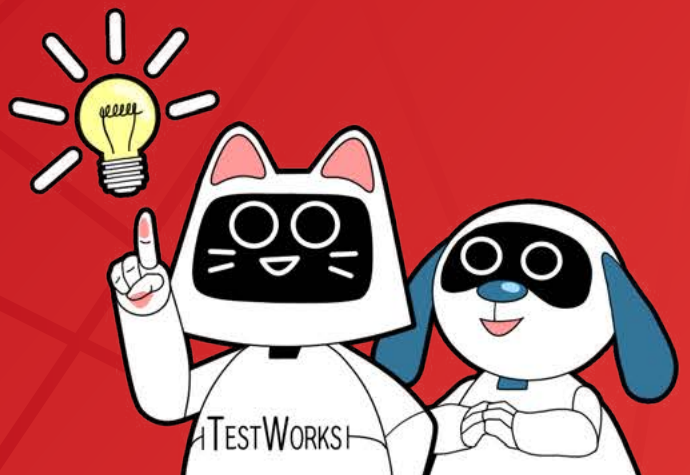


테스트웍스와 함께 알아가는 AI & Data 이야기



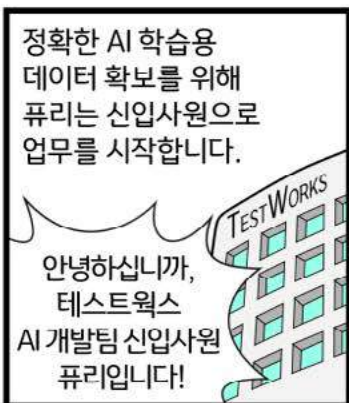
목차

프롤로그	03
Episode.01 인공지능의 발달과 개발 트렌드1	04
Episode.02 인공지능의 발달과 개발 트렌드2	05
Episode.03 AI핵심 역량 양질의 데이터 확보	06
Episode.04 인공지능의 발전과 산업 기술의 혁신1	07
Episode.05 인공지능의 발전과 산업 기술의 혁신2	08
Episode.06 인공지능 지도학습 데이터1	09
Episode.07 인공지능 지도학습 데이터2	10
Episode.08 인공지능 지도학습 데이터3	11
Episode.09 핫해진 데이터 라벨링1	12
Episode.10 핫해진 데이터 라벨링2	13
Episode.11 데이터 가공 기술과 품질의 중요성	14
Episode.12 데이터 전문 가공 기업1	15
Episode.13 데이터 전문 가공 기업2	16
Episode.14 대규모 인공지능 데이터셋 구축사례 사회문제해결1	17
Episode.15 대규모 인공지능 데이터셋 구축사례 사회문제해결2	18



테스트웍스 AI 해결사 티키타카와 함께하는 인공지능 이야기

테스트웍스 AI 해결사 티키와 타카, 그리고, AI 정보 전달사 시 퓨리를 소개합니다.



AI 학습용 데이터 전문기업
테스트웍스

[바로가기](#)

TESTWORKS

최근 혁신적으로 발전한 알고리즘, 컴퓨팅, 빅데이터 기술이 융합되면서 인공지능은 다양한 현실 속에서 실제 서비스로 구현되어 우리의 일상속으로 들어오고 있습니다. 어떤 서비스가 일상에 구현되어 있는지 확인해 보세요.

Episode. 01
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

인공지능의 발달과 개발 트렌드

미래에는 인공지능 로봇인 내가 대세가 되는 건가?

한국판 뉴딜, 시가 차세대 먹거리

어이쿠

콰

시는 미래의 일이 아니라 이미 우리의 생활속에 들어와 있다고!

세계를 놀라게 한

알파고를 기억해?

알파고! 잘 알지~ 최근에 만나서 이야기도 한 걸~

알파고를 기점으로 시는 한단계 더 진화해

그야말로 르네상스

가~

그 정도는 나도 알아!

그 이후로 글로벌 기업들은 적극적으로 연구 분야를 개척하고

빠르게 제품과 서비스로 개발하고 있어

바...벌써?

인지기능과 학습기능

금융 의료 교육 가상비서 ARS 자율주행 지능로봇

추론-행동 지능이 혁신을 이루면서 추천시스템

오호~

시는 우리의 생활속에서 이미 구현되고 있어!

인공지능은 항상 멀게 느껴졌는데 우리 주변에 가까이 있었다니!!!

