

# 인공지능 학습데이터 자발적 확보를 위한 멀티모달 마이데이터 유통시스템 설계

임동현\* · 박대우\*\*

## Designing a Multimodal MyData Distribution System for Voluntary Acquisition of AI Training Data

Dong-Hyun Lim\* · Dea-Woo Park\*\*

### 요약

인공지능의 활용을 위해서는 학습이 필요하고, 학습을 위해서는 데이터가 필요하다. 데이터는 산, 바다, 지형 등 저작권이 없는 것도 있고, 개인정보보호법 및 저작권법 등 각종 법률에 제한받는 데이터도 있다. 본 논문은 정보 주체가 자발적으로 데이터의 수집, 활용, 유통에 동의하고 참여하여, 법률적 제한을 극복하는 방법을 연구한다. 공공장소에 특정한 공간을 만들고, 기업이 참여하여 학습에 필요한 데이터를 정의하고, 국민은 특정한 공간에서 자발적으로 멀티모달 마이데이터의 수집에 참여하여 보상받는 시스템을 설계한다. 또한, 정부에서 운영 중인 마이데이터 플랫폼과 연계해서 생성된 데이터의 인증, 유통, 판매/재판매가 가능한 시스템을 구현한다. 이를 정부 주도로 진행한다면, 학습 도메인별로 학습용 데이터를 법률에 제재받지 않고 새로운 방법으로 데이터의 수집을 할 수 있어, 인공지능 기술 발전과 활용 방안이 더욱 활성화되는 계기가 될 것이다.

### ABSTRACT

AI requires learning, and learning requires data. Some data is copyright-free, such as mountains, oceans, and terrain, while others are restricted by various laws, such as privacy and copyright laws. This thesis investigates how data subjects can voluntarily consent and participate in the collection, utilization, and distribution of their data, overcoming legal restrictions. We design a system that creates specific spaces in public places, engages businesses to define the data needed for learning, and rewards citizens for voluntarily participating in the collection of Multimodal MyData in specific spaces. In addition, a system that enables authentication, distribution, and sale/resale of generated data in connection with the government's MyData platform will be implemented. If this is led by the government, it will be possible to collect data for learning in a new way without legal sanctions for each learning domain, which will further revitalize the development and utilization of AI technology.

### 키워드

Artificial Intelligence, Privacy Protection, Training Data, Multimodal MyData, MyData Distribution Platform  
인공지능, 개인정보 보호, 학습 데이터, 멀티모달 마이데이터, 마이데이터 유통 플랫폼

\* 호서대학교 벤처대학원 융합공학과az062112@naver.com

\*\* 교신저자 : 호서대학교 벤처대학원 융합공학과

· 접수일 : 2024. 07. 27

· 수정완료일 : 2024. 09. 03

· 게재확정일 : 2024. 10. 12

· Received : Jul. 27, 2024, Revised : Sep. 03, 2024, Accepted : Oct. 12, 2024

· Corresponding Author : Professor, Dea-Woo Park

Dept. of Convergence Engineering, Hoseo Graduate School of Venture, Seoul

Email : prof\_pdw@naver.com

## 1. 서 론

인공지능(Artificial Intelligence, “AI”)이 산업과 사회에 적용하기 위해서는 사전 학습이 필요하고, AI 학습을 위해서는 데이터가 필요하다. 데이터는 산, 바다, 지형 등 저작권이 없는 것도 있고, 개인정보보호법 및 저작권법 등 각종 법률에 제한받는 데이터도 있다.

개인식별정보나 저작권이 포함된 데이터는, AI 학습을 위한 활용이 현실적으로 어렵다. 이러한 상황의 개선하기 위해서는, 데이터의 수집을 위한 새로운 방법이 필요하다.

정부는 국민의 데이터 주권을 보장하고 한국 데이터 경제 체질을 근본적으로 개선하기 위해 「국가 마이데이터 추진 전략」을 마련하여 추진하고 있다<sup>1)</sup>.

마이데이터(MyData)란, 정보 주체가 본인 정보의 3자 전송 요구권을 통해 데이터를 원하는 곳으로 이동시켜 혁신적 서비스에 활용할 수 있도록 하는 21세기형 자기 결정권을 말한다<sup>1)</sup>.

개인정보보호법 및 저작권법 등 각종 법률에 제한받는 마이데이터는 AI 학습용 데이터로서 가치가 상대적으로 높고, 산업과 사회의 활용도도 높다.

