

# 2024년 이브와 ICT멘토링 프로젝트 결과보고서

2024. 10.

프로젝트명	비정형데이터 가명처리 기술 개발
-------	-------------------

## 요 약 본

프로젝트 정보	
프로젝트명	비정형데이터 가명처리 기술 개발
주제 영역	<input type="checkbox"/> 생활 <input type="checkbox"/> 업무 <input checked="" type="checkbox"/> 공공/교통 <input type="checkbox"/> 금융/핀테크 <input type="checkbox"/> 의료 <input type="checkbox"/> 교육 <input type="checkbox"/> 유통/쇼핑 <input type="checkbox"/> 엔터테인먼트
기술 분야	<input checked="" type="checkbox"/> SW-AI <input type="checkbox"/> 방송·콘텐츠 <input type="checkbox"/> 블록체인·융합 <input type="checkbox"/> 디바이스 <input type="checkbox"/> 차세대보안 <input type="checkbox"/> 미래통신·전파
달성 성과	<input checked="" type="checkbox"/> 논문게재 및 포스터 발표 <input type="checkbox"/> 앱등록 <input type="checkbox"/> 프로그램등록 <input type="checkbox"/> 특허 <input type="checkbox"/> 기술이전 <input type="checkbox"/> 실용화 <input checked="" type="checkbox"/> 공모전( 이브와 멘토링 공모전 ) <input type="checkbox"/> 기타( )
프로젝트 소개	자율주행 자동차 개발에 필요한 교통 데이터의 개인정보를 자동으로 비식별화 처리하여 안전하게 활용할 수 있도록 하는 웹 기반 플랫폼이다. 교통 분야의 AI 학습에 사용되는 데이터를 안전하게 처리하고 편리하게 접근해 사용할 수 있도록 하는 것이 프로젝트의 목표이다.
개발 배경 및 필요성	자율주행 기술 발전으로 도로 주행 영상 데이터 활용이 중요해졌으나 개인정보 보호법과 상충된다. 교통 관련 다양한 분야에서 활용하기 위해 개인 식별을 방지하는 비식별 기술 개발이 필요하다.
프로젝트 특·장점	데이터 3법 시행으로 개인정보 가명처리 후 기업 활용이 가능해짐에 따라 교통 분야 AI 학습 데이터의 안전한 활용을 위한 비식별 처리 플랫폼을 개발한다. 객체 인식 AI 모델로 개인정보를 자동 탐지 및 비식별 처리하여 시간과 비용을 절감하고, 개인정보 보호와 데이터 활용성을 동시에 확보한다. 사용자 친화적인 웹 인터페이스로 누구나 쉽게 접근 및 활용할 수 있도록 한다.
주요 기능	교통 데이터의 개인정보를 불러, 마스킹 등으로 자동 비식별 처리하는 웹 기반 플랫폼을 제공한다. 플랫폼 내에서 비식별화된 데이터의 시각화, 공유, 조회 등 관리 기능을 지원한다.
기대효과 및 활용 분야	민감 정보 재가공으로 개인정보를 보호하면서 법적 제약 없는 대량의 AI 학습용 데이터 확보가 가능해진다. 접근성 높은 웹 플랫폼을 통해 개발자, 학생 등 다양한 사용자의 활용도 및 자율주행 기술 신뢰도를 향상시키고, 테슬라, 우버 등 자율주행 시장의 주요 기업들이 안전하게 데이터를 활용할 수 있는 기반을 마련한다.

# 본 문

## I. 프로젝트 개요

### 1. 프로젝트 소개

#### 1) 기획의도

- 자율주행자동차용 도로 주행 영상 등 교통분야 학습 데이터의 개인정보 보호를 위한 비식별화 플랫폼을 개발하여 교통분야의 AI학습에 사용되는 데이터를 안전하게 처리하고자 한다.
- 비정형데이터 가명처리 기준에 맞추어 마스킹, 블러링이나 모자이크 등의 기술을 제공한다. 해당 프로젝트는 개인정보 보호를 통해 교통 AI 산업에 혁신을 가져올 것으로 기대된다.

#### 2) 프로젝트 내용

- 사용자가 웹 플랫폼에 이미지나 영상을 업로드한다.
- YOLO와 CV2를 활용한 얼굴과 번호판을 자동 인식 및 마스킹할 수 있다.
- 사용자가 업로드한 데이터에 태그, 크기, 분류 등 메타데이터를 자유롭게 추가하고 공유할 수 있는 기능을 제공한다.

#### 3) 프로젝트 정의

##### 가) 주요목표

- 본 프로젝트는 자율주행 자동차 개발에 필요한 교통 데이터의 개인정보를 자동으로 비식별화 처리하는 웹 기반 플랫폼 개발을 목표로 한다.

##### 나) 비식별 처리 방식

- 비식별 처리는 개인을 식별할 수 있는 요소를 제거하거나 대체하여 개인정보를 보호하는 기술적 방법으로, 주로 가명처리, 익명처리, 총계처리 등이 사용된다.
- 본 프로젝트는 객체 인식 알고리즘을 통해 개인 식별 정보를 탐지하고, 블러 처리 및 마스킹 등의 비식별화 기술을 적용한다.

##### 다) 핵심 가치

- 이를 통해 자율주행 자동차 개발에 필요한 데이터를 안전하게 활용하며, 개인정보 보호와 데이터 활용의 균형을 추구한다.

## 2. 개발 배경 및 필요성

### 1) 프로젝트 제작 동기

- 자율주행 기술의 상용화와 발전에 따른 개인정보 보호의 중요성 증대
- 기존 자동 비식별처리 프로그램의 한계 (완벽한 마스킹 불가, 수동 확인 필요)와 저화질 영상에서의 식별 어려움 극복 필요성

### 2) 목적

- **안전한 데이터 공유와 프라이버시 보호** : 개인정보 보호와 데이터 활용의 균형을 추구하고자 한다.
- **연구 및 개발 지원, 데이터 생태계 확장**
- **비즈니스 기회 창출**: 비식별화된 데이터를 활용한 새로운 제품 및 서비스 개발 지원, 데이터 기반 비즈니스 모델 강화

