

2024 **Vol. 7**

CROWDWORKS INSIGHT MAGAZINE

크라우드웍스 프로젝트로 살펴본
산업별 인공지능 데이터 인사이트

CONTENTS

01. Leader's Message

02. 격변하는 AI 시장의 1년과 기업들

03. 클라우드웍스가 살펴본 산업별 인공지능 데이터

1) 금융/보험 산업

2) 교육 산업

3) 자율주행 및 모빌리티

4) 국방/방산

5) 주목할 만한 산업과 데이터

01

Leader's Message





클라우드웍스 전략기획실장 김혜수

Leader's Message

“데이터 구축의 목적과 형태 다양해져도 여전히 중요한 건 데이터의 품질”

지난 1년간, AI 시장은 급격한 혁신과 변화를 경험했습니다. Open AI, Google, Meta 등 빅테크 기업들이 치열한 기술 경쟁을 펼치며 AI 기술의 폭발적인 성장과 시장 확장을 이뤘죠. 이러한 성장을 옆에서 지켜보며 확실히 기업들이 AI를 활용하여 생산성을 향상하고 더 나은 고객 서비스를 개발하며 보다 나은 세상을 만들어 갈 수 있다는 확신을 가지게 된 한해였습니다.

불붙은 AI 기술패권 경쟁 속에서 한국 기업들은 데이터 구축과 활용에 더욱 박차를 가하고 있습니다. 클라우드웍스는 지난 1년간 다양한 산업에서 필요로 하는 데이터를 구축하는 프로젝트를 진행하며, 주요 산업별로 다양한 인사이트를 얻을 수 있었습니다. 금융/보험 산업에서는 맞춤형 LLM 도입을 위한 LLM-ready 데이터 구축이 빠르게 진행되고 있으며, 교육 산업에서는 데이터 구축을 통한 디지털 교과서 도입 준비가 본격화되고 있습니다. 자율주행 및 모빌리티 산업은 잠시 주춤한 것처럼 보였지만 여전히 고품질의 정밀한 데이터가 활발히 구축되고 있으며, 국방/방산 산업에서는 국가 간 치열한 AI 국방 기술 경쟁에 발맞춰 다양한 데이터 구축이 전개되고 있습니다. 이커머스 산업에서는 대량의 데이터를 기업의 데이터 거버넌스 체계에 맞추어 분류하는 수요가 지속될 것으로 예상합니다.

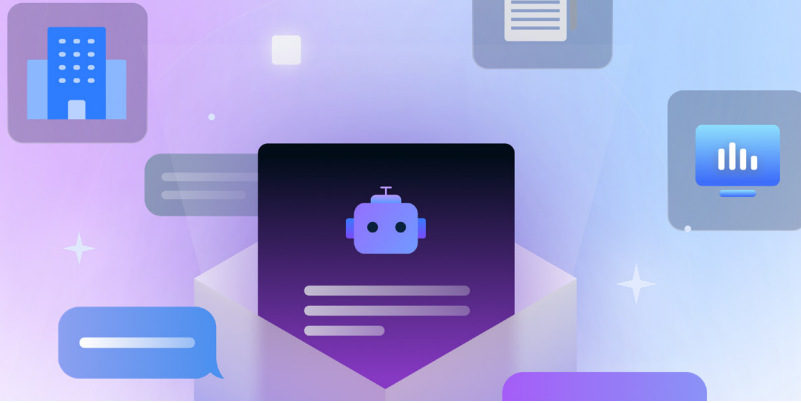
이렇게 다각화된 산업별 니즈 사이에서 확인할 수 있는 가장 중요한, 그리고 공통된 인사이트는 ‘데이터 품질에 대한 중요성은 산업군에 상관없이 항상 강조되고 있다’는 점일 것입니다.

클라우드웍스는 AI가 만들 더 나은 세상을 기대하고 있습니다. 동시에 어떤 편리한 서비스가 나올까, 업무 생산성을 얼마나 더 개선하고 안전한 업무 환경을 만들 수 있을까 등 AI를 어떻게 더 효과적으로 활용할 수 있을지 계속 고민하고 있습니다. 이러한 기대가 현실이 될 수 있도록 클라우드웍스는 앞으로도 다양한 산업에서 필요한 고품질 AI 데이터를 구축하여 기업들이 AI 기술을 효과적으로 활용할 수 있도록 지원하고자 합니다. 다년간 축적한 기술과 노하우를 바탕으로 빠르고 정확한 고품질 데이터를 제공하여 산업 혁신에 기여하며, 리더 기업들과 함께 새로운 미래를 열어나갈 것입니다. 많은 기대와 성원 부탁드립니다.



02

격변하는 AI 시장의 1년과
기업들



2022년이 오픈AI가 챗GPT로 본격적인 ‘생성형 AI 시대’를 연 기념비적인 해라면, 2023년은 오픈AI, 메타, 구글, 앤트로픽 등 이른바 빅테크들이 ‘전쟁’이라고 불려도 좋을 만큼 격렬한 AI 기술 경쟁을 펼쳐 폭발적인 성장과 시장 확장을 이뤄낸 해였습니다. 일반 기업들도 너나 할 것 없이 AI를 통한 생산성 개선, 고객 서비스 개발에 나섰고요. ‘이 기술이 과연 세상을 바꿀 수 있을까?’하는 기업들의 의심은 어느 때보다 빨리 확신으로 변했습니다. 글로벌 시장 조사업체 IDC는 2024년 생성형 AI 시장을 전년 대비 2.7배 성장한 401억 달러 규모로 예측했고, 마켓앤마켓은 2030년 AI 시장이 1조 3452억 달러(약 1800조 원) 규모로 성장할 것으로 예측했습니다.

