을 반대하는 입장의 주된 논거는 관련성을 좁게 해석하는 전제에서 영장주의의 침탈을 초래한다는 데에 있다
- 특정한 범죄군에 한해 제한적으로 온라인 수색을 허용하는 방안을 고려해야 할 것
  - **국민의 범죄화요구가 제기되는 일부 범죄를 신범죄화**하고 이들 제한된 범죄군에 대해서 온라인 수색과 원격지 압수 도입을 고려

## 다크넷 범죄에 대한 오프라인 차원의 대응

- 선진화 · 합법화되고 있는 조직범죄 통제의 관점에서 잠입수사관 제도 도입을 고려할 필요

## 정보협력자 제도의 검토

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/842d1f8d-59b4-4ea9-af37-fa146b8dd2a9/KCI\\_FI002667838.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/842d1f8d-59b4-4ea9-af37-fa146b8dd2a9/KCI_FI002667838.pdf)

- 국제공조 뿐 아니라 잠입수사나 합정수사와 같은 언더커버 수사, 해킹 툴 사용 및 범죄 서버의 일시 운영, 암호화폐 수사 등 특수한 수사방법들이 요구된다. 문제는 그러한 수사방법들을 임의수사에 해당 한다고 보기 어려운 측면이 있기 때문에 **강제수사로서 명확한 입법규정이 필요함**
- 2019년 7월 동부지검에 사이버수사부를, 9월에는 중앙지검에 다크넷 수사팀을 만듦
- 전자영장
  - 원격지에 있는 제3자에게 전자영장을 제시하게 되면 제3자는 해당 전자 영장을 면책 근거로 사용하고, 명확하고 신속한 협조를 할 수 있게 된다
  - 형사소송법의 개정이 요구된다. 예를 들어, 형사소송법 제118조에 전자영장의 제시에 관한 내용이 신설될 필요