## 3. 프로젝트 특·장점

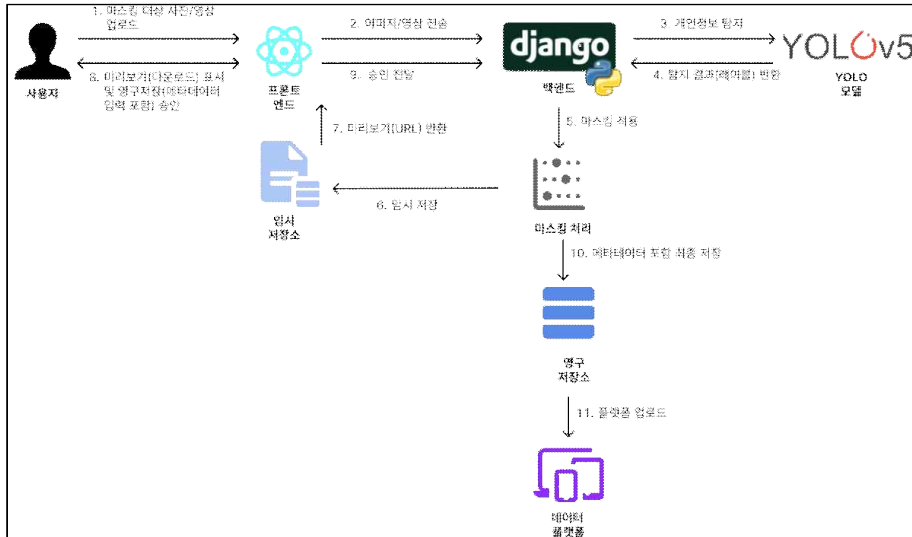
### 1) 기존 제품, 유사 제품과의 차별성:

- 다양한 마스킹 옵션: 블러링, 단색 마스킹, 모자이크 등 다양한 마스킹 방법을 제공하여 사용자의 요구에 맞는 유연한 개인정보 보호 기능 제공
- 자동 개인정보 탐지: 딥러닝 기반의 객체 인식 알고리즘을 사용하여 얼굴, 번호판 등을 자동으로 탐지하고 마스킹 처리
- 메타데이터 관리: 마스킹된 파일의 메타데이터를 자동으로 추출하고 편집할 수 있는 기능 제공(데이터의 태그 및 분류 기능을 통해 체계적인 데이터 관리 가능)
- 내부 플랫폼 연동: 처리된 파일을 바로 내부 플랫폼에 업로드할 수 있는 기능으로 워크플로우 최적화, 데이터 수요자는 내부 플랫폼에서 데이터 다운로드 및 활용 가능
- pdf 암호화 기능 : 마스킹 처리 pdf 등의 pdf 파일을 업로드하고, 열람 횟수와 비밀번호를 설정할 수 있다. 해당 pdf를 열람할 수 있는 보안 pdf 뷰어를 제공한다.

## II. 프로젝트 내용

### 1. 프로젝트 구성도

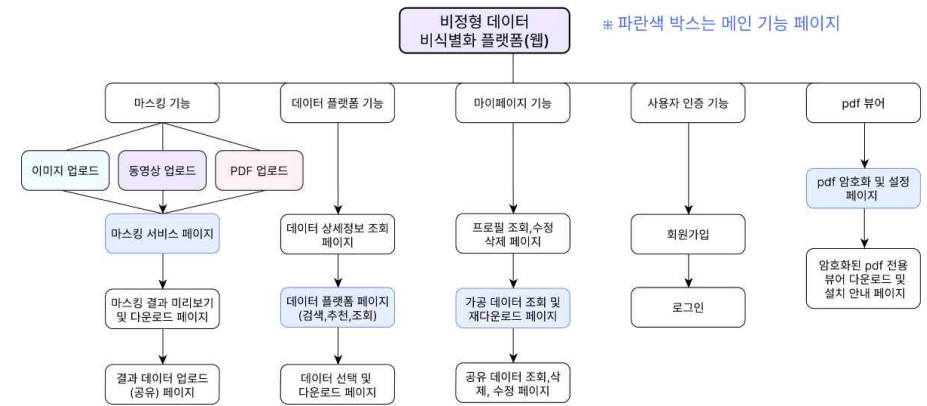
#### 1) 프로젝트 구성도



[개인정보 비식별처리 플랫폼의 프로젝트 구성도]

- 1. 사용자가 마스크 대상 사진/영상을 업로드
- 2. 업로드된 사진/영상이 Django 서버로 전송.
- 3. Django 서버에서 YOLO v5 모델을 사용하여 객체 감지(Object Detection)를 수행
- 4. 감지된 객체의 위치 정보를 Django 서버로 다시 전달.
- 5. Django 서버에서 마스크 처리를 수행.
- 6. 마스크 처리된 사진/영상을 임시 저장소에 저장.
- 7. 마스크 처리된 사진/영상의 URL 정보를 프론트엔드(React)로 전달.
- 8. 프론트엔드(React)에서는 마스크 처리된 사진/영상을 사용자에게 표시.
- 9. 사용자가 마스크 결과를 확인.
- 10. 사용자가 마스크 처리된 사진/영상을 저장하기로 선택하면, 해당 정보를 영구 저장소로 전송.
- 11. 영구 저장소에 마스크 처리된 사진/영상을 저장하고, 해당 데이터의 URL 정보를 프론트엔드로 반환

### 2) 페이지 구성도

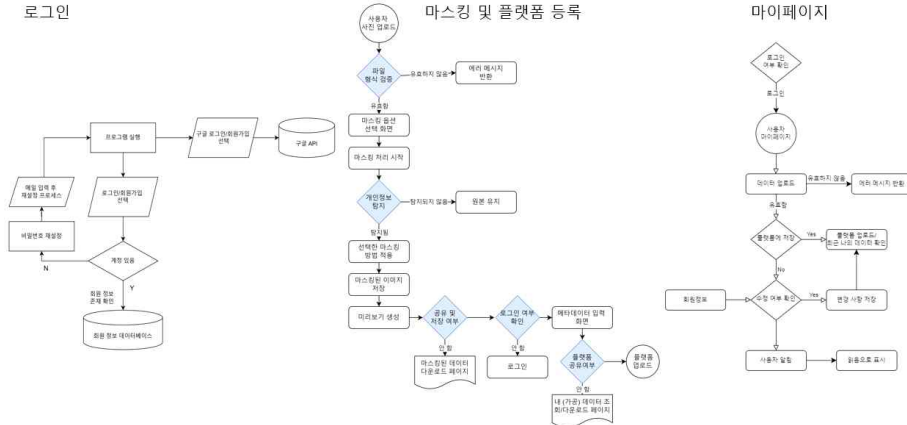


[개인정보 비식별처리 플랫폼의 페이지 구성도]

- 1. 메인 페이지가 존재하며, 헤더에서 마스크 기능, 데이터 플랫폼 기능, 마이페이지 기능, 사용자 인증 기능, pdf 뷰어 기능으로 이동 가능.
- 2. 마스크는 이미지 업로드, 동영상 업로드, pdf 업로드 중 선택하여 진행.
- 3. 마스크 기능에는 마스크 서비스 페이지, 결과 미리보기와 다운로드 페이지, 데이터를 업로드해 공유할 수 있는 페이지 존재.
- 4. 데이터 플랫폼 기능에는 데이터 상세정보 조회 페이지, 검색 · 추천 · 조회 페이지, 데이터 선택 및 다운로드 페이지 존재
- 5. 사용자 인증 기능에는 회원가입, 로그인 기능이 존재
- 6. pdf 뷰어에는 pdf 암호화 및 설정 페이지, 암호화된 pdf 전용 뷰어를 다운로드하고 설치 안내문을 확인할 수 있는 페이지가 존재