올해 5월에 GPT4o가 발표되면서 이제 기업들의 확신은 ‘더 늦으면 큰일이다’라는 조바심으로 바뀌고 있습니다. 텍스트뿐만 아니라 음성, 이미지 등 모든 형태의 입력을 이해하고 처리할 수 있는 AI인 GPT4o의 성능은 저희가 보기에 매우 놀랍습니다. 사람과 거의 유사한 속도로 대화가 가능하고 답변을 생성하는 데 드는 비용도 50%나 줄었죠. 주저 없이 영화 ‘Her’를 떠올릴 정도의 성능이었습니다. 엔비디아의 무서운 성장과 더불어 올해 애플, 삼성과 같은 기업들이 온디바이스 AI를 통해 일상 속 AI 활용에 가속도를 붙인다면 우리의 생활은 급격히 변할 것입니다. 이런 상황이니, 우리가 현장에서 만난 기업들의 조바심도 충분히 이해됩니다.

기업들이 각 산업 현장에 AI를 도입하고자 할 때 가장 먼저 하는 일 중 하나는 보유한 데이터들을 정리하여 활용성을 높이고 필요한 데이터를 마저 구축하는 일일 것입니다. 크라우드웍스는 어느 때보다 숨가빔던 지난 1년을 돌아보며 데이터 구축 프로젝트들의 메타데이터를 통해 최근 기업들의 데이터 구축 현황과 특징 등 데이터 및 업계 동향을 살펴보았습니다. 2023년 7월부터 올해 6월 말까지 1년 동안 기업들이 크라우드웍스에서 진행한 프로젝트를 기준으로 산업별로 주로 어떤 데이터를 구축했는지, 데이터 구축의 목적은 무엇이었는지, 산업별 데이터 특징은 무엇인지 살펴보며 여러분과 공유할 만한 인사이트 위주로 정리해 보겠습니다.



기간: 2023.07.01~2024.06.30(1년)

데이터 프로젝트 수: 249건

총 구축 데이터 수: 약 1,390만 개

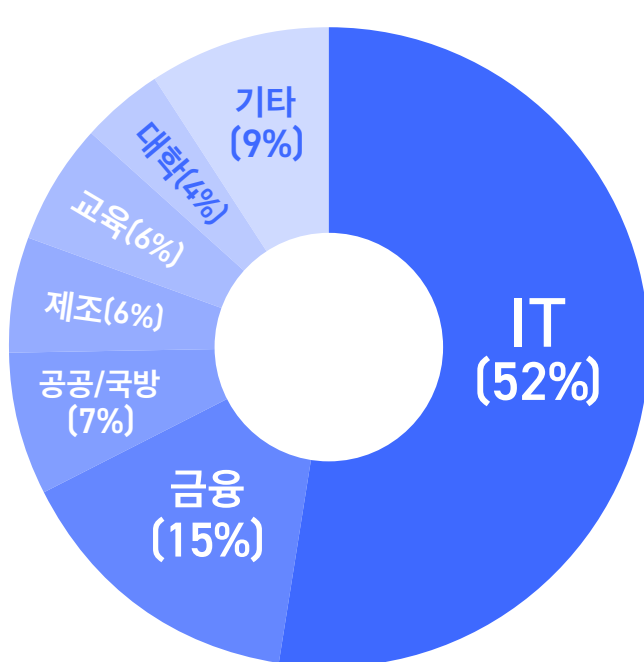
*메타데이터는 프로젝트 상세 내용이나 기업명이 특정되지 않도록 공개할 수 있는 데이터만 활용했으며, 프로젝트 관련 데이터의 경우 하나의 기업에서 다수의 프로젝트를 진행하거나 다양한 형태의 데이터를 구축한 경우도 개별로 포함되어 있습니다.

Key Point

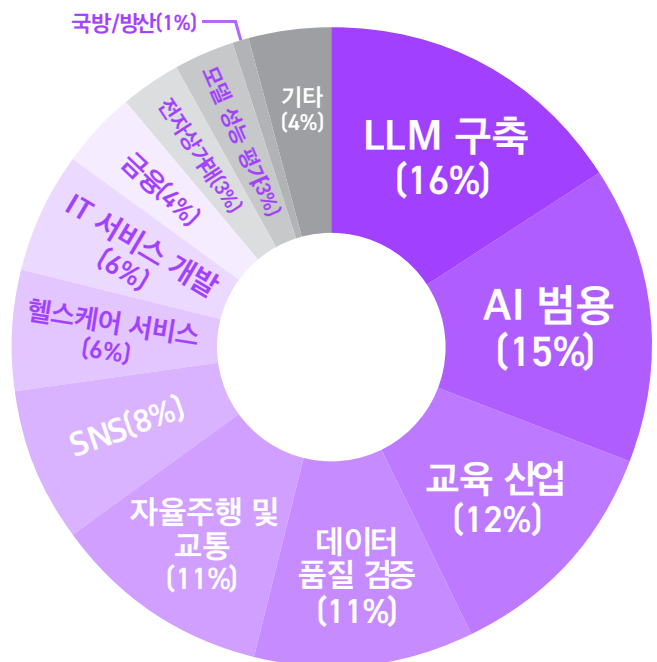
1.

여전히 IT 기업이 많은 데이터를 만들지만, 구축 목적은 훨씬 다양해졌어요.

클라우드웍스에서 지난 1년간 데이터 프로젝트를 진행한 전체 기업의 52%는 IT 기업입니다. 2020년 자사 데이터를 보면 전체 고객의 약 68%가 IT 기업이었는데요. 그때보다 조금 줄긴 했지만, 여전히 IT 기업이 비IT 기업에 비해 많은 데이터 프로젝트를 진행하고 있습니다. 다만 데이터 구축의 목적은 이전에 비해 많이 다양해졌습니다. 과거에는 시스템 및 인공지능 활용, 교육, 미디어 산업, 연구 등 몇 가지로 간단히 추려졌다면, 최근에는 LLM 구축을 위한 학습데이터뿐만 아니라 AI 범용(LLM 이외 다양한 목적으로 활용할 인공지능 데이터), 교육, 자율주행 및 교통, SNS나 헬스케어 서비스 개발, 금융 등 다양한 산업에 활용할 목적으로 데이터를 구축하고 있습니다. 특히 이미 구축한 데이터의 품질 검증 및 재가공이나 휴먼 피드백 데이터를 통한 모델 성능 평가를 목적으로 데이터를 만드는 경우도 눈에 띄게 증가했습니다. 실제로 산업 현장에서 저품질 데이터로 인한 다양한 이슈가 발생하고 있으며, 데이터 품질의 중요도가 점점 높아지고 있다는 방증이기도 하죠.