너무 걱정하지 마! 쉽고 재미있게 알려줄 티키타카가 있으니까!

많이 알려줘~ 나도 시에 대해 더 많이 알고 싶어!

인공지능 트렌드를 알아보는 가장 빠른 방법

바로가기

TESTWORKS

우리에게 편리함을 제공하는 인공지능의 학습은 어떻게 진행되고,
테스트웍스에서는 어떤 서비스를 지원하고 있을까요?

Episode. 02
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

인공지능의 발달과 개발 트렌드 2

인공지능은 알고리즘, 컴퓨팅 파워, 빅데이터 분야의

혁신적 발전으로 생활 속에서 다양한 형태로 구현되고 있어

모든 기술들이 전부 연결되어 있구나

오늘의 추천 음악

인공지능의 여러 분야에서 정보를 인식, 이해, 추론하며 행동하는 수준까지 진행되고 있지

대단하네!!!

자 그럼, 문제! 어떻게 인공지능을 학습 시킬 수 있을까?

QUIZ

데이터 로드중...

지도학습, 비지도학습, 강화학습!!!

정답이야!

그 중 문제를 풀고 답을 맞추는 방법인 지도학습에서부터 인공지능의 혁신이 시작되고 있지!!!

착

시각, 청각, 언어의 인식을 위해 다양한 이미지, 영상, 음성, 텍스트 데이터가 활용되고 있어!!!

착

인식, 이해, 추론, 행동의 학습을 위해 이미지+음성, 이미지+텍스트

자

인공지능에서는 완전 중요한 내용이구나

혹은 더 고도화된 레이블링 구조 및 형태 등 복합적이고 다양한 데이터의 학습이 필요해

짜

메모리 입력!

그래서 인공지능의 개발 및 상용화를 위해서는 고품질의 다양한 학습데이터가 있어야해.

퓨리가 잘 아는 회사가 바로 인공지능 학습데이터 구축 전문기업이야!

내가 잘 아는?

바로 지금 다니고 있는 테스트웍스야!

테스트웍스에서는 데이터 가공을 위해 다양한 솔루션을 제공하고 있지

와~ 그런 중요한 역할을 하고 있었다니 뭔가 자랑스러운걸

AI Data 전문 기업
테스트웍스가
궁금하다면?

바로가기 >

AI 핵심 역량 양질의 데이터 확보

인공지능은 개발하고자 하는 서비스 분야에 따라 다양한 학습 데이터가 필요하며 데이터의 양과 질이 인공지능의 성능을 좌우할 만큼 고품질의 학습 데이터를 확보하는 것은 매우 중요합니다.
데이터 라벨링이라고도 하는 데이터 어노테이션에 대해 살펴 보겠습니다.

Episode. 03
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

AI 핵심 역량! 양질의 데이터 확보

대한민국의 인공지능 기술과 관련 산업은 빠르게 발전하고 있어

카~ 역시 대한민국!!!

하지만 여전히 글로벌 기술장벽은 높기만 하지!

소프트웨어 역량 데이터 확보

아아... 좌절

데이터는 많으면 좋은 거 아니야?

AI 개발 역량의 경쟁력을 높이기 위해서는 반드시 양질의 데이터 확보가 필요해!

인간에게는 익숙한 영상이나 이미지도

AI가 학습하기 위해 정교한 어노테이션 과정이 필요해!!

하늘
나무
산

아~ 데이터 어노테이션! 까다롭지만 매우 중요한 과정이구나.

데이터 어노테이션이란?

인공지능 서비스 구현에 필요한 다양한 데이터를 기계 학습이 가능한 형태로 준비하는 전처리 과정

데이터 전처리 과정은 비용 뿐만 아니라 소요되는 시간에서도 전체 인공지능 서비스 구현 과정의 80% 차지

구글, 페이스북 등이 AI를 선도할 수 있는 건 전처리 과정을 거친 양질의 빅데이터를 확보했기 때문이지

아하~

AI의 시작이자 핵심인 데이터 어노테이션을 꼭 기억해두!

앞으로도 자주 언급될 테니까~

메모리 입력 완료!

AI 학습 데이터 수집·가공 전문 플랫폼
aiworks

바로가기 >

인공지능은 컴퓨터 학습을 가능하게 하는 알고리즘, 하드웨어, 데이터의 발전과 함께 다양한 서비스로 구현되고 있으며 실제 산업에 적용되면서 기술의 진입 장벽을 넘어 새로운 혁신을 일으키고 있습니다. 다양한 산업 분야의 새로운 혁신에 대해 알아보겠습니다.

