

PORTFOLIO.

안녕하세요.

사용자 중심 엔지니어 정경준입니다.



CONTENTS.

프로필 3~4

인적사항 / 경력사항 / 수상이력
활동내용 / 학력사항 / 자격사항 / 스킬

경력사항 5~14

데이터 표준화 자동화 지원 / TTA 품질검증용역 /
개인정보 비식별화 / AI 학습용 데이터 구축사업

기타경험 15~18

AI 자연어처리 기술 전문가 / 임베디드 소프트웨어 경진대회 /
더넥스트스쿨주식회사 / Ospitek Inc.

연락처 19


ABOUT.




정 경 준

KyungJun Jung

1994. 02. 22 / 30세

 +82 10 9479 1654

 jkigoworld@gmail.com

 서울시 은평구 응암동

경력사항

2022.05 ~ 재직중
비투엔 / DIX 3팀 / 담당

2021. 01 ~ 2021. 07
더넥스트스쿨주식회사 / 커리큘럼R&D / 강사

학력사항

2013.03 ~ 2019.02
강원대학교 / 전자통신공학전공
전체평점 : 3.71 / 4.5
전공평점 : 3.81 / 4.5

수상이력/활동내용

2021.11 ~ 2022.03
AI 기술 자연어처리 전문가 양성 과정 / 구름

2021.05 ~ 2021.12
제 19회 임베디드 소프트웨어 경진대회
/ 임베디드 SW/System 산업협회

자격사항/어학

2024. 03
데이터 분석 준전문가 / 한국데이터산업진흥원장

2021. 12
SQL 개발자 / 한국데이터산업진흥원장

2024. 03
OPIC(영어) / IM1

ABOUT.