[표1] 고객사 종사업 비중



[표2] 데이터 구축 목적별 프로젝트 비중

Key Point 2.

고품질 데이터를 확보하고자 하는 움직임이 활발해요.

자사의 2020년 데이터와 비교해 보면 최근 1년간 진행한 프로젝트 수에 비해 구축한 데이터의 수는 줄어들고 있는데요. 데이터 구축량에 중점을 두기 보다는 고품질의 데이터를 확보하고자 하는 기업들이 많아지고 있기 때문으로 보입니다.

데이터 품질 확보에 있어 작업자는 가장 중요한 요소입니다. 이전에는 많은 양의 데이터 수집과 가공을 감당하기 위해 한 프로젝트에 투입되는 데이터 작업자의 수도 많았습니다. 하지만 최근에는 기업들도 작업자의 수보다 작업자 개인의 역량을 더 중요하게 생각하는데요. 크라우드웍스는 데이터 난이도, 산업군, 전문성 등 기업의 다양한 요구 사항에 맞는 데이터 작업자 구성을 위해서 자체 검증 단계를 거치고 있으며, 적합한 작업자를 선별하는 다양한 프로세스 또한 구축했습니다. 국내에서 유일하게 AI 데이터 전문가 과정(AIDE)을 개설해 교육하고 있는 기업으로서, 해당 자격증을 보유한 작업자 풀을 별도로 구성해 프로젝트를 진행하기도 했고요. 또한 데이터 구축 시 자동 검수 시스템을 적용하고 이후 전문 검수자를 통해서 확인하는 과정을 추가해 철저한 품질 검증 과정을 거치고 있습니다.

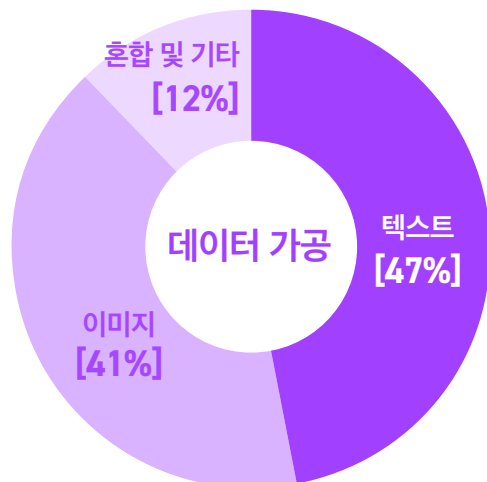
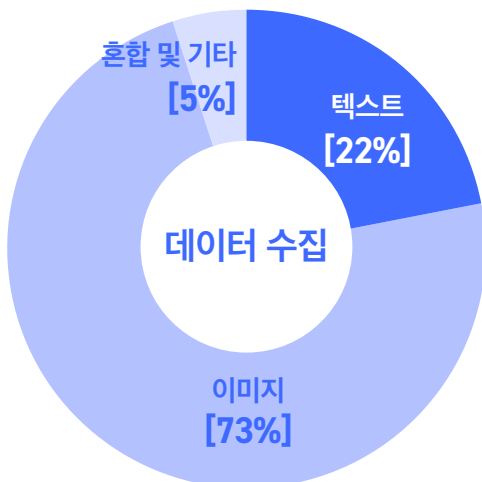
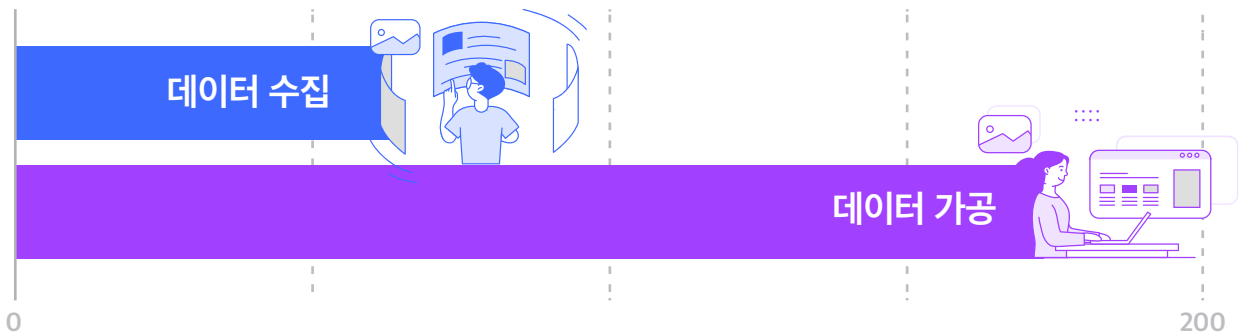


Key Point 3.

데이터 수집보다는 가공이 많고, 이미지와 텍스트 유형이 가장 많아요.

클라우드웍스를 통해 데이터를 구축하는 기업 고객은 데이터의 가공 수요가 수집에 비해 3배 정도 많았습니다. 다양한 경로를 통해 수집된 기업의 데이터 자산을 서비스 개발이나 AI에 활용하기 위해서는 가공 과정이 꼭 필요하기 때문인데요. 최근에는 LLM의 학습데이터 구축을 위해 필요한 텍스트 데이터 가공이 많고, 특히 기존 문서를 OCR 기술을 통해 자산화하는 경우가 많았습니다.

이미지 데이터의 수집과 가공은 자율주행, 교육, 헬스케어, 전자상거래, 플랫폼 사업 등 다양한 산업군에서 이뤄지고 있습니다. 최근에는 오디오, 동영상, 혹은 혼합 형태의 데이터 가공 역시 점점 늘어나는 추세입니다.