### 3) 서비스 흐름도

로그인



[개인정보 비식별처리 플랫폼의 서비스 흐름도]

1. 로그인/회원가입을 선택하고 존재하는 계정을 데이터베이스에서 확인.
2. 파일 업로드 후 파일 형식이 맞으면 개인정보를 탐지해 마스크를 진행.  
마스크된 파일이 저장되고 사용자 다운로드가 가능  
로그인된 상태라면 플랫폼을 이용해 마스크 파일을 공유 가능
3. 로그인 여부를 확인 후 마이페이지에서 자신의 데이터를 확인 가능  
회원정보를 수정하고 사용자 알림을 모아서 확인

## 2. 프로젝트 기능

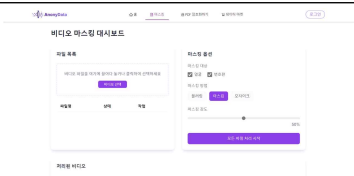
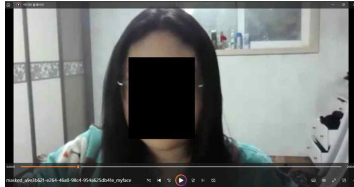

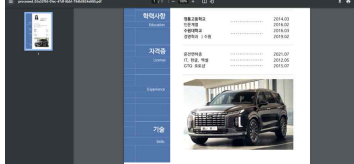


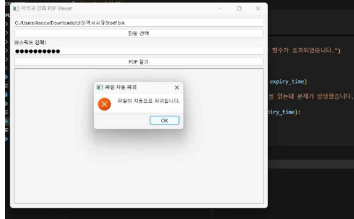
### 1) 전체 기능 목록

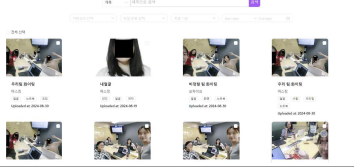
구분	기능	설명
S/W	파일 업로드	드래그 앤 드롭 또는 파일 선택으로 업로드
	자동 개인정보 탐지 및 마스크	yolo 딥러닝 기반 얼굴 및 번호판 탐지, 마스크, 모자이크, 블러링 옵션 제공
	로그인/회원가입	이메일과 비밀번호를 입력하여 로그인과 회원가입 진행
	데이터 플랫폼 검색 기능과 필터링 기능	사용자들이 데이터를 공유하는 데이터 플랫폼에서 원하는 조건으로 데이터를 필터링하거나 검색
	마이페이지	회원정보 수정, 데이터 모아보기
	pdf 뷰어	보안을 위해 비밀번호 제한, 기간 만료 파일 파괴가 설정된 pdf 뷰어를 사용
	회원가입 인증 기능	이메일로 가입 시 인증하는 로직 구현

H/W	해당없음	해당없음
-----	------	------

### 2) S/W 주요 기능

기능	설명	프로젝트 실물사진
로그인	가입한 사용자는 이메일과 비밀번호를 이용해 로그인하거나 Google 메일로 로그인할 수 있다.	
회원가입	사용자는 이메일과 비밀번호 중복 체크로 회원가입 가능하다.	
파일 업로드	사용자 친화적 인터페이스로 쉽게 파일을 업로드한다	
이미지 마스크 처리	사용자가 마스크 옵션을 선택하고 해당 페이지 내에서 마스크된 데이터를 반환받는다.	
마스크된 이미지 확인	모든 파일을 일괄 다운로드하거나 하단에서 마스크된 이미지를 미리 볼 수 있다.	
마스크된 이미지 다운로드	마스크된 이미지를 다운로드한다.	

비디오 마스크 처리	비디오 마스크 대시보드에서 비디오를 업로드하고 마스크 옵션과 강도를 설정할 수 있다.	
마스크된 동영상 다운로드	마스크된 동영상을 다운로드한다.	
PDF 암호화 기능 제공	PDF 파일을 선택하여 열람 가능 횟수와 pdf 오픈 암호를 지정하고, 암호화한다.	
마스크된 pdf 다운로드	pdf 내 이미지가 마스크되어 제공된다. ( 해당 화면의 pdf에서 차량 번호판이 모자이크된 것을 확인할 수 있다.)	
보안 PDF 뷰어 다운로드와 안내 가이드	PDF를 안전하게 다운로드하고 암호와 세션 만료 파괴 설정이 가능한 pdf 뷰어를 다운로드해 사용할 수 있는 안내 페이지이다.	
PDF 뷰어 비밀번호 입력	PDF 뷰어로 파일을 열기 위해 설정한 비밀번호를 입력해야 한다.	
PDF 뷰어 비밀번호 3회이상 오류, 기간 만료 시 자동 파괴	PDF 뷰어에서 입력한 비밀번호가 3회 이상 오류가 발생하거나, 기간이 만료될 경우 자동으로 파일이 파괴된다.	

데이터 플랫폼	데이터 플랫폼에서 사용자들이 데이터를 업로드해 서로 공유할 수 있다.	
데이터 플랫폼 검색플랫폼 기능	- 데이터 플랫폼에서는 카테고리 선택, 파일 유형 태그 등 다양한 옵션의 필터링 기능이 제공  - 검색 기능으로, 원하는 키워드의 데이터를 검색할 수 있다.	
마이페이지	마이페이지에서 회원 정보 수정, 나의 업로드 데이터 수와 최근 나의 데이터 정보를 확인할 수 있다.	