따라서, 개인정보를 보호하고 저작권에 따른 분쟁 가능성을 일소하기 위해, 개인의 멀티모달(Multimodal) 정보를 마이데이터로 정의하여 AI 학습용 데이터로 활용할 수 있는 새로운 데이터 유통시스템에 관한 연구가 필요하다.

본 논문에서는 정부의 마이데이터 유통 표준 절차를 변형하여, 개인이 자발적으로 학습용 멀티모달 마이데이터를 수집/제공/유통하고 보상받는 체계의 구축을 연구한다.

멀티모달 마이데이터 유통시스템 모델을 정부 주도로 진행된다면, 산업 도메인별 AI 학습데이터를 수집할 수 있어, 대한민국의 산업과 사회의 AI 기술이 발전할 계기가 될 것이다.

## II. 관련연구

### 2.1 마이데이터 유통 절차

마이데이터 사업은, 정보 주체 측면에서는 “AI 중심

디지털 심화 시대”에 적극적인 개인 권리를 보장하여 국민 개인정보의 자기 결정권을 강화하고, 데이터 생태계 측면에서는 기업·기관별 칸막이에 막혀 있는 데이터를 정보 주체 의사에 따라 융합하여 데이터 활용 생태계를 한 단계 진일보하기 위해 시도하는 사업이다<sup>2)</sup>.

정부가 제시한 마이데이터 서비스 절차는 그림 1과 같다. 정보주체(Owner)는 마이데이터를 활용하려는 활용자(User)에게 마이데이터 결합/사용/이전 등을 통합(Single Sign-on) 인증을 통해 승인하면, 활용자는 마이데이터 활용이 가능하도록 설계되어 있다.

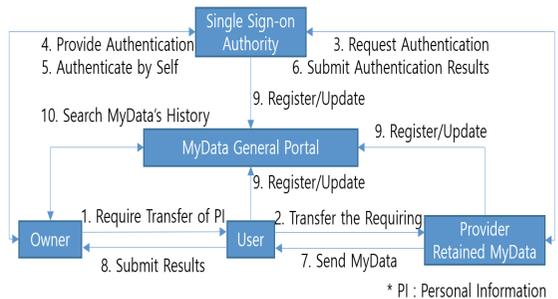


그림 1. 마이데이터 서비스 절차  
Fig. 1 Process of MyData service

### 2.2 정부의 AI 학습용 데이터 구축 현황

「한국지능정보사회진흥원」(이하 “NIA”)은 AIHub 사이트(이하 “AIHub”)를 통해 『AI 학습용 데이터 구축 사업』을 2017년부터 추진하고 있다. 2024년 4월 말 기준으로 구축 완료된 데이터는 703건<sup>2)</sup>으로 투입된 예산은 1.5조원 이상이다.

NIA는 학습용 데이터 구축의 일관성과 기술적 방향 제시를 위해 『AI 학습용 데이터 품질관리 가이드라인』과 『AI 학습용 데이터 구축 안내서』<sup>3)</sup>를 제작 후 매년 업데이트하고 있다.

학습용 데이터를 제작하는 과정은 그림 2와 같다.

단계별로 살펴보면, 데이터 정의 단계에서는 필요한 데이터를 정의하고, 학습을 위한 모델에 적합한 데이터를 구체적으로 규정한다.

데이터 획득 단계에서는 개인정보보호법 또는 저작권법 등 관련한 법률적 제약을 해결할 수 있는 상태

1) <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000 &nttId=9105>

2) <https://www.aihub.or.kr/>

3) <https://aihub.or.kr/aihubnews/qlityguidance/view.do?currMenu=135&topMenu=103&nttSn=10125>

로 원시데이터를 획득 또는 직접 생성한다.

데이터 정제 단계에서는 정의된 형식에 따라 원시 데이터에서 학습이 가능한 데이터를 선택한다. 이때, 중복된 데이터 제거와 개인정보의 비식별화 처리 과정이 포함되기도 한다.

데이터 라벨링 단계에서는 학습할 수 있는 형태로 정제데이터를 가공한다. 이 결과로 생성된 데이터를 학습용 데이터로 정의하며, 이 데이터는 특정한 규칙에 따라 묶음 형태로 구조화되어 AIHub를 통해 공개된다.

데이터 학습/활용 단계에서는 학습이 필요한 개발자가 공개된 학습용 데이터를 사용하여 모델을 학습한다.