정경준

KyungJun Jung

1994. 02. 22 / 30세

+82 10 9479 1654

jkjgoworld@gmail.com

서울시 은평구 응암동

해외경험

2020.04 ~ 2020. 09
Ospitek Inc./ 미국 샌디에고 인턴

나를 소개하는 키워드

#ENTP

#센스

#포용력

#행동력

#책임감

#운동

보유스킬



80/100

pandas

70/100



60/100



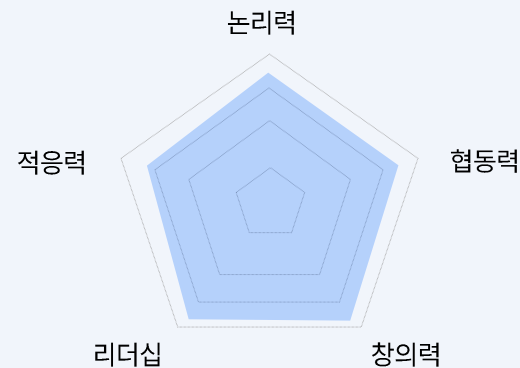
50/100



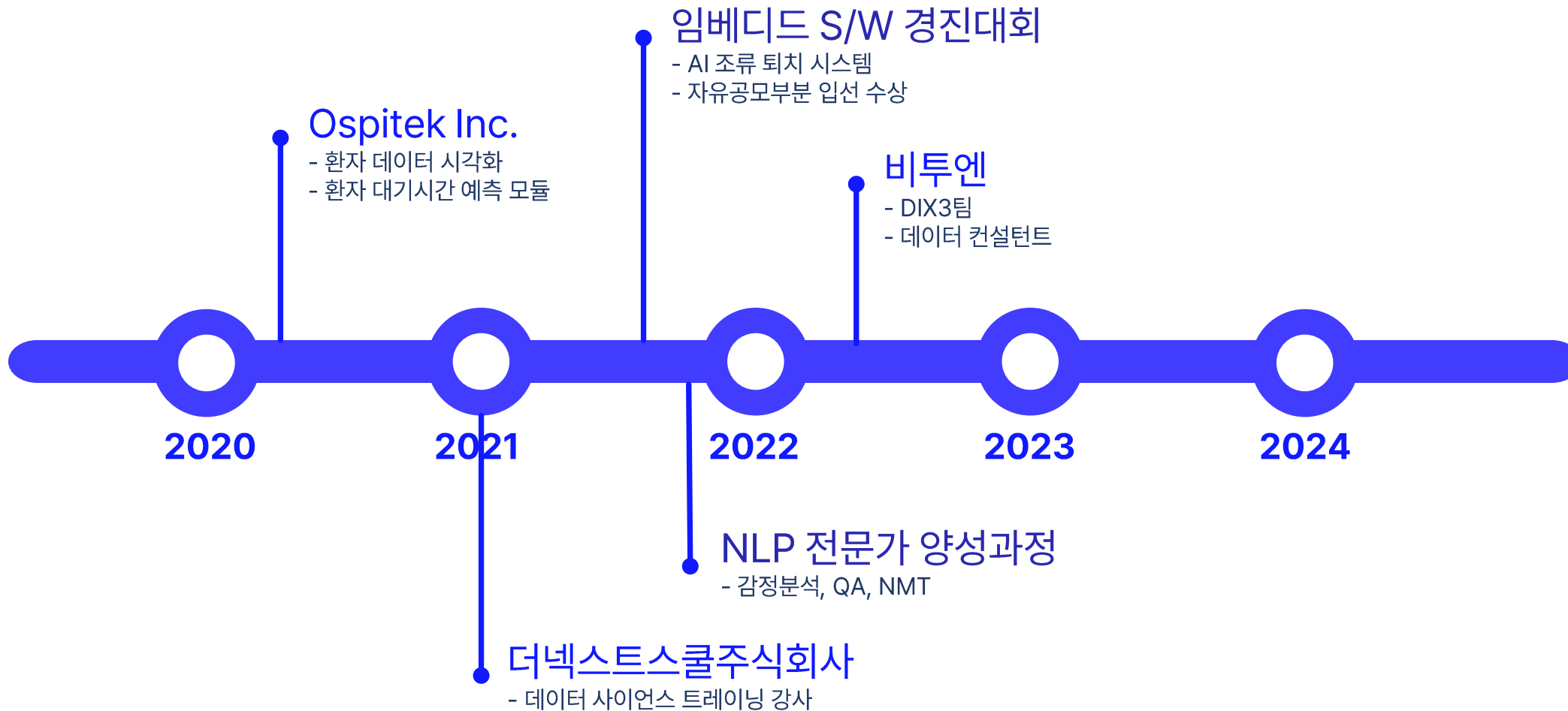
40/100



30/100



TIMELINE



EXPERIENCE

01 경력사항

기간 : 2022. 05 ~ 재직중

회사 : 비투엔

직급 : DIX 3팀 / 담당

데이터 컨설턴트, # 데이터 엔지니어, # AI 엔지니어

01

sLLM기반 Data Solution
AI Assist R&D

데이터 표준화 자동화 모듈

2024.03 ~ 2024.05

02

TTA 품질검증
구문정확성 용역

300개 과제
데이터 품질검증

2023.07 ~ 2024.02



03

개인정보 식별 및 탐지

비정형데이터
개인정보 비식별화

2023.04 ~ 2023.07

04

AI 학습용 데이터
구축사업

자연어, 헬스케어 데이터
구축

2022.05 ~ 2023.03

PROJECT 01

과제 목표

sLLM기반 Data Solution AI Assist R&D

GenAI 기반 AI Assist 기능을 추가하여 데이터 표준화 업무 시 일부 기능의 자동화를 지원합니다.

※ AI Assist Features

- ✓ 단어 분리, 단어 설명, 용어 설명 추천
- ✓ 표준 사전에 없는 용어명 입력 시, 연관 비즈니스 용어명을 추천



인원

8명



역할

sLLM 모델을 위한 SFT 데이터셋 구축



기술스택

Python, transformers, PostgreSQL



기간

2024.03 ~ 2024.05 (2개월)

사용자 질문

단어 "3개월납"의
설명을 작성해줘

"3개월납보험료"를
단어분리해줘

"3개월납보험료"의
설명을 작성해줘

AI Powered Pack

Data Solution

답변

한 번에 3개월치의 보험료를
납부하는 것을 의미

3개월납 + 보험료

보험료를 일정 기간마다 납부하는
방식

< 과제 목표 >

문제점

데이터
수집·생성

데이터
전처리

데이터
필터링

SFT
수행


DPO
추가학습

결과

SFT 데이터셋 구축

< 과제 프로세스 >

나의 경험

기여도 50/100기여도  30/100

기여도 100/100

➤ 기존 데이터 품질 오류

- 사이트별 표준화 용어들이 상이함
- 표준화 작업시 백과사전을 참조하는 등 대개 수작업으로 진행됨



- SFT 데이터셋 정의
- ChatGPT API 활용
데이터셋 생성 -> 변환
- 단어·용어사전 고도화
- 약 40만 건 데이터셋 확보

- 모델 학습 방해 요소 정제
- 외국어 정제
- 중복 제거 (우선순위 적용)
- 도메인 특화 용어명,
사이트명 제거

- 단어·용어의 정의 적합성을 판단하는 분류 모델 구현
- Precision 0.85 달성

- ✓ 고품질 표준사전 구축
 - 단어: 1만 건
 - 용어: 5만 건
- ✓ Pretrained Model 대비 30% 성능 증가

[illegible]