Episode. 04
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

인공지능의 발전과 산업 기술의 혁신

인공지능은 알고리즘, 하드웨어, 데이터의 발전으로

**알고리즘
하드웨어
데이터**

이전의 한계를 극복하고 학습 알고리즘의 성능을 인정받으면서

다양한 서비스 구현 및 산업 분야로 확장하는 중이야!

학습 알고리즘?

인공지능 기계학습 알고리즘의 유형

기계학습 알고리즘은 데이터를 이용해 컴퓨터를 학습시키는 방법론이지. 지도학습, 비지도학습, 강화학습 등으로 구분할 수 있어.

인공지능은 이미 국방, 의료, 교육, 게임, 보안 등 다양한 산업 분야에 응용되고 있지. 참, 자율주행도 기억하라고...

자율주행 분야에서 최근 성공한 스타트업들은 카메라, 초음파 센서, LiDAR, RADAR에서 수집된 정보를 이용한

초음파 센서



LiDAR



RADAR



딥러닝 기반 인식, 판단, 제어 기술을 구현해 빠르게 성장하고 있지

너를 위해 3만개의 주행영상 데이터를 준비했어

자동차 분야 전문가와 컴퓨터 분야의 전문가들이 협업하여 딥러닝 기반 지능형 자율 주행이 가능하다는 건



인공지능이 실제 산업 분야에 적용되면서 기술의 한계를 넘어 새로운 혁신을 가져오고 있다는 얘기지

와~ 굉장한데~

인공지능 데이터 구축 사례가 궁금하다면?

바로가기 ▶



인공지능의 발전과 산업 기술의 혁신 2

인공지능 서비스 중 가장 쉽게 접할 수 있는 인공지능 스피커!
 인공지능 스피커의 핵심인 음성인식 기술과 음성인식 기술을 발전시킨 인공지능 기계학습에 대해 살펴보겠습니다.

Episode. 05
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

인공지능의 발전과 산업 기술의 혁신 2

오호
꿀팁

인공지능이라고 하면 어떤 게 가장 먼저 떠올라?

음... 인공지능 스피커?

맞아! 그동안 음성 언어 인식 분야는 상대적으로 혁신이 지지부진했지

시무룩

하지만, 딥 러닝 기반의 대용량 자동 음성인식으로 탁월한 성능 구현이 예상되고 있어!

Deep learning

음성인식 기술이 더욱 정교해져 사람들의 대화를 완전히 이해하기 위해서는 방대한 양의 정확도 높은 학습 데이터가 절대적으로 필요하지!

음성인식의 핵심 STT와 NLP

- STT (Speech to Text): 주변 소음을 제외한 말소리를 파악하고, 각 발음과 단어를 인식한 후 핵심어와 연결 단어를 인식하는 기술
- NLP (Natural Language Processing, 자연어 처리 과정): 컴퓨터를 통해 사람들이 의사소통을 위해 사용하는 언어를 분석, 이해, 처리하는 과정

Microsoft, Google, Apple

Deep learning

MS 코타나, 스카이프 번역기, 구글 나우, 애플 시리 등 대표적인 음성인식 시스템들이 딥 러닝 기법에 기반하고 있어

AI스피커 -> 서비스

현재의 음성 인식 지능의 발전은 아마존, 구글, 애플 등 글로벌 기업과 국내 기업들의 다양한 음성 인식 서비스 경쟁으로 확산되고 있지

자율주행과 음성 인식 분야를 통해 알아본 인공지능 기계학습은

자율주행 음성인식

TESTWORKS

다양한 학습 알고리즘의 발전과 새로운 시도를 통해 지속적으로 발전해 현재는 학습 경쟁의 시대에 돌입했다고 볼 수 있어

와~ 앞으로가 더 기대돼!

영상으로 알아보는
인공지능 데이터
가공 기술 적용 사례

바로가기 >

TESTWORKS

인공지능은 개발하려는 서비스 분야에 따라 다양한 학습 데이터가 필요합니다.

예를 들어, 차량용 도로와 다른 특성을 가진 보행자용 인도의 경우 자율주행과는 다른 데이터셋이 필요합니다. 데이터셋이 무엇인지 함께 알아보겠습니다.