개방된 AIHub 데이터를 살펴보면, 개인정보보호법, 저작권법 등 각종 법률 및 법규 준수로 인해 데이터로 만들어지지 않는 데이터도 있다. 즉, 도로시설물과 동식물 등 개인정보가 없는 데이터는 종류가 다양하나, 얼굴, 차량번호판 등 개인정보인 데이터는 제한적인 이미지에 대해 모자이크 등 비식별화 처리하여 공개거나 전문 배우 등의 인위적 재연을 통해 촬영되었다[3].

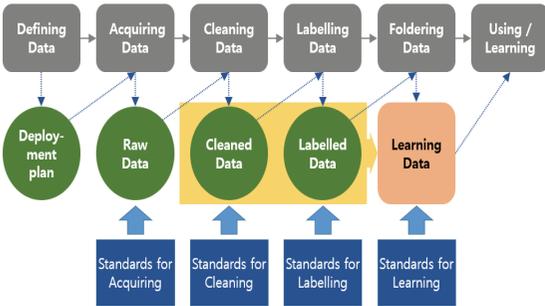


그림 2. 학습데이터 구축 절차  
Fig. 2 Process of learning data building

### 2.3 AI 학습용 데이터 구축에 관한 요구 분석

AIHub 데이터 중 하나인 “자동차 차종/연식/번호판 인식용 학습용 데이터 영상”을 살펴보면, 개인정보 문제를 해결을 위해 차량과 차량번호판을 분리하여 데이터를 구축하였다.

이런 데이터를 사용하여 차량 및 차량번호 인식 학습을 위해서는 네트워크모델의 신뢰도를 낮추는 방향의 “후처리 결합 공정”이 추가 되어야 한다.

즉, 한 단계의 네트워크모델로 학습할 과정을 두 단계 이상의 학습과 두 개 이상의 네트워크모델로 검증하여야 하고, 이때, 모델 중첩으로 인해 학습의 신뢰도는 낮아진다. 그림 3은 AIHub에 등록된 글이다.

또한, 트랜스포머(Transformer)[4] 기술을 시작으로, ChatGPT 등 텍스트 기반의 AI 서비스 성능이 향상되고, 텍스트가 음성, 이미지, 영상과 상호 변환되는 기술이 발전하면서[5] “멀티모달”에 대한 학습 도메인 데이터 수요가 증가하고 있다.

이에 따라, 2024년 인공지능 학습용 데이터 구축 사업의 세부 사업 대부분이 LLM(Large Language Model) 기반 학습용 데이터 구축이다<sup>4)</sup>.

엄태\* 2021.09.24

안녕하세요, 저는 프로젝트 이름에 "인식"이라는 단어가 포함된 관계로 자동차번호판 인식 실험을 진행하기 위하여 데이터를 다운로드 하였어요.

현재 제공하고 있는 자동차 이미지에서 번호판은 마스크 처리 되어 있고 이 자동차 정보를 기록한 json 파일 중 번호판의 위치는 원본 이미지에서의 위치입니다.

자동차 정보를 기록한 json 파일을 보시면 video 이름, frame index가 모두 기록되어 있는데 영상과 이미지도 제공되어 있지 않습니다. 혹시 원본 이미지/영상을 추가로 제공해줄 수 있나요?

---

AI허브 2021.09.24

안녕하세요. AI허브입니다.

문의하신 자동차 차종/연식/번호판 인식용 영상 데이터는 자동차 번호판 OCR데이터, 차종 분류 데이터로 나누어 구축되었습니다.

차종 분류 데이터는 자동차의 차종을 분류하기 위한 목적으로 데이터를 구축했고 자동차 번호판 OCR데이터는 번호판 식별을 위한 목적으로 구축했기 때문에 원천 데이터 이미지에서 차량 및 번호판을 그룹하여 비식별화 조치 및 라벨링 작업을 했습니다.

본 데이터셋은 원본 동영상상 정제 및 가공하여 AI에 학습시키기 위한 라벨링된 이미지 데이터를 제공하며, 원본 데이터를 추가 제공해 드리지 않습니다.

그림 3. AIHub에 등록된 글  
Fig. 3 Posts registered in AIHub

### 2.4 비식별화 처리된 영상데이터 학습의 효과성 분석

개인정보의 삭제 처리를 위해 사용할 수 있는 기술로는 그림 4와 같이 삭제와 모자이크(mosaic) 처리가 있다. 개인정보를 포함하는 특정 영역을 모자이크 처리하여 학습하면, 왜곡된 데이터로 인해 학습모델의 정확도가 20%~40% 낮아지게 된다는 연구 결과가 있다[6].