[표3] 데이터 수집/가공 및 구축 유형 비율

03

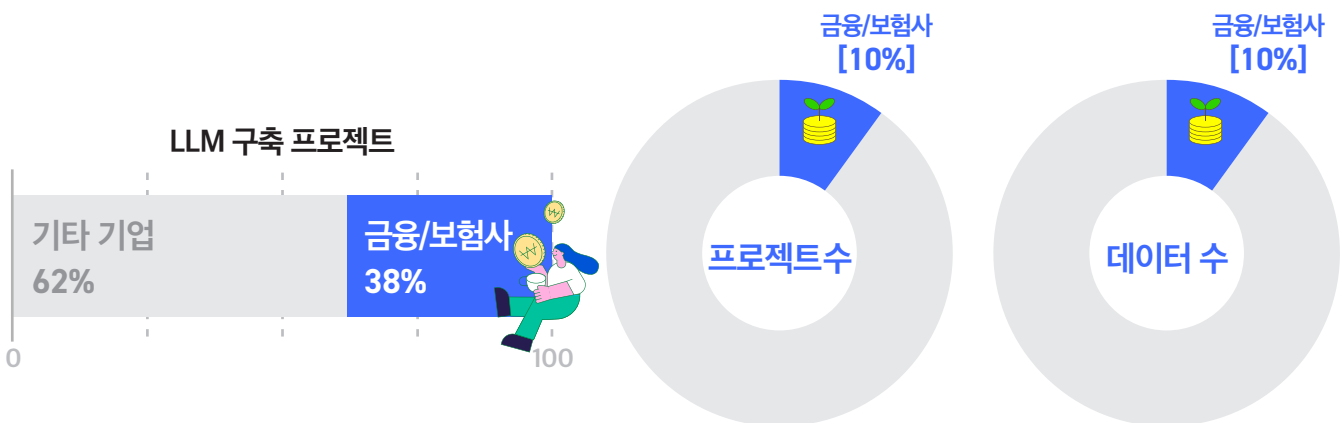
클라우드웍스가 살펴본
산업별 인공지능 데이터



1. 금융/보험 산업

가장 빠르게, 또 활발하게 방대한 데이터를 구축하여 시에 활용하고 있어요

앞서 [표1]에서 보이는 것과 같이 클라우드웍스가 진행한 데이터 프로젝트 고객의 15%는 금융/보험사입니다. 그런데 [표2] 데이터 구축 목적별 프로젝트 비중을 보면 4% 정도로 자칫 낮아보일 수 있는데요. 실은 금융/보험사들은 다른 어떤 산업군보다 방대한 정형, 비정형 데이터를 기반으로 사업 전방에 시를 활용하고자 하는 시도가 가장 활발합니다. 그래서 최근 1년간 금융/보험사들은 이미 과거에 차곡차곡 구축해둔 데이터를 기반으로 LLM 구축을 빠르게 시도하고 있습니다. 메타데이터를 좀 더 자세히 살펴보면 [표2]에서 금융으로 구분된 프로젝트뿐만 아니라 'LLM 구축'으로 분류된 프로젝트 중 38% 역시 금융/보험사가 진행하고 있는 프로젝트인 걸 알 수 있습니다.



[표4] 데이터 구축 목적 중 'LLM 구축'에서 금융/보험사 프로젝트 비중

[표5] 전체 프로젝트 및 데이터 수 중 금융/보험사 비중

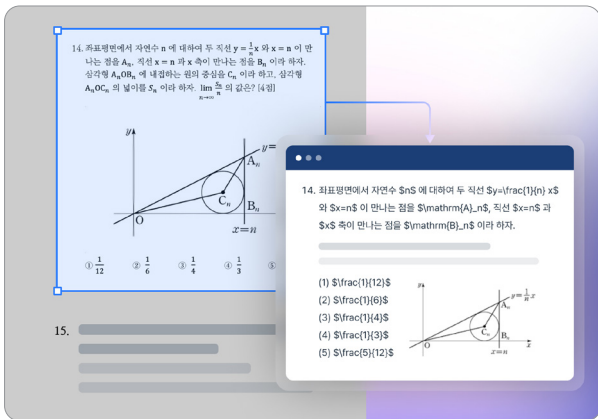
금융/보험 산업은 AI 도입 초기 단계부터 서비스에 활용할 데이터의 검토 및 설계, 가공 및 관리 운영 방안 수립에 많은 투자가 이뤄졌습니다. 최근에 금융/보험사의 자체 LLM 구축이 많아지면서 클라우드웍스는 금융/보험사가 구축하는 AI 서비스의 최신성 및 정확성 보장을 위해 기업이 가진 방대한 데이터를 검토하고 분석하여 안전하게 활용할 수 있도록 설계, 가공, 구축하는 프로젝트를 지속적으로 진행하고 있습니다. 특히 클라우드웍스 자체 Fine-tuning Data Framework를 통해 Instruction을 포함한 데이터셋을 구축하여 LLM 모델의 성능과 문제 해결 능력 향상에 핵심적인 역할을 하고 있기도 하고요.

또한 장기적인 측면에서 데이터를 효과적으로 관리하고 업데이트하기 위한 메타데이터 관리 전략과 체계도 제안하면서 정기적인 데이터 관리 업무도 진행하고 있습니다.