Episode. 06
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

인공지능 지도학습 데이터 1



인공지능이라고 다 같은 인공지능이 아니야!

잉?

도로를 예로 들어볼까?

도로? 자율주행을 말하는 거야?

도로에는 자율주행에 필요한 차량용 도로와 이와는 다른 특성을 가진 보행자용 인도가 있어



자율주행 서비스를 구현하기 위해서는 도로객체인식을 위한 학습 데이터가 필요하지



그렇지

보행자를 대상으로 한 인공지능 기술 개발을 위해서는 자율주행과 비슷하지만 다른 데이터셋이 필요해



아하

인공지능은 개발하고자 하는 서비스 분야에 따라 다양한 학습 데이터를 필요로 해

고품질의 데이터 = 인공지능의 성능

고품질의 데이터를 얼마나 많이 확보하고 있느냐가 인공지능의 성능을 좌우하겠구나

그렇지. 바로 그 데이터의 주인공이 테스트웍스지!

고품질의 학습 데이터가 곧 경쟁력! OK~

특히 인도 보행에 대한 인공지능 기술은 다양한 사회 문제 해결의 토대를 마련한다는 의미가 있어



와~ 굉장한데~

‘인도 보행 영상’ 데이터셋 구축 사례

바로가기 >

TESTWORKS

인공지능 데이터셋은 서비스 개발 분야에 따라 다양한 형태의 학습 데이터가 필요합니다.
인공지능 서비스 개발의 예시와 함께 확인해 보겠습니다.

Episode. 07
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

인공지능 지도학습 데이터 2

궁금해~

서비스 분야에 따라 다양한 학습 데이터가 필요하다는 말 기억해?

응, 같은 도로라도 자율주행과 보행자를 위한 서비스는 다르니까

이번에는 멀티 모달 학습데이터 구축의 경우를 예로 들어 볼게!

Multi Modal

멀티... 모달?

컴퓨터는 이제 시각, 청각, 촉각, 미각, 후각 등 다양한 센서를 통해 꼭 말로 하지 않아도 인간의 마음을 이해할 수 있어

시각
후각
촉각

미각
청각

멀티 모달이란 다양한 입력 방식을 융합하여 인간과 컴퓨터 사이에 자연스런 의사소통이 가능한 사용자 친화형 기술이야

외부 침입 감지, 대화를 통한 감정 분석, 영상 상황/의도 분석 등 멀티 모달 기술을 활용할 수 있는 곳은 무궁무진하지!

침입 감지 감정 분석 영상 분석

의료용 학습데이터도 서비스 개발에 따라 다양한 데이터가 필요해

이미 지금도 사람의 관절 데이터셋을 기반으로 나이를 추정하거나, 자세 교정 등의 인공지능 서비스가 구현되고 있어

버... 벌써?

최근에는 반려동물 데이터에 대한 관심도 높아 강아지나 고양이에 대한 데이터를 수집해 다양한 인공지능 서비스 개발이 진행되고 있지

데이터 수집
전문 플랫폼 aiworks

바로가기 >

TESTWORKS

아~ 학습 데이터가 필요하지 않은 분야가 없구나!

인공지능 지도학습 데이터 3

대기업을 중심으로 연구가 진행되던 인공지능 기술은 IT 인프라 확보와 인공지능 알고리즘 개발로 인하여 스타트업, 소상공인, 1인 기업까지 확산되고 있습니다. 인공지능 서비스 개발 활성화를 위한 오픈 데이터셋에 대해 알아보겠습니다.

Episode. 08
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

인공지능 지도학습 데이터 3

이제 대부분의 서비스 분야에 인공지능이 기본으로 자리잡겠어

반짝이는 아이디어만 있으면 되지!

하지만 대기업이 아니라면 역시 어렵겠지?

노노~~

IT 인프라 확보와 인공지능 알고리즘이 개발되면서

스타트업, 소상공인, 1인 창업까지 확산되고 있어

**스타트업
소상공인
1인 창업**

기술의 고도화에 따라 실제 산업군에 맞게 다양한 서비스들이 구현되고 있지

국가 차원에서 다양한 산업군의 인공지능 서비스 개발 활성화를 위해

**데이터바우처 사업
오픈 데이터셋 구축
정부지원**

정부가 지원하는 데이터 바우처 사업이나 오픈 데이터셋 구축 사업이 활발히 진행되고 있어

산업별/서비스별 인공지능 개발과 관련된 다양한 고품질의

인공지능 서비스의 빠른 상용화

학습 데이터를 확보하는 것이 인공지능 서비스의 빠른 상용화를 가져오지

고품질 학습데이터 확보

오호~ 그렇게네!