표정의 인식을 위한 유전자 알고리즘 기반 심층학습 네트워크 최적화[7] 또는 보안 감시를 위한 심층학습 기반 다채널 영상 분석[8] 등의 신경망 활용의 전처리로 개인정보를 비식별화하여 학습하기 전에, 모델의 입력데

4) [https://www.nia.or.kr/site/nia\\_kor/ex/bbs/View.do?cbIdx=99835&bcIdx=26446&parentSeq=26446](https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=99835&bcIdx=26446&parentSeq=26446)

이터 특성에 따라, 어떤 개인정보를 어느 정도로 비식별화할지, 그에 따라 신뢰도 있는 결과가 도출될지 등을 사전에 연구하고, 데이터를 만들어야 하나, AIHub의 데이터 구축 과정에는 이러한 단계가 없다.

따라서, AI 예측 정확도 향상을 위해 비식별화 처리되지 않은 원본 데이터를 사용한 학습과 그에 따른 모델 개발이 필요하며, 이에 관한 선행 연구[9]가 진행되기도 하였다. 또한, 원본 데이터 학습에 의한 자율주행 기술 구현[10]과 학습 데이터 기반의 2D-3D 매핑(mapping) 라이브러리 연구[11], 그리고 다중 네트워크모델 처리를 위한 경량 인공지능 하드웨어 기반 통합프레임워크[12] 등도 연구되어 있다.



그림 4. 차량번호판의 삭제, 원본, 모자이크 이미지 예시  
Fig. 4 Number plate in deleted, original and mosaic

### III. 멀티모달 마이데이터 유통시스템의 개념

본 연구에서 정보 수집/가공 주체를 역할별로 정의하면, ‘정보 주체’는 개인정보를 제공하고 제공과 활용을 승인하는 개인을 나타낸다. ‘정보사용자’는 멀티모달 마이데이터를 사용하여 학습하는 AI 기업 또는 AI 연구소를 표현한다. ‘정보관리자’는 정보사용자의 요구에 맞는 멀티모달 마이데이터의 수집/촬영 환경 제공하는 지방자치단체(이하 “지자체”)를 의미한다. ‘마이데이터 포털’은 마이데이터의 투명한 유통/관리를 위해 정부가 운영하는 플랫폼을 표현한다.

그림 5는 멀티모달 마이데이터 유통시스템에서 정보 주체의 역할별 데이터 생성 과정을 표현한다.

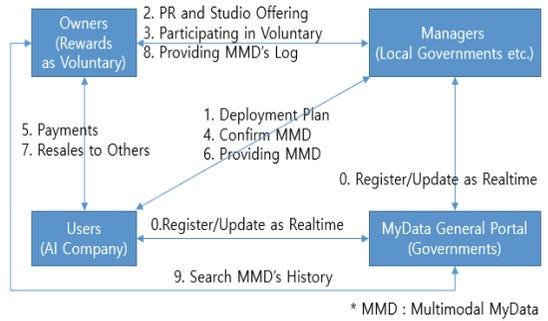
멀티모달 마이데이터 유통시스템의 효과적인 운영을 위해서는 공공기관의 역할이 명확히 구분되어야 한다. 정부는 관리 감독을, 지자체는 구축, 운영, 관리, 홍보를 담당하는 것이 적절하다.

텍스트 기반의 일반적인 마이데이터 제공 방법과 본 연구에서 제안한 방법의 차이는, Text-Based API(Application Programming Interface) 제공뿐 아니라, 정보관리자의 저장소에 보관된 영상을 포함한 멀티모

달 데이터가 유통된다는 점이다. 또한, 마이데이터의 소비자가 정보 주체가 아니라 정보사용자라는 점이다.

이러한 차이로 인해, 정보소비자는 정보 주체와 거래관계가 성립되게 된다.

수집되는 데이터는 그림 6과 같이 일반적으로 획득하기 어려운 발, 발가락 등 신체 부위 사진과 러닝머신 등에서 달리는 자세 동영상, 운동 시 혈압의 변화와 호흡의 소리 등 정보사용자가 요구하는 멀티모달 마이데이터가 수집된다.