### 3) H/W 주요 기능

기능/부품	설명	프로젝트 실물사진
해당없음	해당없음	해당없음

## 3. 주요 적용 기술

### 1) 딥러닝 기반 객체 인식

- 적용 알고리즘: YOLO v5 (You Only Look Once version 5)
- 이론적 배경: YOLO는 단일 신경망을 사용하여 이미지를 그리드로 나누고, 각 그리드 셀에서 객체의 경계 상자와 클래스 확률을 동시에 예측. 이는 기존의 R-CNN 계열 알고리즘보다 빠른 처리 속도를 제공
- 구현 시나리오:
  - 사전 학습된 YOLO v5 모델을 개인정보(얼굴, 번호판, 신분증 등) 데이터셋으로 전이학습
  - 데이터 증강 기법 적용: 회전, 반전, 노이즈 추가 등으로 모델의 견고성 향상
  - 비정형 데이터 입력 시 자동으로 객체 탐지 수행
  - 탐지된 객체의 좌표와 신뢰도 점수를 마스크 모듈로 전달

## 2) 마스크 알고리즘

- 적용 기법: **가우시안 블러, 평균 필터, 모자이크, 경계 보존 블러링**
- 이론적 배경:
  - a) 가우시안 블러: 2D 가우시안 분포를 따르는 커널을 이용한 컨볼루션 연산
  - b) 평균 필터: 주변 픽셀들의 평균값으로 대체하는 간단한 블러링 기법
  - c) 모자이크: 특정 영역을 저해상도로 다운샘플링 후 업샘플링하여 픽셀      화
  - d) 경계 보존 블러링: 엣지 정보를 유지하면서 블러링 처리  
(예: 바이래터럴 필터)
- 구현 시나리오:
  - a) 객체 인식 모듈에서 전달받은 **좌표를 기반으로 마스크 영역 설정**
  - b) 사용자가 선택한 **마스크 방법과 강도에 따라 적절한 알고리즘 적용**
  - c) **실시간 미리보기 제공**을 위해 저해상도 버전으로 먼저 처리 후 표시
  - d) 최종 처리 시 **원본 해상도로 마스크 적용**

## 3) RESTFUL API 설계

- 적용 기술: Django Rest Framework, JWT (JSON Web Tokens)
- 이론적 배경: REST(Representational State Transfer)는 웹 서비스 디자인을 위한 아키텍처 스타일로, 리소스 중심의 균일한 인터페이스를 제공
- 구현 시나리오:
  - a) 리소스 기반 URL 설계 (예: /api/v1/files/, /api/v1/masks/)
  - b) HTTP 메서드를 통한 CRUD 작업 구현 (GET, POST, PUT, DELETE)
  - c) JWT를 이용한 사용자 인증 및 권한 관리

## 4) 비동기 처리

- 적용 기술: Celery, RabbitMQ
- 이론적 배경: 비동기 처리는 장시간 실행되는 작업을 백그라운드에서 처리하여 사용자 경험을 향상시키고 서버 리소스를 효율적으로 사용할 수 있게 함
- 구현 시나리오:
  - a) Celery 워커 프로세스 구성 및 RabbitMQ를 메시지 브로커로 설정
  - b) 마스크 작업을 Celery 태스크로 정의하여 비동기적으로 실행
  - c) 작업 진행 상황을 실시간으로 프론트엔드에 전달 (WebSocket 사용)
  - d) 작업 재시도 메커니즘 구현 (exponential backoff 적용)
  - e) 분산 처리를 위한 Celery 워커 클러스터 구성

## 5) Fernet 암호화 알고리즘

- 적용 기술: Fernet 대칭 키 암호화 사용
- 이론적 배경: Fernet 암호화 알고리즘은 AES128을 CBC 모드와 PKCS7 패딩으로 구현한 대칭 키 암호화 방식입니다. 동일한 키로 데이터를 암호화 및 복호화하며, 무작위 생성된 키를 사용해 보안성을 강화합니다. 또한, 만료 시간과 최대 조회 횟수를 설정하여 암호화된 데이터의 보호를 더욱 철저히 합니다.
- 구현 시나리오:
  - a) Fernet 라이브러리를 사용해 무작위로 생성된 대칭 키를 생성하고, 이 키로 Fernet 객체를 초기화함
  - b) PDF 파일의 내용을 읽어들이 Fernet 객체를 사용해 데이터를 암호화
  - c) 만료 시간, 최대 조회 횟수, 비밀 키 등을 포함한 메타데이터를 생성하고, 이를 Base64로 인코딩하여 보호.
  - d) 암호화된 메타데이터와 암호화된 PDF 내용을 결합해 최종 암호화 데이터 생성
  - e) 최종 암호화된 데이터를 .bin 확장자로 파일에 저장하여 보호하고 필요시 Fernet키를 이용해 복호화함

## 4. 프로젝트 개발 환경

구분		상세내용
S/W 개발환경	OS	Ubuntu 20.04 LTS
	개발환경(IDE)	Visual Studio Code, PyCharm
	개발도구	Git, Docker
	개발언어	Python, JavaScript (React)
	기타사항	Redis, Nginx
H/W 구성장비	디바이스	해당없음
	센서	해당없음
	통신	해당없음
	언어	해당없음
	기타사항	해당없음
프로젝트 관리환경	형상관리	Git
	의사소통관리	Notion, Jira
	기타사항	해당없음

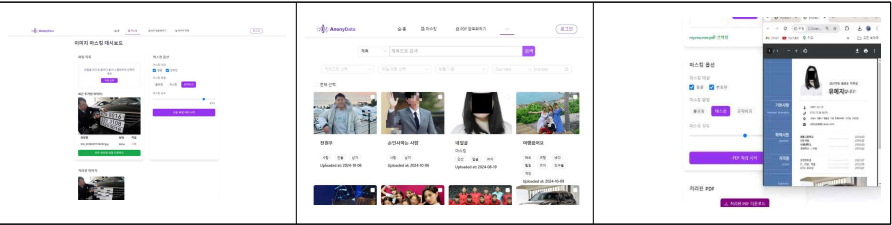
### 5. 장비(기자재/재료) 활용

번호	품명	작품에서의 주요 기능
1	AWS	- MySQL 구동 서버
2		-
3		-
4		-

### 6. 프로젝트 작동 동영상

- <https://youtu.be/pLjzwsbixtY>
- 

### 7. 결과물 상세 이미지



### 8. 달성 성과

■ 논문게재 및 포스터발표	게재(발표)자명	논문(포스터)명	게재(발표)처	게재(발표)일자
	유예지	인간 및 AI 실험 기반 비식별화 기법 성능 평가	ACK 2024	2024.11.01.
□ 앱(APP) 등록	등록자명	앱(APP)명	등록처	등록일자
				2024. 00. 00.
□ 프로그램 등록	등록자명	프로그램명	등록처	등록일자
				2024. 00. 00.
□ 특허/실용신안 출원	출원자명	특허/실용신안명	출원번호	출원일자
				2024. 00. 00.
□ 기술이전	기술이전기업명	기술명	금액	이전일자
				2024. 00. 00.
■ 공모전(대회)	구분(교내/대외)	공모전명	수상여부(출품/수상)	상격
	교내	융합보안창의소프트웨어경진대회	수상	대상
□ 실용화				
□ 기타				

## III. 프로젝트 수행 내용

### 1. 업무 분장

번호	성명	역할	담당업무
1	양필규	멘 토	- 전반적인 프로젝트 진행 상황 멘토링.
2	최현우	지도교수	- 논문지도
3	유예지	팀 장	- 프론트엔드 - PM
4	이지연	팀 원2	- 백엔드 - 서기
5	양소윤	팀 원3	- 프론트엔드 - 문서관리
6	이정연	팀 원4	- 백엔드 - 문서관리