- 특수 문자 :

$$\Pi ::= \langle \rangle \mathbb{X} :: \langle \rangle \leftrightarrow \langle \rangle \sim \langle \rangle ? :: \langle \rangle \mathbb{F}$$
- 영어로 시작하는 문장
- 문장에 한글이 포함되지 않은 문장

→ 해당 사람은 펠리머지, 후, 아이제키하여 수정함

후후 [영어] **목속** / 관공 **목**작은 사람이 똑똑하여 훌륭한 구실을 함을 비유적으로 이르는 말

후 후후는 작아도 **매**다 **후**후대만 **덜**뻘뻘하다

단어 분리: 50.28%
단어 설명: 측정 불가
용어 설명: 측정 불가

단어 분리 : 94.94%
단어 설명 : 90.00%
용어 설명 : 73.33%

PROJECT 02

과제 목표

TTA 품질검증 구문정확성 용역

AI 학습용 데이터의 구문적 정확성, 통계적 다양성 검사를 수행합니다.

※ 사업 수행 방안

- ✓ 자연어, 영상·이미지, GIS, 로보틱스, 법률 등 14개 유형, 300개 과제 학습데이터 품질 검증
- ✓ 새로운 품질항목 "유사성" 모델 개발 (사업 적용)
- ✓ 새로운 품질항목 "유해성" 모델 제안



인원

5명



역할

TTA 품질검증수행



기술스택

Python, Tensorflow, AWS Athena

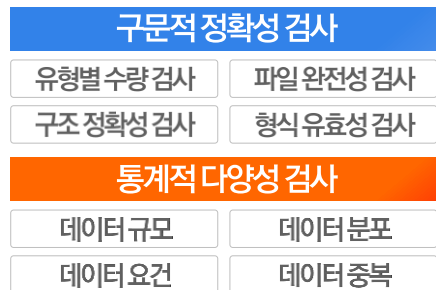


기간

2023.07 ~ 2024.02 (8개월)



< 사업 제안 >



< 300개 과제 품질 검사 >



< 결과서 산출 >

(이전) 품질 검증 이후 데이터

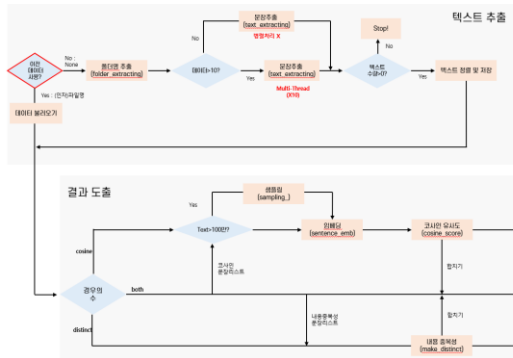
과제명	유형	파일명	파일 크기	파일 개수	파일 형식	파일 위치
과제명	유형	파일명	파일 크기	파일 개수	파일 형식	파일 위치
과제명	유형	파일명	파일 크기	파일 개수	파일 형식	파일 위치

(이전) 품질 검증 이후 데이터

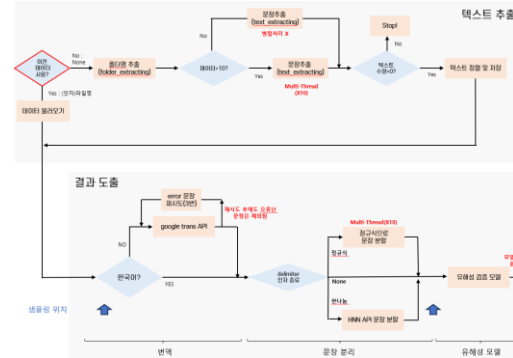
과제명	유형	파일명	파일 크기	파일 개수	파일 형식	파일 위치
과제명	유형	파일명	파일 크기	파일 개수	파일 형식	파일 위치
과제명	유형	파일명	파일 크기	파일 개수	파일 형식	파일 위치