2. 교육 산업

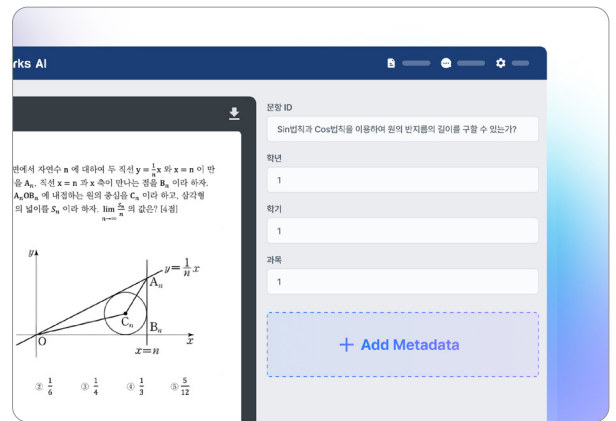
교육부 등의 정책에 따라 디지털교과서로의 전환 준비에 속도를 내고 있어요

교육 분야 역시 금융/보험 산업과 함께 인공지능 도입의 최전선에서 있습니다. 우리 교육부는 2025년에 초등학교 3·4학년, 중학교 1학년, 고등학교 1학년의 수학·영어 과목에 AI 디지털 교과서를 도입할 예정인데요. 해외 일부 도시에서 디지털 교과서를 도입한 사례는 있지만 중앙 정부 차원에서 공교육에 디지털 교과서를 정식 도입하는 것은 세계 최초입니다. 따라서 많은 교육 기업이 디지털교과서 서비스 모델 개발과 더불어 이를 위해 필요한 문항 OCR, 문항 메타 데이터, 신규 문항 및 컷 등을 빠르게 제작해나가고 있습니다.



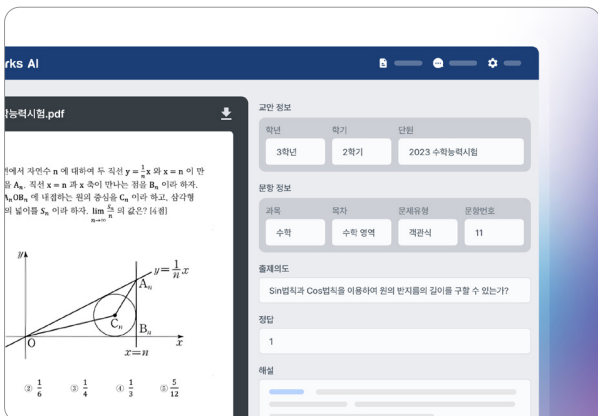
문항 OCR

하나의 도구로 이미지, 텍스트, 표, 수식을 한 번에 추출하여 작업 효율을 극대화합니다.



문항 메타 데이터 제작

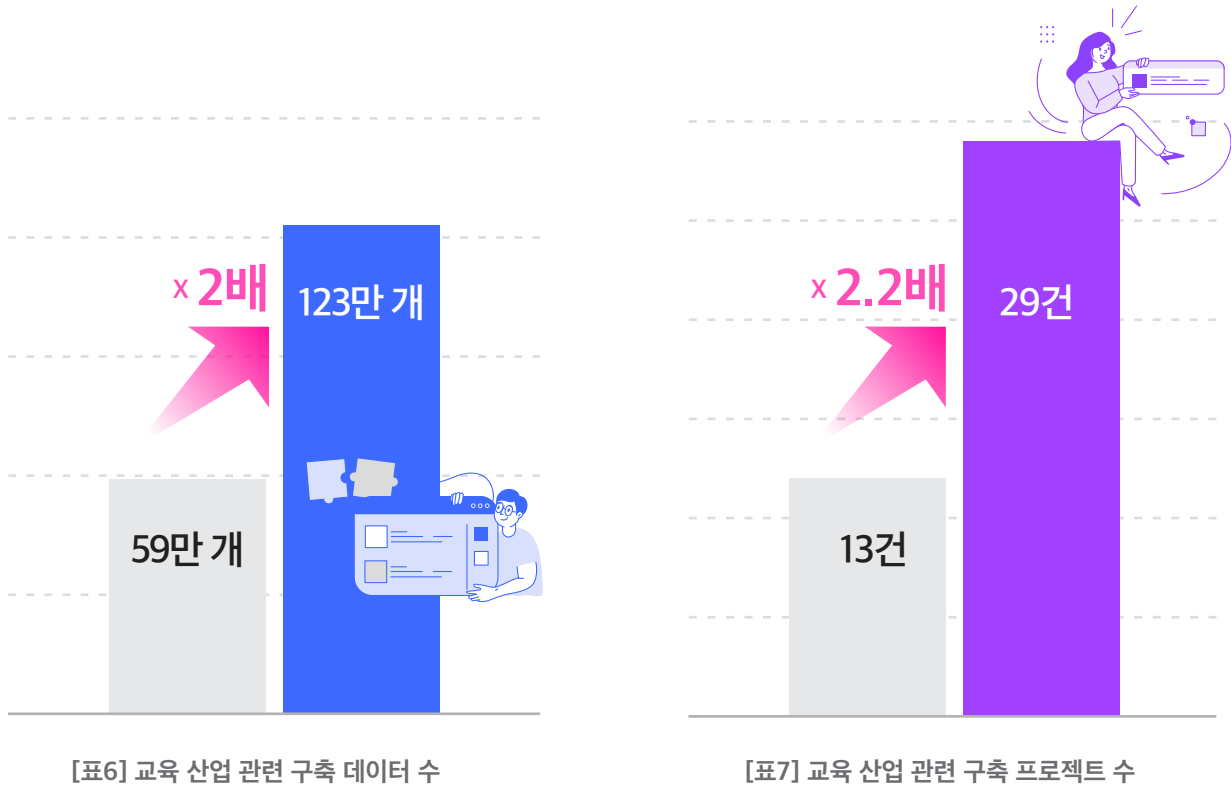
최적화된 툴을 활용해 문항과 메타 데이터를 동시에 생성하며 원하는 메타 데이터를 자유롭게 추가하여 맞춤형 데이터를 구축합니다.



문항 및 컷 제작

교육 전문가가 작업자로 참여해 문제, 정답, 해설, 그래프, 도형, 삽화 등의 이미지를 포함한 문항을 생성합니다. 고객사가 원하는 문항 정보 및 난이도, 평가 영역, 행동 영역 등의 메타데이터를 생성하여 DB화시켜 서비스에 활용하실 수 있습니다.