\* MMD : Multimodal MyData  
그림 5. 멀티모달 마이데이터의 흐름도  
Fig. 5 Flowchart of multimodal MyData



그림 6. 수집된 마이데이터의 예시 이미지  
Fig. 6 Examples of MyData collected

### IV. 멀티모달 마이데이터 유통시스템의 설계

#### 4.1 물리적 환경

멀티모달 마이데이터의 수집을 위한 장소는, 유동 인구가 많아 자발적 참여 분위기가 조성될 수 있는 장소여야 한다. 서울 도심에서의 경우 강남역, 홍대입구 인근, 각 지하철역 등이 해당한다.

선정된 장소의 환경 구축을 위해서는 그림 7과 같이 이동이 가능한 컨테이너 등을 활용한다.

그림 8은 컨테이너 내부의 촬영실 배치이다. 동시에 정보 주체 수 명이 다중 데이터의 제공이 가능하도록 여러 개의 스튜디오를 갖춘 환경으로 설계한다.



그림 7. 컨테이너로 환경 구현된 예시  
Fig. 7 Examples of environments built with container



그림 8. 컨테이너 내부 배치 예시 이미지  
Fig. 8 Example of internal space allocation

데이터의 수집을 위한 구성은 그림 9와 같다. 영상 촬영을 위해서는, 고속 촬영, 여러 방향에서 세밀한 촬영 등이 가능해야 하며, 조도와 조명을 다양한 부위에 여러 단계로 조절할 수 있도록 설계한다.

음성과 텍스트 등의 수집을 위해서는, 정보사용자와 협의 후 마이크와 여러 가지 센서 등을 동시에 운영할 수 있는 환경을 구현하여야 하며, 특히 시간동기화, 무소음과 무진동 등에 대해서도 설계에 반영한다.

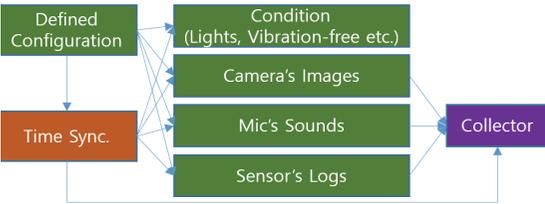


그림 9. 마이데이터 수집 설정  
Fig. 9 Configuration for MyData collection

멀티모달 마이데이터의 보호를 위해, 네트워크 구성은 그림 10과 같이 네 부분으로 세분화하여 연결한다.

촬영부에서는 정보 주체가 정보사용자가 요구하는 촬영을 수행한다.

확인부에서는 WIFI와 앱(App)을 이용하여, 정보 주체가 본인의 멀티모달 마이데이터의 등록을 결정한다.

수집된 마이데이터는 클라우드(Cloud)로 이관 후 보관되며, 정보 주체는 원격지에서 멀티모달 마이데이터의 판매/재판매를 진행하게 된다.

이러한 네트워크 구성에서 정보보호를 위해 IEEE

802.1x 인증 기반의 방화벽과 DDoS, VPN, IDS/IPS를 통합한 UTM( Unified Threat Management) 시스템과 AI 기반 이상 행위 탐지시스템 등을 운영하여야 하며, 실시간 모니터링 및 사고 대응 체계를 설계에 반영한다.

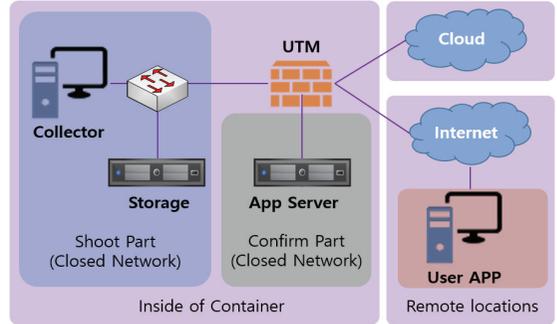
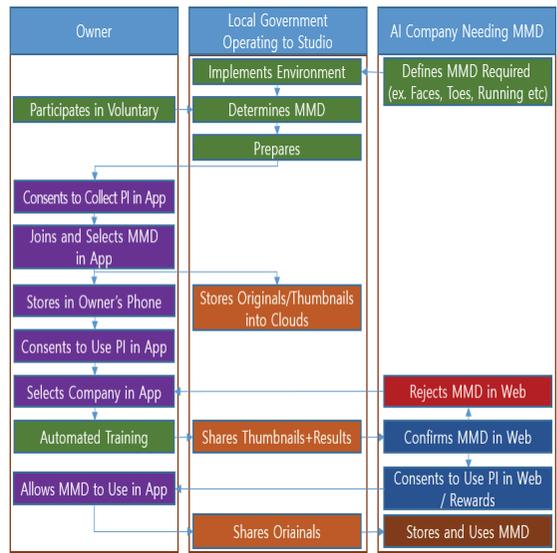