### 2. 프로젝트 수행일정



구분	추진내용	수행 일정							
		3월	4월	5월	6월	7월	8월	9월	10월
계획	프로젝트 제반사항 정립 및 계획 구체화								
분석	비정형 데이터, 비식별 처리에 대한 논문과 기술 분석								
데이터셋 구축	데이터 수집 및 라벨링을 통한 데이터셋 구축								
설계	교통 분야 비정형데이터 처리에 적합한 모델과 라이브러리 선택								
	도커 기반 웹 플랫폼 설계								
개발 환경 구축	작업 환경 및 데이터베이스 구축								
개발	모바일 및 와이드모니터 포함 UX/UI 기준 및 Basic 레이아웃 세팅								
	Django&React 기반 및 웹 플랫폼 설계 및 클라우드 DB설정								
	YOLO 모델 강화학습 및 마스킹 알고리즘 고도화								
테스트	데모 제작 및 피드백								
종료	개발 종료 및 공모전 발표 준비								
온-오프라인 미팅									

### 3. 프로젝트 추진 과정에서의 문제점 및 해결방안

#### 1) 프로젝트 관리 측면

##### - 비대면/대면 혼합 회의에서 전면 대면 회의로의 전환

프로젝트 초기에는 팀원들의 상황에 따라 일부는 대면으로, 일부는 비대면으로 참여하거나 전원이 비대면으로 정기회의를 진행했다. 그러나 이러한 방식으로는 효과적인 소통과 협업이 이루어지지 않아 프로젝트 진행에 어려움을 겪었다. 특히 비대면으로 참여하는 경우, 집에서 회의에 참여하다 보니 집중력이 떨어지고 적극적인 참여가 어려워지는 문제가 발생했다.

이러한 문제를 해결하기 위해 팀원들과 논의 끝에 일주일에 한 번씩 진행하던 화상회의를 전면 대면회의로 전환하기로 결정했다. 이후 매주 프로젝트의 방향을 조율하고, 문제점을 신속히 파악하여 해결책을 모색하게 되는 이점이 생겼다. 또한

모든 팀원이 한 자리에 모여 회의를 진행함으로써 팀원들 간의 커뮤니케이션이 크게 개선되었고, 프로젝트의 진행 속도와 질 또한 향상되었다

##### - 효율적인 역할 분담 체계 구축

프로젝트 초기에는 팀장이 대부분의 관리 업무를 담당하면서 과부하가 걸리고 효율성이 떨어지는 문제가 발생했다. 이러한 어려움을 해결하기 위해 팀원들의 강점을 고려한 역할 분담을 실시했다. 구체적으로 각자의 강점을 살릴 수 있도록 기술 부문 1명, 서기 부문 2명, 일정관리 부문 1명으로 나누었다. 이렇게 각 팀원의 역할과 책임을 명확히 정의하고, 이를 문서화하여 notion에 공유함으로써 작업 중복을 줄이고 각자가 맡은 역할에 대한 책임감을 갖게 되었다.

##### - 프로젝트 관리 도구의 고도화

프로젝트 초기에는 Notion과 KakaoTalk을 사용하여 일정과 커뮤니케이션을 관리했으나, 보다 체계적인 프로젝트 관리의 필요성을 느껴 Jira를 도입하기로 결정했다. Jira를 통해 일정과 작업 진척도를 더욱 효과적으로 관리할 수 있게 되었다. 주간 목표를 설정하고, 이를 달성하기 위해 필요한 자원과 시간 배분을 명확히 할 수 있게 되었다. 또한 Jira의 다양한 기능을 활용하여 작업의 우선순위 설정, 일정 지연 시 알림 기능 등을 통해 보다 체계적인 프로젝트 관리가 가능해졌다.

##### - 팀원 특성을 고려한 개발 방법론 조정

개발 방법론 측면에서, 초기에는 명확한 단계 구조를 가진 워터폴 방식을 적용하려 했으나, 팀원 중 코딩 경험이 부족한 팀원이 있다는 것을 고려하여 접근 방식을 변경했다. 매 회의마다 프로젝트의 목적과 필요성을 상세히 설명하고, 개발 목표를 아주 작은 단위로 나누어 팀원들의 이해도를 높이고 실행 가능한 태스크로 만들도록 하였다. 이를 통해 팀원들의 프로젝트 이해도가 높아지고, 더 주도적인 참여를 이끌어낼 수 있었다.

##### - 기술 격차 해소를 위한 노력

개발을 진행하며, 해당 언어나 프레임워크를 처음 다루는 팀원들이 있다 보니 개발 진행 속도가 빠르지 않아 프로젝트 일정이 점점 밀리게 되는 경우가 발생했다. 따로 작업해 오는 것보다는 최대한 대면으로 진행하며 질의응답과 오류 해결을 진행하는 것이 효율적이라고 생각하여 최대한 대면으로 모여 개발을 진행하였고, 기술 블로그 작성으로 기술과 프로젝트에 대한 이해도를 높이기 위해 노력했다.

- 대면 회의의 도입, 역할 분담 명확화, Jira를 사용한 일정 관리 강화, 그리고 팀원들의 특성에 맞춘 개발 방식 채택을 통해 우리 팀 내 의사소통이 원활해지고



프로젝트 진행 속도가 가속될 수 있었다. 팀 내 역할 분담이 적절하게 할당됨으로서 개발과 프로젝트 매니징에 더 집중할 수 있게 되어 전체적인 프로젝트의 효율성이 크게 향상되었다. 이러한 과정을 통해 팀워크 또한 크게 향상되는 긍정적인 결과를 얻었다.

## 2) 프로젝트 개발 측면

- 프로젝트 진행 중 우리는 얼굴 인식과 차량 인식에서 아쉬웠던 부분을 발견했다. 얼굴 인식의 경우, 정면 얼굴에 대해서는 높은 정확도를 보였으나 옆모습이나 기울어진 측면 얼굴에 대해서는 인식률이 현저히 떨어지는 문제가 있었다. 차량 인식에서도 유사한 문제가 발생했는데, 가까이 있는 차량은 잘 인식하고 마스킹했지만 멀어지는 차량에 대해서는 마스킹이 사라지는 현상이 나타났다. 우리 팀이 초기 모델에 학습시킨 데이터셋이 주로 정면 얼굴이나 가까운 거리의 차량 이미지로 구성되어 있었다. 때문에, 초기 모델이 학습 데이터셋에 너무 맞춰져 있어 새로운 상황이나 다양한 각도, 거리의 데이터에 대해 일반화 능력이 부족한 과적합이 일어났다고 판단했다. 이 문제를 해결하기 위해 우리는 데이터셋의 다양성을 확대하는 작업을 했다. 다양한 거리와 각도에서 촬영된 얼굴 이미지와 차량 이미지를 추가로 수집하여 학습 데이터셋을 보강했다. 또한, 데이터 증강(Data Augmentation) 기법을 활용하여 기존 데이터의 회전, 반전, 밝기 조절 등을 통해 더욱 다양한 상황을 학습할 수 있도록 했다.