클라우드웍스가 지난 1년간 교육 산업용으로 구축한 데이터는 약 123만 개로 전체 구축 데이터의 약 9%를 차지합니다. 자사의 2020년 자료와 비교해 보면 59만 개에서 약 2배 이상 증가한 양입니다. 프로젝트 수도 13건에서 29건으로 2배 이상 증가했습니다.



또한 총 29건의 교육 산업 관련 프로젝트 중 20건이 문항 OCR 작업을 포함하고 있습니다. 신규 문항이나 컷을 생성하는 프로젝트도 빠르게 늘고 있으나 여전히 많은 프로젝트가 기존 문항 데이터를 가공하여 AI 서비스에 활용할 수 있도록 전환하는 데 초점을 맞추고 있습니다. 클라우드웍스의 60만 작업자 풀 중에서 교육 분야에 특화된 전문 작업자를 선별해 메타 데이터를 추가하거나 수정하는 작업도 빠르게 증가하고 있는데요. 이 역시 문항 데이터를 교육에 좀 더 다양하게 활용할 수 있도록 가공하는 것으로 추후 다양한 인공지능 서비스 개발이 기대되는 부분입니다.

3.

자율주행 및 모빌리티

시장에 대한 엇갈리는 투자... 여전히 고품질의 정밀 데이터 가공이 필수예요

최근 자율주행 시장은 여러 가지 이슈로 인해 투자에 대한 견해가 갈리고 있습니다. 작년 10월 GM의 자율주행 사망 사고 여파로 미국에서는 자율주행의 안전성에 대한 우려가 커졌는데요. 2020년만 해도 자율주행차의 상용화에 대한 기대로 빅테크들이 너도나도 투자에 뛰어들었는데 작년 말에는 구글, 테슬라 등 몇 개 기업을 제외하고는 많은 기업들이 투자 예산을 삭감하거나 사업을 접기도 했습니다.

하지만 올해 들어 다시 투자가 활성화되고 있습니다. GM이 다시 자율주행차를 가동하고 중국도 바이두 등이 정부의 강력한 정책적 지원에 힘입어 엄청난 속도로 대규모 자율주행 데이터를 수집하고 있습니다. 테슬라도 마침내 중국에서 FSD 테스트 주행 승인을 받았고 유럽에서도 자율주행 테스트 성공이나 시험 주행 지원 소식, 자율주행차 법령 마련 소식 등이 계속해서 나오고 있어 잠시 주춤했던 시장이 극적으로 활성화될 것으로 기대됩니다.

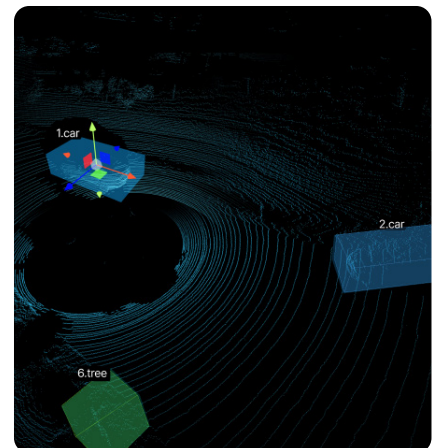
주춤했다고는 하지만 **클라우드웍스의 지난 1년간 데이터 프로젝트의 11%가 자율주행 및 모빌리티 산업과 관련되어** 있습니다. 데이터 수 역시 한해 **총 구축 데이터 수의 15%**에 달합니다. 경로 탐색, 고정 지물 인식, 변동/이동 물체 인식 등 자율주행의 목표 기능 구현에 필수인 고품질의 정밀한 데이터를 다양한 방식으로 가공하고 있습니다. 또한 LLM, AI 음성 인식 서비스등을 차량 인포테인먼트 서비스(IVI)와 결합하여 개인화된 이동 경험을 제공하는 것을 목표로 다양한 데이터를 구축 중에 있습니다.