그림 10. 물리적 연결 설정  
Fig. 10 Physical configuration

#### 4.2 논리적 데이터 흐름



\* MMD : Multimodal MyData

\* PI : Personal Information

그림 11. 마이데이터 처음 생성 후 판매 절차  
Fig. 11 Flowchart for making first sale

멀티모달 마이데이터 유통플랫폼은 확장성을 고려한 HW-SW 링크 시스템으로 설계한다. 그림 11은 플랫폼에서 멀티모달 마이데이터의 수집 및 판매하는 단계별 절차이다.

정보사용자의 데이터 무단 사용을 방지하기 위해,

정보사용자는 정보 주체가 마이데이터 유통 App에서 전송하는 썸네일(Thumbnails) 이미지와 사전 제공된 모델로 학습한 결과를 조회하여 데이터 품질을 확인하고 거래하도록 설계하며, 정보 주체는 정부의 마이데이터 포털을 통해 정보사용자의 데이터 사용 상태를 정기적으로 확인하도록 설계한다.

정보 주체가 멀티모달 마이데이터 유통시스템에 수집 보관되고 있는 데이터를 다른 정보사용자에게 판매하고 싶다면, 이때는 클라우드에 보관 중인 본인의 멀티모달 마이데이터를 그림 12와 같은 절차로 재판매한다.

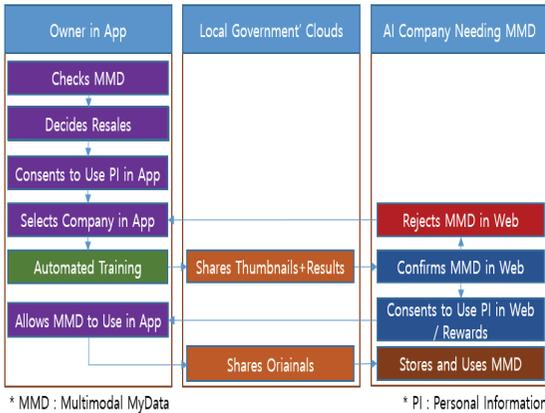


그림 12. 마이데이터 재 판매를 위한 절차  
Fig. 12 Flowchart for reselling

### 4.3 App을 활용한 UX/UI 구현

멀티모달 마이데이터 유통플랫폼 정보 주체는 마이데이터 유통 동의와 모든 거래 기록 및 거래 상태를 직관적이고 단순한 UX/UI의 App으로 관리한다. 이때 필요한 기능을 그림 13과 같이 정의할 수 있다. 또한, 정보관리자도 App을 이용해 마이데이터 유통 상황을 조회할 수 있도록 설계한다.

그림 14는 멀티모달 마이데이터 시스템이 정부의 마이데이터 종합 포털과 연결을 통해 취득하는 정보의 종류의 예를 표현한다.

앱을 통해, 클라우드에 업로드된 멀티모달 마이데이터의 범주/유형 등의 조회, 데이터를 사용 중인 기업 정보와 상태 정보, 데이터의 수집 장소 정보와 수집 중인 데이터의 소개 정보, 현재까지 데이터 유통플랫폼에 참여한 정보 주체의 통계 등의 정보를 실시간으로 조회할 할 수 있어야 한다.

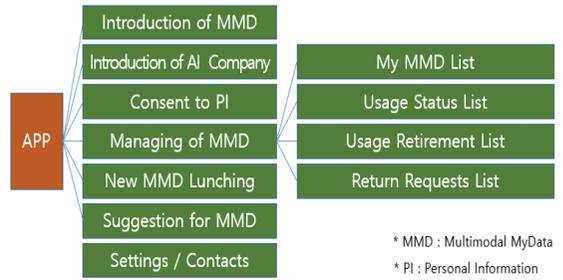


그림 13. 마이데이터 운영 앱의 기능 정의  
Fig. 13 Defining app features

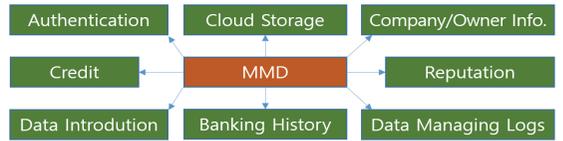


그림 14. 마이데이터 플랫폼에 연결되는 정보  
Fig. 14 Connected information in multimodal platform

### 4.4 취득 가능 데이터의 다양화

보편적인 데이터 학습을 통해, 특정 도메인에 적합한 높은 정확도를 갖는 학습 결과를 도출하기는 어렵다. 도메인 데이터를 이용한 도메인 내의 학습이 신뢰도를 높일 수 있다. 따라서, 수집되는 데이터는 섬세한 요구 기반의 다양한 데이터가 취득되어야 한다.