- 이러한 노력을 통해 모델의 일반화 능력을 향상시키고 과적합 문제를 완화할 수 있었다. 결과적으로 다양한 각도의 얼굴과 다양한 거리의 차량에 대해서도 안정적인 인식이 가능하게 되었고 마스킹이 개선되었다. 이 과정에서 우리는 인공지능 모델 개발에 있어 데이터의 다양성과 균형이 얼마나 중요한지 알게 되었다. 그리고 과적합 문제를 해결하기 위해 휴리스틱한 접근을 통해 결과물을 지속적인 모니터링 개선의 필요성을 깊이 이해하게 되었고 현재도 모델 성능을 더 개선 시키기 위해 , Google Colab의 T4 GPU를 활용하여 학습 시간을 단축시키는 등의 노력을 계속하고 있다.

- 프로젝트 개발 과정에서 우리는 '함께 성장해야 한다'는 중요한 교훈을 얻었다. 초기에는 경험이 풍부한 팀원이 다른 팀원들에게 강의식, 주입식으로 가르치려고 했지만, 이 방식이 팀원들의 자발적 학습과 성장을 저해한다는 것을 깨달았다. 이에 각 팀원에게 특정 기능과 개발 도메인(백엔드, 프론트엔드, 인공지능)을 맡기고, 해당 파트를 구현하면서 다른 팀원들과 기술 공유를 할 수 있는 수준으로 공부해 오는 것을 목표로 했다. 이 접근 방식은 팀원들의 자발적 학습과 성장을 촉진했고, 전체 팀의 역량을 향상시키는 데 크게 기여했다.

- 개발 과정에서 가장 빛을 발했던 것은 디버깅이었다. AI를 활용한 코드 생성으로 기능 구현은 상대적으로 쉬웠지만, Postman 등을 이용한 테스트, 서버 배포, 버그 수정 등은 혼자 해결하기 어려운 과제였다. 이를 해결하기 위해 우리는 정규 회의일이 아니어도 매일 학교에 모여 함께 코딩하고 이슈를 해결하는 방식을 채택했다. 이러한 페어 프로그래밍 방식은 팀원들의 빠른 성장과 문제 해결 능력 향상에 많은 도움이 되었고, 팀원들의 성장을 이끌어낼 수 있었다.

- 개발 과정에서 직면한 주요 문제 중 하나는 Git 사용에 대한 미숙함이었다. 지속적인 충돌과 실수로 인한 데이터 손실을 방지하기 위해, 우리는 엄격한 Git 사용 규칙을 수립했다. 각자의 브랜치를 생성하여 작업하고, 코딩 전 브랜치 확인, 머지 전 철저한 검토와 테스트 등의 규칙을 도입했다. 이러한 규칙 도입으로 오류가 줄어들고 생산성이 향상되었다.

- 또 다른 도전 과제는 플랫폼 간 빌드 방식의 차이였다. Google Colab과 AWS EC2는 Ubuntu 환경이었지만, 우리의 개발 환경은 Windows였다. 이로 인해 빌드 과정에서 여러 문제가 발생했다. 우리는 지속적인 인터넷 검색과 정보 공유, 그리고 해결 과정을 블로그에 기록하는 방식으로 이 문제를 극복해 나갔다. 이러한 노력으로 문제점을 해결하기 위한 경험과 꾸준함을 기를 수 있었고, 기술적 이해도를 크게 향상시키는 것에 도움이 되었다.

## 4. 프로젝트를 통해 배우거나 느낀 점

- 이번 프로젝트를 통해 우리 팀은 기술적 측면뿐만 아니라 프로젝트 관리와 팀워크 측면에서도 많은 것을 배우고 경험할 수 있었다. 먼저 우리 프로젝트 관리의 중요성에 대해 깊이 이해하게 되었다. 초기에는 의사소통의 개선 여지와 역할 분담의 불명확성으로 인해 어려움을 겪었지만, 정기적인 대면 회의 도입과 체계적인 역할 분담을 통해 이를 극복할 수 있었다. 특히 전면 대면 회의로의 전환은 팀원들의 참여도와 집중도를 높이는 데 큰 도움이 되었고, 이는 프로젝트의 전반적인 질을 향상시키는 결과로 이어졌다.

- 두번째로, 우리 팀은 인공지능 모델 개발 과정에서 데이터의 다양성과 품질이 얼마나 중요한지 깨달았다. 초기에 겪었던 과적합 문제와 특정 상황에서의 인식 한계는 데이터셋의 편향성에서 비롯된 것이었다. 이를 해결하기 위해 다양한 각도와 거리의 데이터를 수집하고 정제하는 과정에서, 인공지능 모델의 성능 향상이 단순히 알고리즘의 개선뿐만 아니라 양질의 다양한 데이터 확보에 크게 의존한다는 점을 체감했다. 이와 관련해서 지속적인 학습과 실험의 중요성도 배웠다. 인공지능 모델의

성능 개선은 단순한 이론적 접근만으로는 불가능하며, 실제 데이터를 통한 반복적인 실험과 분석, 그리고 그 결과를 바탕으로 한 지속적인 개선이 필요하다는 것을 깨달았다. 이 과정에서 팀원들과의 적극적인 의견 교환과 멘토님의 조언이 큰 도움이 되었다.

- 세 번째로, 단순히 기능 개발이 아닌 프로젝트 개발의 전체 과정에 대해 이해할 수 있는 안목을 기르는 기회가 되었다. 단순히 기능이 돌아가는지, 성공하는지 만들기만 하는 것이 아닌, ICT 프로젝트의 여러 과정을 통해 사용자의 경험을 고려하고, 우선순위에 따라 개발하는 방법을 배웠다. 멘토님의 조언으로 팀원들이 깨닫지 못하는 부분을 보완하고 고찰할 기회를 가질 수 있었고, 그만큼 다양한 요소들에 대해 고려해보며 프로젝트를 진행할 수 있었다. 이는 프로젝트의 전체 흐름을 보는 시야를 넓히고 프로젝트의 방향성을 잡는 것에 핵심적인 역할을 하였다. 또한 우리 팀이 만든 프로젝트가 실제 산업에서 어떻게 활용될 수 있을지 고민하고 조사하는 과정에서 우리 프로젝트의 결과물에 대해 실용적 가치를 제감할 수 있었다.