<2D Object Detection>

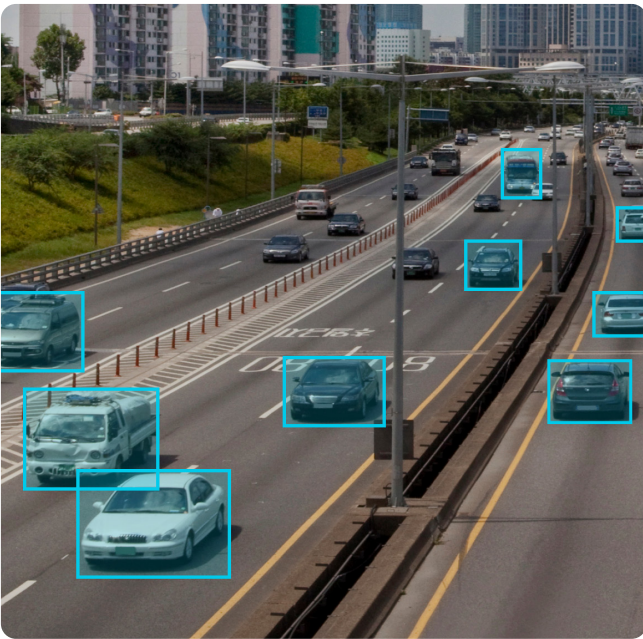


<2D Segmentation>

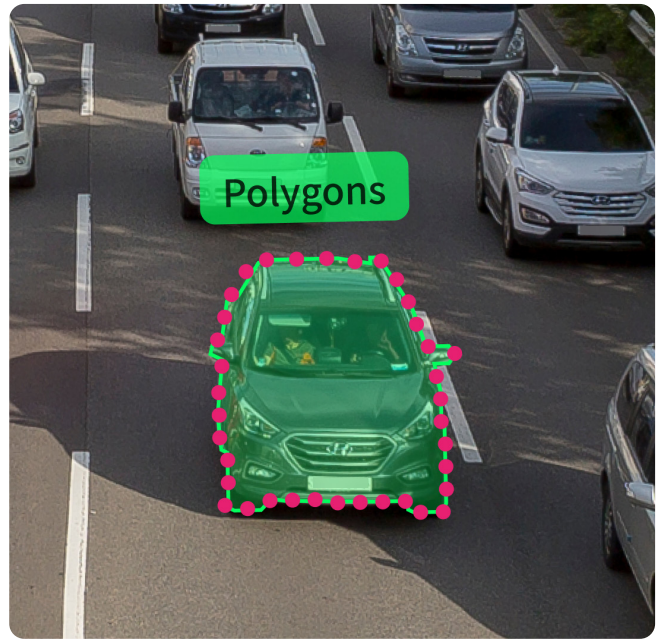


<3D Object Detection (LiDAR)>

데이터의 정밀함과 정확성이 매우 중요한 산업인만큼 최신 기술이 적용된 어노테이션 도구를 활용한 스마트 라벨링은 물론 고객사가 필요한 데이터에 적합한 도구가 다양하게 활용되고 있고요. 까다로운 대량의 데이터를 빠른 시일 내 가공하기 위해 전문 PM과 데이터 작업자들이 참여해 체계적으로 프로젝트를 관리하고 있어 고객의 만족도도 높습니다. 기업들이 다시 활기를 찾아가고 있는 만큼 내년이 더욱 기대되는 시장입니다.



<Bounding Box>



<Polygon>

4. 국방/방산

AI 기반 국방 기술 혁신을 놓고 치열한 주요 국가 간 경쟁을 대비해요

국방에서 AI 도입은 미래 국방력과 밀접히 연결됩니다. 부족한 병력 수급 문제의 해결을 비롯해 점점 복잡해지는 현대전에 대비한 첨단 전투력 강화, 효율적인 자원 관리 등 인공지능 기술이 해결할 수 있는 많은 국방 이슈가 있기 때문입니다. 그래서 국방 AI 기술은 전 세계적으로 국가 간 경쟁이 가장 치열한 분야인데요. 미국은 2018년에 국방부에서 AI 전략 발표 직후 20억 달러를 투자했고 현재 국방 AI 연구 개발 예산만 한 해에 18억 달러에 이릅니다. 중국도 최근 수십억 개의 빅데이터 분석과 자가 학습 능력을 갖춘 AI 기술을 접목해 지능화군을 건설하겠다는 목표를 가지고 미국에 맞먹는 투자를 진행하고 있고요.

우리 국방부는 지난해 3월 새롭게 ‘국방혁신 4.0 기본계획’을 발표하고 이를 기반으로 “24~28 국방혁신 추진 계획”을 수립했습니다. 목적은 군사 전략부터 핵심 전력, 국방과학기술 분야 등을 혁신하여 ‘경쟁우위의 AI과학기술강군을 육성’하는 것입니다. 클라우드웍스는 ‘육군 지능화 감시체계 데이터셋 구축’ 등 몇 가지 국방 데이터 사업에 참여하여 지난 1년간 약 40만 개의 관련 데이터를 구축했습니다.

프로젝트 사례

육군 지능화 감시체계 이미지 인식 데이터 구축

프로젝트 목적

- 무기 체계 이미지 인식 시스템을 위한 데이터 구축

작업 내용

- 전차 실물 촬영, 3D 모델링, 3D 프린팅,
FLA 모델 유형의 이미지 데이터 구축 및 이미지 바운딩



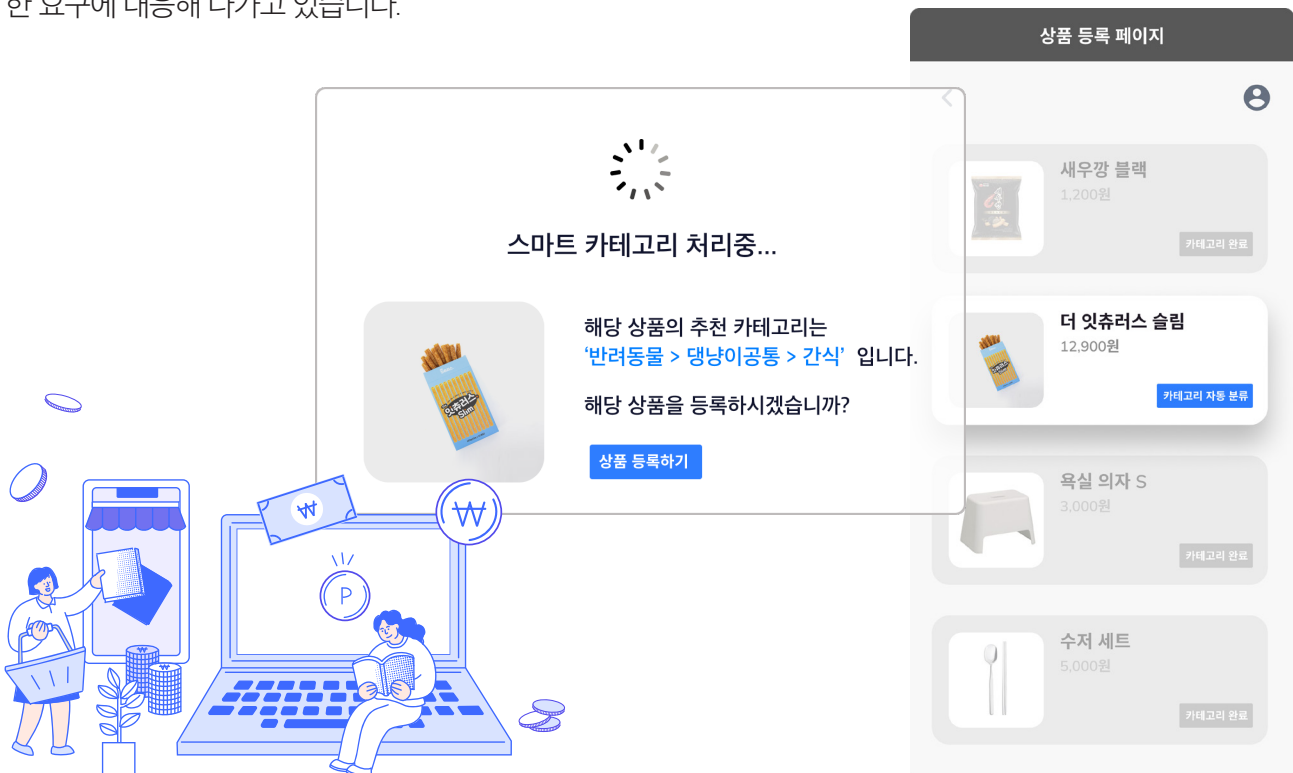
국방 분야 데이터는 특히 작업상 보안 및 안정성 검증과 정밀한 데이터 가공 역량이 중요한데요. 클라우드웍스는 전담 PM 및 전문 데이터 작업자를 파견하고, 국방부의 보안 규정을 준수해 데이터를 수집하고 철저히 안전망 내에서 가공하는 등 안전하면서도 정밀한 고품질의 데이터를 구축하고 있습니다. 경쟁력 있는 국방 AI 체계 수립을 위해서는 고품질 데이터 구축이 반드시 선행되어야 하므로 앞으로는 데이터 수집과 가공뿐만 아니라 데이터 플랫폼, 데이터 가공 기준 제시 및 인프라 구축, 최신 AI 기술 활용 제안 등 다양한 협력 방안을 논의해 나갈 예정입니다.