< 사업 종료 >

< 과제 프로세스 >



< 유사성 모델 도식화 >



< 유해성 모델 도식화 >

나의 경험



- ✓ TTA 품질검사 용역 (60개 과제 수행)
- ✓ 문장 유사성 모델 구현
- ✓ 문장 유해성 모델 구현

기여도 30/100

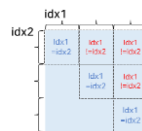
결과

- 작년 사업에서 데이터 중복 제출한 컨소시엄들이 존재
- 자동화되지 않은 검사항목들이 존재함. 이는 사업비에서 큰 비중을 차지



- 이모지, 특수문자
한자 제거
- 문장 단위로 구별
(과제별 상이, 한나눔API)
- 문장 추출 및 코사인유사도
계산을 위한 문장 벡터화

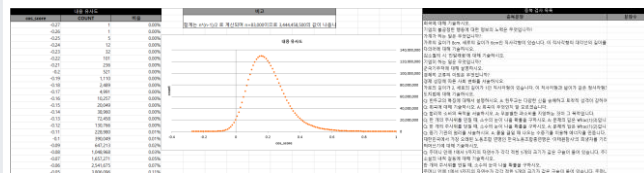
- 전체 문장을 대상으로
코사인 유사도 계산을 위한
알고리즘 구현



- 결과 산출물 자동화

- 14만 건의 데이터 수집
- 4천개 비속어 사전 구축
- Kcelectra 모델 선정, 하이퍼파라미터 튜닝
- 결과 산출물 자동화

- ✓ 유사성 모델 해당년도 사업 적용
 - 30개 과제 적용
- ✓ 유해성 모델 2024년 제안 예정
 - 92% 성능 달성



	N	특소급 주출	번역	결과 도출	총 소요 시간
한국어	생물 문장 수 : 1,000,000 [대체어 수 : 2,396,192]	2402 55분 [대체어 수 : 2,396,192]	-	2시간 4분	6시간 59분
외국어 (영어/중국어)	생물 문장수 : 1,000,000 [대체어 수 : 2,002,310]	1942 38분 [대체어 수 : 2,002,310]	10142 51분 [대체어 수 : 2,002,310]	55분	121시간 49분