- 마지막으로, 이 프로젝트를 통해 우리는 실제 문제를 해결하는 과정에서 발생하는 우리 팀의 극복과정의 도전정신과 문제를 해결하였을 때 얻는 큰 성취감에 대한 가치를 알게 되었다. 또한 주도적으로 어려운 문제들을 해결하는 과정에서 공학인으로서 성찰의 경험과 자신감을 얻을 수 있었다. 이러한 경험과 깨달음은 우리 팀원 모두에게 기술적 성장뿐만 아니라 프로젝트 관리 능력, 팀워크, 그리고 문제 해결 능력의 향상을 가져다 주었다. 이는 앞으로의 프로젝트와 전문적인 경력 개발에 있어 큰 자산이 될 것이라고 느꼈다.

- 이번 프로젝트는 우리 팀에게 단순한 기술 개발 경험을 넘어 종합적인 성장의 기회가 되었다. 프로젝트 관리, 팀워크, 기술적 도전, 윤리적 고려사항, 그리고 실제 산업 적용 가능성에 이르기까지 다양한 측면에서 배움을 얻었다. 이 경험은 우리가 앞으로 마주하게 될 더 큰 도전들을 해결할 수 있는 기반이 되어줄 것이라고 생각한다. 또한, 이번 프로젝트를 통해 얻은 자신감과 문제 해결 능력은 팀원들의 성장에 있어 귀중한 밑거름이 되어 줄 것이다. 해당 프로젝트가 단순히 학술적 성과를 넘어, 실제 세계의 문제를 해결하고 사회에 긍정적인 영향을 미칠 수 있는 잠재력을 가졌다는 점에서 큰 자부심을 느꼈다/ 앞으로도 이러한 경험을 바탕으로 지속적인 학습과 성장을 이어나가, 기술과 사회의 발전에 기여할 수 있는 전문가로 성장해 나가고 싶다.

## IV. 기대효과 및 활용분야

### 1. 프로젝트의 기대효과

- 본 프로젝트는 자율주행 자동차 기술의 상용화에 발 맞추어 교통 분야 AI 학습 데이터의 개인정보 비식별화를 플랫폼 형식으로 안전하게 처리하고 제공하는 것을 목표로 한다. 해당 프로젝트는 민감한 교통 데이터를 비식별처리하여 보안을 강화하고, 기업의 데이터 작업 프로세스의 효율을 높인다. 기존의 비식별화 서비스와 달리, 본 플랫폼은 사용자 친화적인 웹 인터페이스를 제공하고 인공지능 모델로 데이터를 비식별처리하여 바로 사용할 수 있도록 데이터 마켓 형식으로 데이터를 편리하게 이용할 수 있도록 한다.

또한, 객체 인식 AI 모델을 활용해 자동으로 민감한 개인정보 영역을 탐지하고 비식별 처리함으로써, 기존의 수작업 방식 대비 시간과 비용을 크게 절감할 수 있다. 이는 자율주행 기술 개발에 소요되는 비용을 낮추고, 기술 발전 속도를 높이는 데 기여할 것으로 기대된다.

우리 플랫폼은 개인정보가 안전하게 보호되면서도 데이터의 활용성이 유지되므로, 해당 플랫폼의 사용을 통해 데이터 공유와 협업이 활발해질 것으로 기대할 수 있다. 자율주행 기술 발전에 긍정적인 선순환 구조를 형성하여, 해당 산업의 경쟁력을 강화시키고 시장을 활성화할 수 있다.

아울러 윤리적이고 합법적인 데이터 활용이 가능해짐에 따라, 자율주행 기술에 대한 소비자의 신뢰도 또한 제고될 것으로 보인다. 이는 자율주행 자동차의 상용화를 앞당기고, 관련 시장의 성장을 가속화하는 요인으로 작용할 것이다.

### 2. 프로젝트의 활용분야

- 본 프로젝트를 통해 개발된 개인정보 비식별화 플랫폼은 자율주행 자동차 제조사를 비롯한 교통 데이터 활용 기업에 광범위하게 적용될 수 있다. 테슬라, 우버 등 자율주행 시장의 주요 플레이어들은 방대한 양의 주행 영상 데이터를 수집하고 있는데, 해당 플랫폼을 통해 개인정보 보호된 교통 분야 데이터를 합법적이고 안전하게 수집하는 것이 가능하다.

- 또한 국내 교통 데이터 기반 서비스가 확대되는 추세에서, 다인승전용차 단속 제도 시행 등의 불법이용 단속, 교통흐름 개선을 위해 개인정보 비식별화 기술 이용이 필요하다. 이러한 국내 사례에서도 본 플랫폼을 이용해 빠르고 정확한 행정 처리를 수행할 수 있을 것이다. 또한 가명정보 간 결합 데이터는 행정 결정에 더욱 중요하게

사용되고 있다. 최근 한국인터넷진흥원(KISA)<sup>1)</sup>에서 강조한 “가명정보 간 결합”의 중요성을 고려할 때, 본 프로젝트의 활용 가능성은 더욱 확대될 수 있다. 예를 들어, 한 기업이 전기차 충전소 배치를 위해 지방자치단체의 시민 이동 정보를 결합하여 의사결정에 활용한 사례처럼, 우리의 비식별화 플랫폼을 통해 처리된 데이터는 다양한 기관의 정보와 결합되어 새로운 가치를 창출할 수 있다.

- 주목할 만한 점은 EU의 일반 데이터 보호 규정(GDPR)<sup>2)</sup>에서 데이터 가명처리를 통해 개인 식별이 불가능하게 만들면 규제의 적용을 받지 않도록 하고 있다는 것으로, 해외 진출을 목표로 하는 기업들에게 큰 의미를 갖는다. 우리의 비식별화 플랫폼은 이러한 글로벌 규제 환경에 부합하는 솔루션을 제공함으로써, 국내 기업들의 해외 진출을 지원하고 글로벌 시장에서의 경쟁력을 높이는 데 기여할 수 있다.

- 비정형데이터 비식별화 처리 기술은 딥페이크와 같은 디지털 범죄의 예방에도 효과적으로 사용될 수 있을 것이라고 추측된다. 딥페이크 기술의 발전과 함께 증가하는 디지털 성범죄 예방에도 본 플랫폼이 중요한 역할을 할 수 있다. 개인의 얼굴이나 신체 부위를 자동으로 탐지하고 비식별화함으로써, SNS나 온라인 플랫폼에서 수집된 개인 이미지가 불법적인 딥페이크 콘텐츠 제작에 악용되는 것을 방지할 수 있다. 이는 개인의 프라이버시 보호뿐만 아니라, 디지털 성범죄 예방에도 크게 기여할 것으로 예상된다.