자율주행, 인도보행, 멀티모달, 의료, 반려동물 등을 통해 알아본 인공지능 지도학습 데이터 이야기 어땠어?

인공지능과 더 친해진 느낌이야~ 유익한 시간이었어

그리고 지도학습 데이터가 인공지능에 얼마나 중요한 지도 알게 되었어

TESTWORKS

맞아! 그 중심에 바로 테스트웍스가 있다는 것도 기억해~

TESTWORKS

데이터바우처 지원사업
성공 파트너 테스트웍스

[바로가기](#)

TESTWORKS

국내 중소-벤처기업들이 인공지능 학습용 데이터를 자체 구축하기에는 많은 시간과 비용이 소요되며 원천 데이터 확보에도 어려움이 있습니다. 이를 해결하기 위해 대규모로 데이터를 구축하여 민간이 참여하는 선순환 생태계 조성 및 확대를 위해 과학기술정보통신부와 한국지능정보사회진흥원(NIA)이 추진하는 AI 학습용 데이터 구축 사업을 확인해 보겠습니다.

Episode. 09
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

핫해진 데이터 라벨링 1

지난 4년 동안 국가 차원에서 심혈을 기울인 사업이 있어

그... 그게 뭔데?

바로 AI 데이터 구축 사업이지!

연도	투자액 (억 원)
2017	40
2018	90
2019	195
2020	3315

정부예산이 무려 82배나 늘어났다고!

대박인데~

'한국판 뉴딜'에 최적화된 AI 데이터 구축 사업은 차세대 먹거리인 'AI 육성'과 '일자리 창출'이라는 두마리 토끼를 동시에 잡는 방법이 될 수 있어

과학기술정보통신부에서도 AI 데이터 구축 사업을 4차 산업혁명 시대 기계 학습을 위한 대표 사업으로 꼽고 있지

4차 산업혁명 대표 사업

일자리 창출 효과에 대한 기대도 높겠구나

당연하지

일자리

재택근무가 가능한 클라우드 소싱 방식을 기준으로 10억원 당 200명의 일자리 창출이 가능하다는 연구 결과가 있어

200명
10억원

정부는 AI데이터셋을 구축해 개방함으로써

AI 성장의 토대를 구축하고 AI 발전의 선순환 생태계를 만들겠다는 취지야

AI 발전

↓

AI 성장 토대 마련

데이터 구축

↓

데이터 개방

대규모 데이터셋 가공에 최적화된 All-in-One 솔루션
blackolive

[바로가기](#)

AI 데이터 구축을 위해서는 원천 데이터에 설명을 추가하는 라벨링 작업이 기초가 되어야 합니다. 인공지능 서비스 개발에 활용이 가능한 학습 데이터 구축을 위한 데이터 라벨링에 대해 알아보겠습니다.

Episode. 10
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

핫해진 데이터 라벨링 2

국가 차원에서 그 정도로 심혈을 기울이고 있다는 건 그만큼 사업이 어렵고 복잡하다는 의미겠지?

그렇지도 않아~

‘AI 데이터 구축 사업’은 간단해 이름 그대로

디지털 데이터에 ‘라벨’을 부착하는 작업이야

그래서 ‘데이터 라벨링’이라고 부르는구나~

기술/산업에서 필요로 하는 인공지능 기술 구현에 활용성과 범용성이 높은 데이터를 얼마나 많이 확보하는가 하는 점이 핵심이지

데이터 라벨링의 핵심

기술/산업적으로 유망
AI 응용개발에 공통적으로 활용
범용성 높은 AI 데이터
(이미지/동영상 등)

대한민국도 중요성을 인식하고 노력하고 있지만 AI 부문 선진국이라 할 수 있는 미국에 비하면 아직 많이 부족한 것이 사실이야

1600만개의 이미지 데이터 라벨링 공개 스탠포드 대학	오른 이미지 데이터 900만개 구글
---	---------------------------

사업 자체가 어려운 것 보다는 막대한 인건비가 문제겠구나.

데이터 숫자가 더더덜덜

그래서 정부도 발벗고 나섰어 지난 3년간 21종에 불과했던 데이터 라벨링을 170종까지 대폭 확대하기로 했지

데이터 라벨링 170종

우와~ 2020년은 대한민국 AI의 새로운 전환점이 되겠는데!