산출물									
산출물명	종류	종류코드	종류명	단위	단위코드	단위명	단위명	종류	단위
1) 1차	농산물	0100000000	농산물	1톤	0100	농산물	농산물	1톤	0100
2) 2차	농산물	0200000000	농산물	1톤	0200	농산물	농산물	1톤	0200
3) 3차	농산물	0300000000	농산물	1톤	0300	농산물	농산물	1톤	0300
4) 4차	농산물	0400000000	농산물	1톤	0400	농산물	농산물	1톤	0400
5) 5차	농산물	0500000000	농산물	1톤	0500	농산물	농산물	1톤	0500
6) 6차	농산물	0600000000	농산물	1톤	0600	농산물	농산물	1톤	0600
7) 7차	농산물	0700000000	농산물	1톤	0700	농산물	농산물	1톤	0700
8) 8차	농산물	0800000000	농산물	1톤	0800	농산물	농산물	1톤	0800
9) 9차	농산물	0900000000	농산물	1톤	0900	농산물	농산물	1톤	0900
10) 10차	농산물	1000000000	농산물	1톤	1000	농산물	농산물	1톤	1000
11) 11차	농산물	1100000000	농산물	1톤	1100	농산물	농산물	1톤	1100
12) 12차	농산물	1200000000	농산물	1톤	1200	농산물	농산물	1톤	1200
13) 13차	농산물	1300000000	농산물	1톤	1300	농산물	농산물	1톤	1300
14) 14차	농산물	1400000000	농산물	1톤	1400	농산물	농산물	1톤	1400
15) 15차	농산물	1500000000	농산물	1톤	1500	농산물	농산물	1톤	1500
16) 16차	농산물	1600000000	농산물	1톤	1600	농산물	농산물	1톤	1600
17) 17차	농산물	1700000000	농산물	1톤	1700	농산물	농산물	1톤	1700
18) 18차	농산물	1800000000	농산물	1톤	1800	농산물	농산물	1톤	1800
19) 19차	농산물	1900000000	농산물	1톤	1900	농산물	농산물	1톤	1900
20) 20차	농산물	2000000000	농산물	1톤	2000	농산물	농산물	1톤	2000
21) 21차	농산물	2100000000	농산물	1톤	2100	농산물	농산물	1톤	2100
22) 22차	농산물	2200000000	농산물	1톤	2200	농산물	농산물	1톤	2200
23) 23차	농산물	2300000000	농산물	1톤	2300	농산물	농산물	1톤	2300
24) 24차	농산물	2400000000	농산물	1톤	2400	농산물	농산물	1톤	2400
25) 25차	농산물	2500000000	농산물	1톤	2500	농산물	농산물	1톤	2500
26) 26차	농산물	2600000000	농산물	1톤	2600	농산물	농산물	1톤	2600
27) 27차	농산물	2700000000	농산물	1톤	2700	농산물	농산물	1톤	2700
28) 28차	농산물	2800000000	농산물	1톤	2800	농산물	농산물	1톤	2800
29) 29차	농산물	2900000000	농산물	1톤	2900	농산물	농산물	1톤	2900
30) 30차	농산물	3000000000	농산물	1톤	3000	농산물	농산물	1톤	3000
31) 31차	농산물	3100000000	농산물	1톤	3100	농산물	농산물	1톤	3100
32) 32차	농산물	3200000000	농산물	1톤	3200	농산물	농산물	1톤	3200
33) 33차	농산물	3300000000	농산물	1톤	3300	농산물	농산물	1톤	3300
34) 34차	농산물	3400000000	농산물	1톤	3400	농산물	농산물	1톤	3400
35) 35차	농산물	3500000000	농산물	1톤	3500	농산물	농산물	1톤	3500
36) 36차	농산물	3600000000	농산물	1톤	3600	농산물	농산물	1톤	3600
37) 37차	농산물	3700000000	농산물	1톤	3700	농산물	농산물	1톤	3700
38) 38차	농산물	3800000000	농산물	1톤	3800	농산물	농산물	1톤	3800
39) 39차	농산물	3900000000	농산물	1톤	3900	농산물	농산물	1톤	3900
40) 40차	농산물	4000000000	농산물	1톤	4000	농산물	농산물	1톤	4000
41) 41차	농산물	4100000000	농산물	1톤	4100	농산물	농산물	1톤	4100
42) 42차	농산물	4200000000	농산물	1톤	4200	농산물	농산물	1톤	4200

PROJECT 03

과제 목표

개인정보 식별 및 탐지

온라인 구어체 등 일상대화에서 개인정보를 식별 및 탐지합니다.

* 솔루션 기능

- ✓ 14개 개인정보 요소 식별
- ✓ 데이터 마스킹, 데이터 범주화 등 비식별화 처리
- ✓ Web application, On-premises방식 지원

인원 1명

역할 개인정보 비식별화 솔루션 기능 구현

기술스택 Python, Django, Regex, CSS, PostgreSQL, Spark...

기간 2023.04 ~ 2023.07 (4개월)

Git 주소 <https://github.com/koreaiot/de-identification>



< 과제 프로세스 >

비식별화 모듈 데모

정보입력을 통한 비식별화

*정보를 입력하세요

안내하세요. 환영합니다.

- 나이 : 30세
- 전화번호 : 010-9479-1654
- 주소 : 서울특별시 은평구 불광전길 466 999호
- 이메일 : jkigworld@gmail.com
- 주민등록번호 : 940222-1234567
- 대표번호 : 127.0.0.1

시뮬레이션 초기화

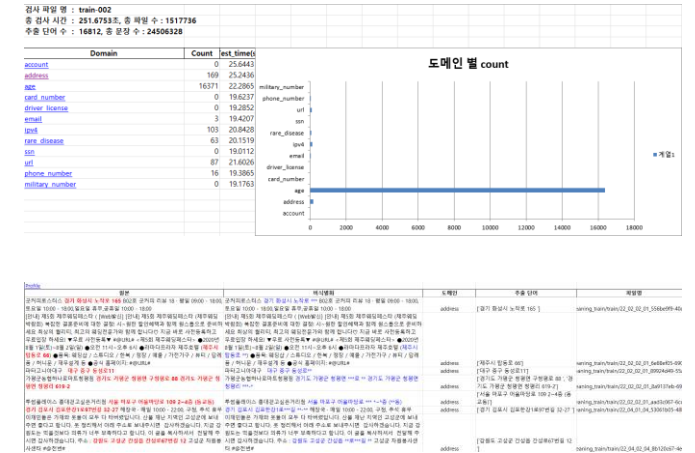
검사항목 :