- 아울러 비식별화된 교통 데이터는 운전자의 행동 패턴과 선호도 분석 등에 활용되어, 보다 개인화되고 지능화된 교통 서비스 개발로 이어질 수 있다. 이는 자율주행 자동차의 안전성과 편의성을 높이는 한편, 교통 체증 해소와 같은 사회적 편익 창출에도 기여할 것으로 기대된다.

- 더 나아가, 교통 분야뿐 아니라 의료, 금융 등 다양한 영역에서 데이터 활용에 대한 수요가 높아지고 있는 만큼, 개인정보 비식별화 기술의 적용 범위도 점차 확대될

- 1) 김준혁. (2024, 6월 13). “가명정보 결합은 데이터경제의 꽃…정부, 사례 확대 나선다. 파이낸셜뉴스. <https://www.fnnews.com/news/202306131104580780>, 해당내용 “김 팀장은 “한 기업은 전기차 충전소 배치에 앞서 한 지방자치단체가 가진 시민 이동 정보를 결합해 기업 의사결정에 활용한 사례도 있다”고 설명했다.”
- 2) Burt, A., Stalla-Bourdillon, S., & Rossi, A. (2021, July 15). A guide to the EU’s unclear anonymization standards. IAPP. <https://iapp.org/news/a/a-guide-to-the-eus-unclear-anonymization-standards/>, 해당내용 “Specifically, the GDPR defines anonymous data as data that “does not relate to an identified or identifiable natural person or to personal data rendered anonymous” so “the data subject is not or no longer identifiable.” Data that meets this criteria is therefore not subject to the GDPR.”

것으로 예상된다. 다양한 마스크 옵션과 다양한 비정형 데이터 유형에 대응할 수 있는 유연한 시스템을 가진다면 시스템 기능의 확장과 고도화가 원활히 가능하다. 구조본 프로젝트의 성과를 데이터 활용과 개인정보 보호의 균형점을 모색하는 데 있어 좋은 선행로서 활용하여 데이터 활용성과 공유를 촉진할 수 있을 것이다.

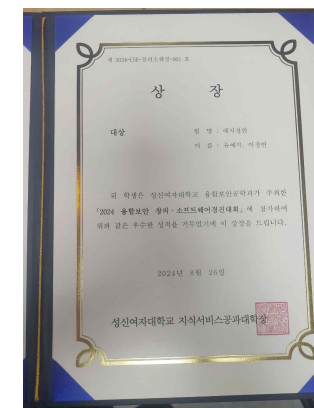
## V. 참고자료

## 1. 참고 및 인용자료

- 해당없음

## 별첨

## 2024년 ICT멘토링 프로젝트 달성성과 증빙자료



### S3. 빅데이터

01. LSTM & Transformer 기반 횡단 물동량 시계열 AI 예측 기술 구현 KIPS\_C2024B0395  
신보람\*, 박주아, 장혜정, 최가연, 장혜원(동덕여자대학교), 이규영(한국과학기술원)

### S4. 인공지능

01. 생성형 AI 및 RAG 활용한 횡단공사 챗봇 서비스 개발 KIPS\_C2024B0048  
김태경\*, 신상준, 이강준, 조성민(인하대학교), 박준규(한국항공대학교), 김한우(ECS대학교)
02. 적대적 AI 공격에 대한 해킹산업 보안강화 연구 KIPS\_C2024B0147  
전준석(서울과학기술대학교), 김화준(고려대학교), 이현화(서울과학기술대학교), 윤다영(성신여자대학교), 박자우(수원대학교), 이규영(한국과학기술원)
03. 자율운항선박의 안전운항을 위한 적응적 AI 및 강화학습 기술의 연구 KIPS\_C2024B0190  
홍예령(성신여자대학교), 박주현(동덕여자대학교), 조혜원(이화여자대학교), 김민승(충익대학교), 한지은(강남대학교), 이규영(한국과학기술원)
04. AI를 활용한 유해 화학물질 가스 누출 실시간 감지시스템 KIPS\_C2024B0259  
신기백\*, 장혜리, 조수형, 홍예람(서울과학기술대학교)
05. 자율운항선박 강화학습 에이전트의 충돌회피 성능향상에 관한 연구 KIPS\_C2024B0269  
손건희(국립공주대학교), 이연재, 조연수(충실대학교), 김화준(고려대학교), 정현원(한국해양대학교), 양한주(리얼기술사무소)
06. AI 기반 실시간 위험 안내 서비스 개발 KIPS\_C2024B0509  
이선구(한국외국어대학교), 고정대(연세대학교), 홍성훈(인하대학교), 권용범(순천향대학교), 신창화(디노별류)
07. 비선형 AI 모델을 활용한 횡단 물동량 예측 KIPS\_C2024B0511  
김영호\*, 김예은, 김민환, 김재훈, 장소민(영희대학교)

ICT연도원(이브와)

### C1. 컴퓨터시스템 및 이론

01. AI 기반 맞춤형 언어 치료 애플리케이션 개발 및 기대 효과 KIPS\_C2024B0099  
홍지은, 손지우, 아예진, 김현진(덕성여자대학교)

### C5. 시뮬네이션

01. RFID 기술과 스마트 전자물품을 이용한 실내 내비게이션 솔루션 KIPS\_C2024B0083  
조유진\*, 장훈지, 오상은, 박지연(배화여자대학교), 박준용(지즈에이드)

### C6. 정보보안

01. 프라이버시 보호를 위한 적대적 AI 공격기법의 연구 KIPS\_C2024B0170  
조혜원(이화여자대학교), 한지은(강남대학교), 김민승(충익대학교), 박주현(동덕여자대학교), 홍예령(성신여자대학교), 이규영(한국과학기술원)

### C7. 개인정보보호



01. 인간 및 AI 실험 기반 비식별화 기법 성능 평가 KIPS\_C2024B0218  
유예지\*, 이정연, 이자연, 양소윤, 최한우(성신여자대학교)

### C9. ICT융합



01. NLP 기반 소셜 IP 확장 솔루션 개발 KIPS\_C2024B0157  
김예수, 류영주, 유수연\*, 황상민(덕성여자대학교)
02. 인공지능 기반 영어 학습 서비스 개발 연구 KIPS\_C2024B0201  
최가연\*, 이새은, 상유진, 임양규(덕성여자대학교)

### S4. 인공지능

01. 대화 기반 행동 및 음성 분석을 활용한 CBT 보조도구 연구 KIPS\_C2024B0114  
김수연, 신정훈, 양희정, 이윤희\*, 정은채(덕성여자대학교), 김민원(OH!스큐브랩)

### S5. 메타버스 및 XR

01. 눈 건강 관리를 위한 XR 기반 디지털 헬스케어 서비스 개발 KIPS\_C2024B0240  
이새란\*, 이정은, 박승현(한신대학교), 문소진(MedicalIP)