‘한국판 뉴딜’의 발판이 될 데이터 라벨링을 잘 기억해 두라고~

‘데이터 라벨링’하면 ‘테스트웍스’라는 것도 잊지 말고!

TESTWORKS

데이터 자동화-가공-관리 솔루션
blackolive의
최신 데이터 라벨링 기술

바로가기

인공지능의 알고리즘 성능이 아무리 탁월해도 저품질의 데이터를 입력한다면 AI 서비스는 무용지물이 될 수 있습니다. 데이터 품질이 왜 중요한지 알아보겠습니다.

Episode. 11
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

데이터 가공 기술과 품질의 중요성

왜 그래? 어디 아파? 아무래도 저품질의 데이터를 입력한 모양이야

아이고 배야~~

저품질 데이터?

인공지능 알고리즘 성능이 우수해도 품질이 낮은 데이터를 이용한다면 AI 서비스는 무용지물이 될 수 있어

그러니까 아무거나 주워먹지 말라고~

머신러닝과 딥러닝을 위한 학습 데이터는 그 양도 중요하지만 품질도 매우 중요해

정교한 인공지능 구현의 필수요소인 고품질 학습 데이터는 가공단계부터 시작하기 때문이지

AI의 경쟁력 = 양질의 데이터

양질의 데이터를 제공하기 위한 다양한 데이터 가공 솔루션과 서비스가 있어

그 중에서도 'Web-based Annotation Platform'은

오차 없이 고품질의 데이터를 대량 생산하는 기술 프로세스를 성공적으로 구축했어

Web-based Annotation Platform

이 학습 데이터 가공 기술은 단순히 Raw-Data를 가공하는 것이 아니라, AI 개발 주체가 원하는 AI 솔루션/엔진의 방향성에 대한 이해에서 출발하지

AI모델의 목적과 특성을 고려한 프로젝트 설계 → **작업자 매니지먼트** → **맞춤 툴과 직관적 UI/UX 개발** → **검수 솔루션**

다수의 어노테이터가 장소에 상관없이 '다양한 환경'에서 다양한 데이터 가공이 가능해

초기 AI 모델을 활용해 라벨링 작업을 자동화하는데, 이 부분이 AI 예측에 대한 검수 작업으로 변환되면서 작업량이 확 줄어드는 거야

효율성이 크게 향상돼

아! Web-based Annotation Platform이 대량의 순도 높은 데이터를 얻는 가장 혁신적인 방법이라는 이야기이지?

테스트웍스의 대규모 고품질 데이터셋 구축 기술

바로가기 ▶

데이터 전문 가공 기업 1


AI 데이터 시장의 성장과 규모에 따라 인공지능 학습 데이터셋 구축 서비스를 전문적으로 제공하는 데이터 가공 및 공급 기업들이 주목을 받고 있습니다. 특히 그 중심에는 다양한 산업군 내에서 대규모 데이터셋 구축 경험 노하우를 보유하고 있는 테스트웍스가 있습니다.

Episode. 12
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

데이터 전문 가공 기업 1


궁금해~

2020년의 전세계 AI 데이터 시장규모는 2100억 달러 (약 242조원)에 달하고



\$210 billion

데이터의 양도 조만간 10배 이상 증가할 것으로 예상하고 있어



Year	Volume
2016	16ZB
2025	163ZB

10배

AI 데이터 산업 규모가 엄청나구나! 그런데 기업들은 왜 데이터가 부족하다고만 하지?

느낌적인 느낌?

그건 데이터 자체가 부족한 것보다 '어노테이션'된 데이터가 부족하다는 의미야

데이터 어노테이션, 즉 데이터의 수집/분석에서 가공까지의 과정은 결코 쉬운 일이 아니야

기업의 일반적인 인적 자원만으로는 데이터 어노테이션을 전문적으로 수행하기에 턱없이 부족하지

[AI 개발 프로세스]
데이터 확보/관리
데이터 어노테이션
AI 학습
서비스화

➔

[기업의 인적자원]
개발조직
연구조직
사업조직

AI 개발 과정에서 비용과 시간이 가장 많이 소용되는 건 뭐라고 했지?

정답이야! 그 중에서 어노테이션이 무려 25%를 차지해~

데이터 전처리 과정! 전체 과정의 약 80% 정도? 맞지?