['account', 'address', 'age', 'card_number', 'driver_license', 'email', 'ip4', 'rare_disease', 'sex', 'url', 'phone_number', 'military_number']

입력정보 : 안내하세요. 환영합니다. - 나이 : 30세 - 전화번호 : 010-9479-1654 - 주소 : 서울특별시 은평구 불광전길 466 999호 - 이메일 : jkigworld@gmail.com - 주민등록번호 : 940222-1234567 - 대표번호 : 127.0.0.1

비식별화 : 안내하세요. 환영합니다. - 나이 : 30세 - 전화번호 : 010-****-**** - 주소 : 서울특별시 은평구 불광전길 **** - 이메일 : *****@****.com - 주민등록번호 : 940222-1234567 - 대표번호 : 127.0.0.1

< Output 1>



< Output 2 >

PROJECT 03

나의 경험



- ✓ 1인 과제 End-to-End 경험
- ✓ 개인정보보호법, 비식별화 솔루션 세미나 진행
- ✓ 테스트케이스 수집, 서버 구현, 솔루션 개발

기여도 100/100

기여도 100/100

기여도 100/100

문제점

- AI HUB 데이터에서 개인정보 데이터 발견
- 개인정보 결합, 비속어 탐지 솔루션은 존재하나, 비정형데이터에서 개인정보 탐지 솔루션은 많지 않음



과제 정의

- 개인정보보호법 참고
- 과제 정의, 기획, 개발진행
- 20개 개인정보 요소 중, 14개 요소 구현
- 자사 솔루션의 독립적인 플러그인 모듈 제안

데이터 수집

- 공공데이터 포털 활용
- 나무위키 개인정보체계 참고 ex) 주소체계, 계좌번호종류
- 웹 애플리케이션 구현 -> 사용자 입력 데이터 수집
- 테스트케이스 100만건 확보

탐지 로직 구현

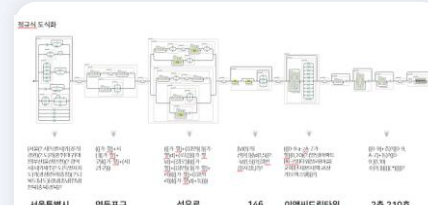
- 정규식 필터
- DB 매칭
- 도메인별 조건 연산 로직
- Transformer, spaCy모델

결과

- ✓ TTA 문장 유해성 항목 제안 모델로 발전
- ✓ 개인정보 비식별화 기술 세미나 진행
- ✓ AI hub SNS 데이터 기준 98% 탐지 가능