영상의 경우, 각도의 다양성과 인위적 촬영 여부가 학습에 있어서 중요한 변위 요인이 될 수 있다. 예를 들어, 웃는 얼굴 촬영을 위해서는 코미디 프로그램을 재생하면서, 정보 주체가 웃는 시점에 데이터의 수집을 진행하는 등 수집 방식이 사전에 확정되어야 한다.

### 4.5 사전 준비된 AI 네트워크모델로의 검증

요구사항 정의 단계에서, 정보사용자는 멀티모달 마이데이터의 검증모델을 정의하고, 자동 검증 도구를 정보관리자에게 제공한다. 정보 주체의 촬영 직후, 해당 데이터는 자동 검증 도구에 의해 검증되고 수집데이터는 결과와 함께 저장된다.

정보사용자는 썸네일(Thumbnails) 이미지와 함께, 모델의 학습 결과를 확인하고 정보 주체와 거래한다.

### 4.6 클라우드 기반의 보관/유통/폐기 관리

신뢰할 수 있는 유통플랫폼의 운영을 위해, 정보사용자의 마이데이터에 대한 투명한 관리체계가 공공기관에 의해 확보되어야 한다.

디지털 파일의 특성상 정보 주체가 제공한 데이터는 학습을 위해서만 사용되어야 하고, 학습 이후에는 동의 없이 배포 불가능하도록 DRM(Digital Rights Management) 등을 적용하여 관리되어야 하며, 정보 주체가 요청 시 해당 파일은 시스템 내에서 완벽하게 파괴되어야 한다. 이에 대한 방법과 책임은 지자체에 있다.

사용이 허가된 데이터는 정보사용자의 학습 방법에 따라 다양한 형태로 사용될 수 있다. 다만, 정보사용자를 제외한 정보사용자의 제공과 인터넷 등을 포함한 정보 공개 및 배포는 불가하도록 계약한다.

정보 주체의 즉각적인 변심을 막기 위해, 데이터 사용 허가 후 정보사용자는 데이터 사용을 일정 시간 동안 보장받고, 이후에는 정보 주체의 의견에 따라 폐기 될 수 있는 방어 장치가 필요하다.

정보 주체는 같은 데이터를 다른 정보사용자에게도 판매할 수 있다. 이를 위해 정보관리자인 지자체는 멀티모달 마이데이터 정보사용자를 위한 전용 페이지를 제공하고, 이곳에서 정보사용자는 요구사항 게시를, 정보 주체는 보유 중인 데이터의 샘플데이터 홍보를 할 수 있게 하여, 데이터의 거래가 이루어질 수 있는 환경을 제공한다. 이들의 거래를 보증하는 역할은 지자체의 몫이다.

## V. 결론

본 논문에서는 멀티모달 마이데이터 유통플랫폼을 구축하여, 정보 주체의 데이터를 데이터가 필요한 AI 기업에, 각종 법률을 준수하며, 학습용 데이터로 제공하는 방안을 연구하였다.

AI 학습용 데이터를 확보하기 위하여 자발적으로 수집/유통할 수 있는 플랫폼을 분석 및 설계하여, 수집 환경 구축, 플랫폼과 App을 이용한 유통, 정보 주체와 정보사용자가 거래하는 절차를 제안하였다.

“AI 학습데이터를 자발적으로 확보하기 위한 멀티모달 마이데이터 유통시스템”이 정부와 지자체 주도로 설치/운영된다면 AI 기술의 발전을 기대할 수 있으며, 세계 최초로 공공주도의 데이터 유통 모델을 구현하여 전 세계에 우리나라에 대한 위상을 드높일 수 있을 것이다.

향후 연구로는, AI 실시간 블록체인 기술을 적용한 마이데이터 유통시스템의 보안 강화 문제, 정보사용자

가 자동으로 취득된 학습데이터에 대한 신뢰도를 검증할 수 있는 경량화된 AI 학습 모델에 관한 연구가 필요하다.