데이터 어노테이션 작업은 최소 수만개의 데이터에 라벨링을 해야 하는데 기술력이 없으면 단순 노동 집약적 업무가 되어버려

백만스물하나~
백만스물둘~
그런데 더 이상은 우리야


너무 걱정하지마!
빠른 시간에 저비용으로도

어노테이션이 가능한
꿀팁이 따로 있으니까

빨리 알려줘~
현기증 난단 말이야

인공지능 데이터 구축
원스톱 서비스

바로가기 >




15


데이터 전문 가공 기업 2

테스트웍스는 AI 학습용 데이터 라벨링 및 관리에 최적화 된 자동화 가공 관리 솔루션 블랙올리브를 보유하고 있습니다. 자체 데이터셋으로 더 고도화 되어 사용자 별 강점으로 편리해진 블랙올리브에 대해 알려 드립니다.


Episode. 13
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

데이터 전문 가공 기업 2





궁금해~



아! 데이터 어노테이션을 위한 꿀팁 알았다! 전담부서를 만들면 되는거지?


때!

AI 기업에서 데이터 어노테이션까지 직접 진행하려면 프로젝트 마다 어노테이션 도구를 개발하거나~~

작업 인력을 일시적으로 총원/교육/관리해야 하는 등 수많은 어려움이 발생해

작업 인력 어노테이션 도구 기업 후들

정답은 바로 빠른 시간에 저비용의 어노테이션 작업을 위해 AI를 활용하는 거야!



뭣이라? AI를 위한 AI 활용? 소름~

하지만 AI는 연구만으로는 불가능하고 학습 데이터가 많을수록 정확도가 높아지

동일한 AI 모델이라도 데이터 투입량에 따른 정확도 차이가 무려 18%에 달한다고

데이터 투입량에 따른 정확도 차이 18%

AI가 점차 고도화되면서 세분화된 어노테이션의 필요성이 증가하고

그 주기도 계속해서 짧아질 전망이야

경쟁이 치열해질수록 어노테이션의 수요는 확대되겠구나

테스트웍스는 이러한 상황을 효과적으로 개선했지

데이터 수집 가공에 적합한 환경을 갖추고 있어 빠른 시간에 대량의 데이터를 고품질로 어노테이션 할 수 있다고!

TESTWORKS

데이터 자동화 가공 관리 도구

데이터 수집 가공 클라우드 소싱 플랫폼

내부 및 외부 데이터 전문 인력 보유

고객 만족 고품질 관리 프로세스

테스트웍스의 블랙올리브는 다양한 비정형 데이터를 어노테이션 할 수 있지

이미지

비디오

블랙올리브 (데이터 자동화 가공 관리 도구)

텍스트

음성

고객사의 AI 솔루션 및 엔진의 방향에 대한 이해에서 출발 하기 때문에

로우데이터를 단순 가공하는 것이 아니라 다양한 기업의 요구 사항을 효과적으로 반영할 수 있지

오~ 데이터 전문 가공 기업 테스트웍스~

더 향상된 인공지능 서비스를 위한 고품질 데이터셋 구축의 시작과 끝

바로가기 >

TESTWORKS

과학기술정보통신부와 한국지능정보사회진흥원(NIA)이 2019 인공지능 학습용 데이터 구축 사업의 일환으로 추진한 ‘인도 보행 영상 시 데이터’ 과제에서 테스트웍스가 주관 기업으로 데이터셋을 구축했던 사례를 소개해 드립니다.

Episode. 14
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

대규모 인공지능 데이터셋 구축사례 사회문제해결 1

궁금해~

테스트웍스와 같이 사회적 가치 창출을 추구하는 기업을 뭐라고 하는지 알아?

딩동댕~

그 정도야~ 소셜 벤처, 사회적기업!!!

테스트웍스는 사회 문제해결 관련 데이터 구축 사업에 꾸준히 참여하고 있어

사회문제해결?

대표적인 예로 한국지능정보사회진흥원(NIA)의 ‘인도 보행 영상 시 데이터’ 과제를 들 수 있지.