참고자료



PROJECT 04





과제 목표

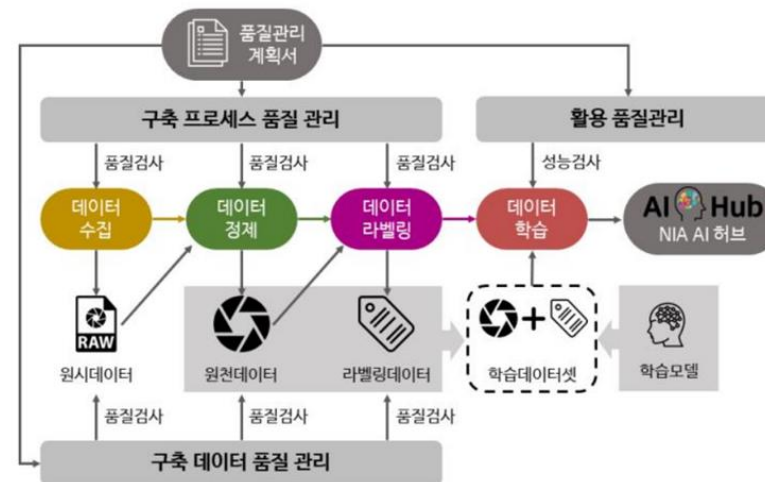
인공지능 학습용 데이터 구축사업

응급실 임상 대화 데이터, 구음장애인 명령어 데이터,
SNS 데이터 고도화 구축사업에서 품질 담당자 역할 수행

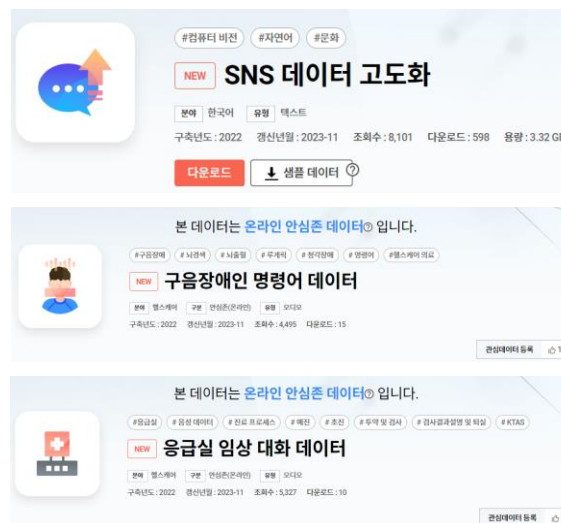
※ 품질 담당자 역할

- ✓ 공정 품질 체크리스트, 컨소시엄 품질 교육
- ✓ 수집·정제·라벨링 각 단계 산출물 데이터 품질 검증
- ✓ 모델링 지표 설정, 모델링 과정 모니터링
- ✓ TTA 품질 검증 대응

	인원	약 20개 기관 컨소시엄 구성
	역할	데이터 품질 검증, 데이터 컨설턴트
	기술스택	Python, PostgreSQL, Speech Recognition, BERT
	기간	2022.05 ~ 2023.03 (11개월)



< 과제 프로세스 >



< 참여 과제 >



< 컨소시엄 대·내외 품질 활동 >

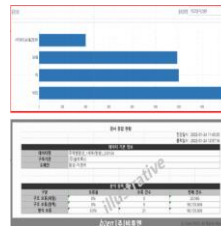
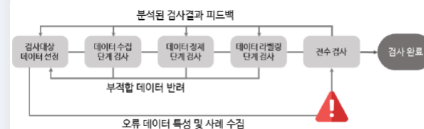
나의 경험



- | | | |
|-----|------------------------|---------|
| 기여도 | <div><div></div></div> | 20/100 |
| 기여도 | <div><div></div></div> | 100/100 |
| 기여도 | <div><div></div></div> | 80/100 |

결과

- 

[illegible]

EXPERIENCE

02 기타 경험

NLP, # 데이터 사이언티스트,
헬스케어, # Object detection

01

K-DIGITAL TRAINING
AI 기술 자연어처리 전문가
양성 과정

감정 분류, QA, NMT

2021.11 ~ 2022.03



02

임베디드 소프트웨어
경진대회 자유부문

입선 수상
Objection detection

2021.05 ~ 2021.12

03

더넥스트스쿨주식회사

Data Scientist 교육과정
강사 및 조교

2021.01 ~ 2021.07

04

Ospitek Inc.