## References

- [1] J. M. Choi, and Y. Jo, “A Study on the Right to Data Portability and MyData Industry,” *J. of Law & Economic Regulation*, vol. 13, no. 2, 2020, pp. 92-107.  
doi.org/10.22732/CeLPU.2020.13.2.92.
- [2] J. Park, “Study on methods for establishing legislation on data protection and distribution,” *Sogang Law J.*, vol. 9, no. 2, 2020, pp. 3-41.  
doi.org/10.35505/slj.2020.06.9.2.3.
- [3] B. R. Choi, and Y. Kim, “A study on data construction for artificial intelligence learning,” In *Proc. Symp. of the Korean Institute of Communications and Information Sciences*, Jeju, Korea, 2023, pp. 337-339.
- [4] C. Subakan, M. Ravanelli, S. Cornell, M. Bronzi, and J. Zhong, “Attention Is All You Need In Speech Separation,” In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (IEEE)*, Toronto, Canada, 2021, pp. 21-25.  
doi.org/10.1109/ICASSP39728.2021.9413901.
- [5] C. C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, and M. Bacchiani, “State-of-the-Art Speech Recognition with Sequence-to-Sequence Models,” In *Proc. 2018 Int. Conf. on Acoustics, Speech and Signal Processing(IEEE)*, Calgary, Canada, 2018, pp. 4774-4778.  
https://doi.org/10.1109/ICASSP.2018.8462105.
- [6] H. Jung, N. Lee, S. Seol, and K.-S. Han, “A Study on the Prediction Accuracy of Machine Learning using De-Identified Personal Information,” *J. of Korean Institute of Information Scientists and Engineers*, vol. 47, no. 10, 2020, pp. 906-910.

doi.org/10.5626/JOK.2020.47.10.906.

- [7] J.-S. Park, "Optimization of Deep Learning Model Based on Genetic Algorithm for Facial Expression Recognition," *J. of the KIECS*, vol. 15, no. 1, Feb. 2020, pp. 85-92.  
doi.org/10.13067/JKIECS.2020.15.1.85.
- [8] J.-S. Park, M. Wiranegara, and G.-Y. Son, "Multi-channel Video Analysis Based on Deep Learning for Video Surveillance," *J. of the KIECS*, vol. 13, no. 6, Dec. 2018, pp. 1263-1268.  
doi.org/10.13067/JKIECS.2018.13.6.1263.
- [9] D. H. Lim, and D. Park, "The design and operation of artificial intelligence image analysis support center using social safety identification Information," Master's Thesis, *Hoseo Graduate School of Venture*, 2022.
- [10] G.-E. Park, C. U. Hwang, S. R. Lim, and H. S. Jang, "CNN-LSTM based Autonomous Driving Technology," *J. of the KIECS*, vol. 18, no. 6, 2023, pp. 1259-1268.  
doi.org/10.13067/JKIECS.2023.18.6.1259.
- [11] G.-H. Seok, and S.-H. Lee, "Research on object detection library utilizing spatial mapping function between stream data in 3D data-based area," *J. of the KIECS*, vol. 19, no. 03, 2024, pp. 551-562.  
doi.org/10.13067/JKIECS.2024.19.3.551.
- [12] S.-H. Jeon, J.-H. Lee, J.-S. Han, and B.-S. Kim, "A Study of Unified Framework with Light Weight Artificial Intelligence Hardware for Broad range of Applications," *J. of the KIECS*, vol. 14, no. 5, Oct. 2019, pp. 969-976.  
doi.org/10.13067/JKIECS.2019.14.5.969.

## 저자 소개



### 임동현 (Dong Hyun Lim)

2021년 호서대학교 벤처대학원 융합공학과 석사 졸업(공학석사)  
2024년 호서대학교 벤처대학원 융합공학과 재학(공학박사)

※ 관심분야 : 무선공학, 통신공학, 스마트시티, CCTV, 영상분석, 도시통합정보운영



### 박대우 (Dea-Woo Park)

1998년 숭실대학교 컴퓨터학과 (공학석사)  
2004년 숭실대학교 컴퓨터학과 (공학박사)

2004년 숭실대학교 겸임교수  
2006년 정보보호진흥원(KISA) 선임연구원  
2007년~현재 호서대학교 벤처대학원 교수  
※ 관심분야 : 인공지능, AI로봇, Hacking, CERT/CC, Forensic, 사이버안보, 네트워크 보안