2019 NIA ‘인도 보행 영상 시 데이터’

- 장애인 보행에 위험 요소인 각종 장애물과 파손 등으로 위험한 보행 노면에 대한 데이터셋 구축
- 장애인의 이동권 신장을 위한 인공지능 서비스 개발의 기반 마련

인도 보행 영상 시 데이터셋은 기존 자율주행차량 데이터셋과 달리 국내 현실에 맞는 인도를 중심으로 구축되었지~

인도 보행 중심의 데이터셋

주체나 규모면에서 국내외에 유래가 없다는 점에서 국내 공개 데이터셋의 독창성과 다양성 확보에 크게 기여했지

최초로 국내 환경에 맞는 인도 보행 영상 데이터셋을 구축 및 보행 보조 시스템과 같은 시범 서비스 활용 사례를 통해, 기술을 통한 장애인 이동권 문제 해결의 토대를 마련했어



전방 주시 스마트폰





모바일 어플을 통한 실시간 탐지/음성 안내

아~ 테스트웍스는 사회적문제 해결을 위한 계획이 다 있구나~

다음회에서는 ‘인도 보행 영상 시 데이터’ 구축 관련

비하인드 스토리와 실제 활용사례를 살펴볼 생각이야. 더욱 흥미진진하겠지~

NIA ‘인도 보행 영상 시 데이터’ 구축 사례

[바로가기 >](#)



테스트웍스가 주관 기업으로 수행한 NIA '인도 보행 영상 AI 데이터' 과제는 최종 수요자가 데이터셋 구축에 직접 참여하였으며, 국내외 유례없는 데이터셋을 구축하였다는 의미가 있습니다.

Episode. 15
테스트웍스와 함께 알아가는 AI(인공지능) 이야기

대규모 인공지능 데이터셋 구축사례 사회문제해결 2

테스트웍스는 소셜벤처라고 하던데...

기업의 이념도 중요하지만 NIA '인도 보행 영상 AI 데이터' 구축 사업에 참여하여 사회적 가치 창출을 실현하는 점도 중요하지 않아요

특히, 이 사업은 척수장애인협회와 연계해 기술의 수혜자가 직접 기술 개발에 참여했다는 점을 주목해야해

테스트웍스는 '인도 보행 영상 데이터셋'을 활용해 보행 보조 시스템을 시범 서비스로 개발했으며, NIA 인공지능 학습용 데이터 활용 우수사례로 선정되었습니다!

구분	산출물	성과 기준(A)	결과(B)	달성율(B/A)
인도 보행 영상 데이터셋	Bounding Box	350,000장	352,810장	100.8%
	Polygon	100,000장	100,712장	100.7%
	Surface Masking	50,000장	51,318장	102.6%
	깊이 인식 Depth Prediction	150,000세트	174,080세트	116.1%

STEM 분야의 다양성 - 4차 산업의 고용모델

테스트웍스는 IT기술 기반의 사회 혁신을 추구하며, STEM (과학, 기술, 공학, 수학) 분야에서 다양성을 실천하고 있어 경력단절여성, 발달장애인, 미취업 청년, 시니어 대상 교육 및 기회를 제공하여 4차 산업혁명 시대의 노동환경 변화에 능동적으로 대응하는 새로운 고용모델을 제시했지

또한, 자체 개발한 솔루션의 자동화와 비식별 처리 기술 및 웹 기반 데이터 수집 도구와 전문 검수 프로세스를 통해 단기간에 대규모의 고품질 데이터셋 구축 역량을 보여주었지

테스트웍스는 기술 기반의 성장과 함께 사회적 가치 창출을 위해 지속적으로 ~

사회 문제 해결을 위한 데이터셋 구축 및 고용 창출에 기여하고자 고민하고 있다고~

기술 기반의 사회 혁신을 추구하는 소셜벤처 테스트웍스

[바로가기](#)

테스트웍스와 함께 알아가는 AI & Data 이야기

참조 링크

과기정통부, 디지털 뉴딜 핵심 '데이터 댐' 구축에 나서다
<https://www.korea.kr/news/pressReleaseView.do?newsId=156396946>

'3년새 예산 82배 폭증' 데이터 라벨링이 뭐길래
http://it.chosun.com/site/data/html_dir/2020/06/04/2020060404335.html

기획/작성

이은영(Testworks)
김형복(Testworks)

그림

김형렬

디자인

(주)유니마인드



2021 Testworks. All rights reserved.


TESTWORKS

(주)테스트웍스 www.testworks.co.kr
서울 송파구 백제고분로 41길 42-19(더블유123빌딩) 3, 4층
Mail : cs@testworks.co.kr | Tel : 02-423-5178 | Fax : 02-424-5178