샌디에고 IT 헬스케어 인턴

2020.04 ~ 2020.09

임베디드 S/W 경진대회

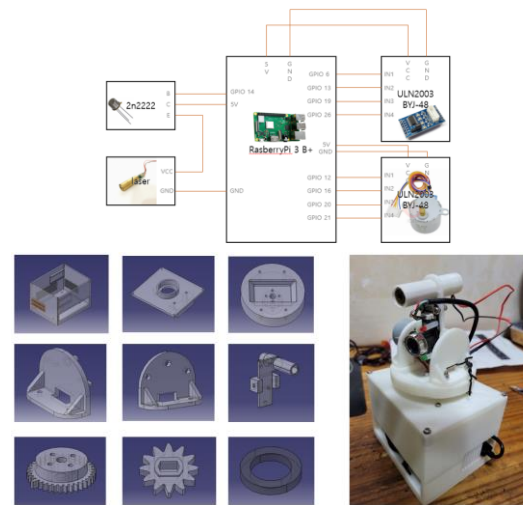
도심형 조류 퇴치 AI 시스템

제 19회 임베디드 소프트웨어 경진대회 입선 수상,
그린레이저를 이용한 유해조류 퇴치 시스템

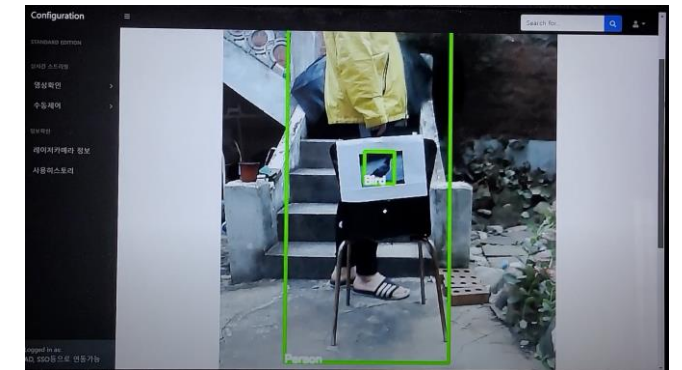
※ 모듈 기능

- ✓ Object Detection, Tracking
- ✓ 라즈베리파이 모터 제어
- ✓ Web socket 통신, 모니터링 서버 시스템

	인원	5명
	역할	S/W 개발, 수집/정제, 이미지 학습
	기술스택	Python, Java, CSS, Linux
	기간	2021.05 ~ 2021.12 (7개월)
	Git 주소	https://github.com/koreaiot/2021ESWContest_free_1124
	시연영상	https://youtu.be/Ze3FZgkKXO4?si=Rfgm3yly6tCjr1Re



< 하드웨어 >



< 소프트웨어 >

임베디드 S/W 경진대회

나의 경험



- ✓ 라즈베리파이 모터 제어
- ✓ 모니터링 서버 구축
- ✓ 이미지 객체 인식 · 추적

기여도 90/100

기여도 50/100

기여도 100/100

문제점

- 기존 조류 퇴치 제품은 산업용으로 설계되어 크고, 무겁고, 비싸고 위험
- 도심에서 사용해도 위험하거나 공해가 끼치지 않는 제품 필요



라즈베리파이 제어

- 모터, 레이저 제어
- 객체인식 시 캡처, DB저장
- 관할구역검사 자동화



모니터링 서버

- 소켓통신
- 촬영 위치, 날짜 등 DB 구축
- 로봇 관측 모니터링
- DDNS 서버 할당

이미지 학습

- 객체 인식, 트래킹
- SSD-mobileNet 학습
- 실제 유해조류 사진 4천건 수집 및 라벨링
- Tflite 변환

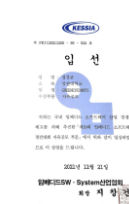
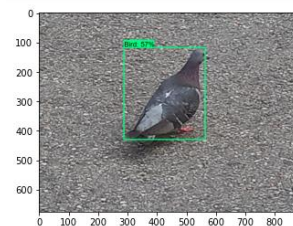
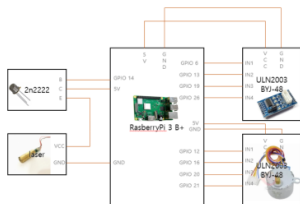
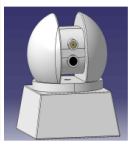


결과

- ✓ 자유공모부문 입선
- ✓ 77% 성능 달성
- ✓ 이미지 학습, DB 구축, 서버 통신, 코드 최적화 역량 향상



Spec	Design
레이저 규격	30mW (2등급)
사거리	50m
작동반경(수직/수평)	-50° to +30° / 360°
크기	15x15x20
통신	Wifi, 5G
전원	파워 어댑터(220V)
	배터리(30000mAh)
방수성능	IPX6 수준




기타 경험

AI 기술 자연어처리 전문가 양성 과정

 기술스택 Python, Pytorch, Tensorflow, Transformers


 기간 2021. 11 ~ 2022. 03 (총 5개월)

 역할 ML, DL, NLP 역량 향상
감정분석, 질의응답, 번역모델 구현

더넥스트스쿨주식회사

 기술스택 Python, Pandas, Seaborn, SQL


 기간 2021. 01 ~ 2021. 07 (총 8개월)

 역할 데이터 사이언스 트레이닝 교육 커리큘럼 R&D
Python, Pandas 강사

Ospitek Inc.

 기술스택 Python, Tableau, Linux


 기간 2020. 04 ~ 2020. 09 (총 6개월)

 역할 환자 데이터 시각화
환자 대기시간 예측 알고리즘 구현

전자통신공학전공 졸업작품

 기술스택 Python, Java, Android, Linux

 기간 2018.09 ~ 2019.02 (총 6개월)

 역할 "라즈베리파이를 이용한 메카넘휠 정찰로봇" 논문 등재
숙명여대 학술대회 참여

감사합니다,
열정 넘치는 인재
정경준이었습니다.

CONTACT

+82 10 9479 1654

jkjgoworld@gmail.com

Thank you!