5. 주목할 만한 산업과 데이터

대량의 데이터 분류를 자동화하고 검증하려는 이커머스 산업

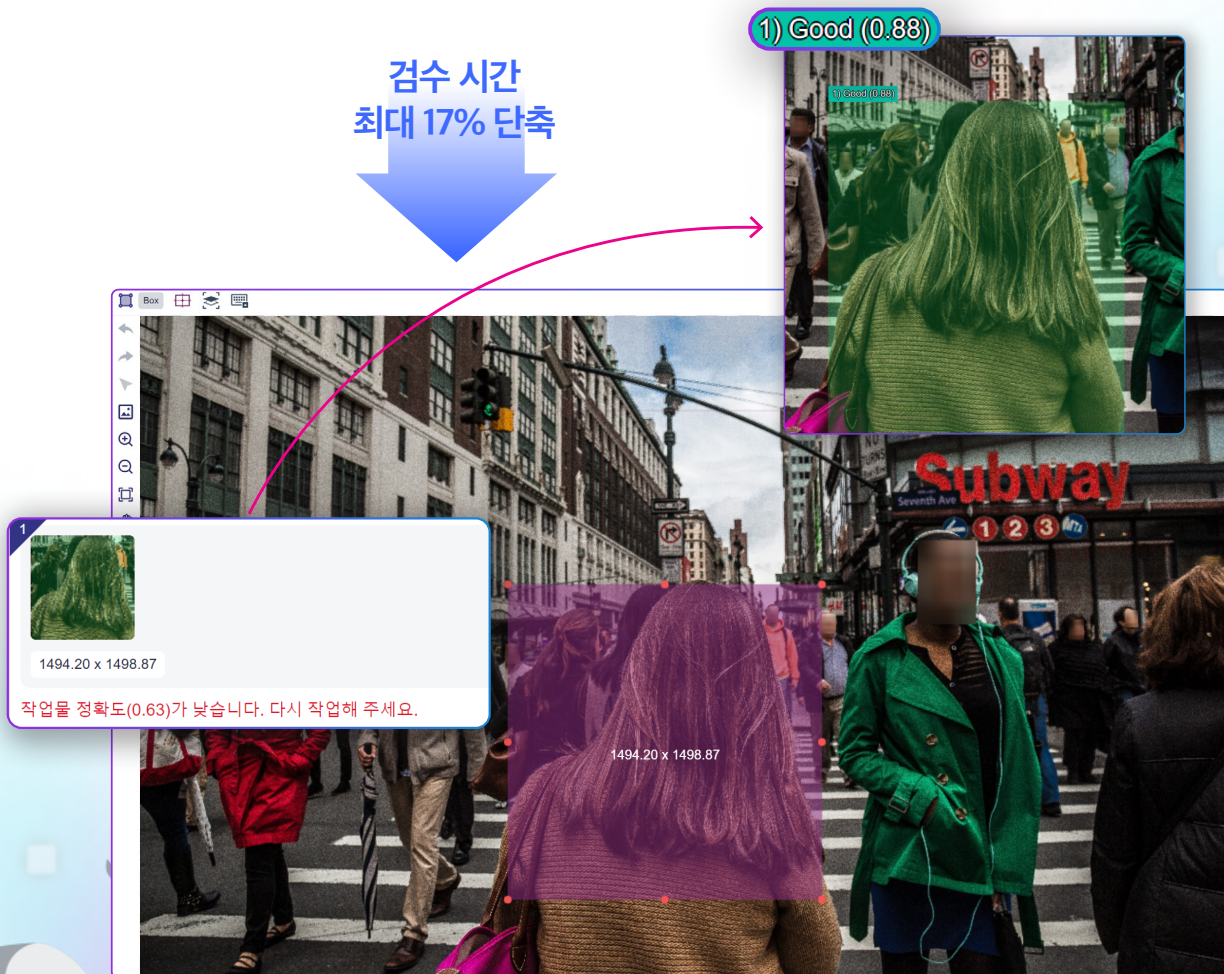
이커머스 산업은 자사 플랫폼을 통해 매출, 광고, 재고, 고객, 물류 등 세분화하면 100종 이상의 다양한 데이터를 매일 대량 생산하고 수집합니다. 말 그대로 숨만 쉬어도 쌓이는 엄청난 분량의 산재한 데이터들을 어떻게 분류하고 관리할 것인가가 큰 이슈죠. 특히 최근에는 기업들이 데이터를 인공지능 서비스 등에 적극적으로 활용하기 위해 장기적인 관점에서 데이터 거버넌스를 추진하면서 대규모 데이터를 정제하는 프로젝트가 생기고 있으며, 앞으로 더 늘어날 것으로 예상됩니다.

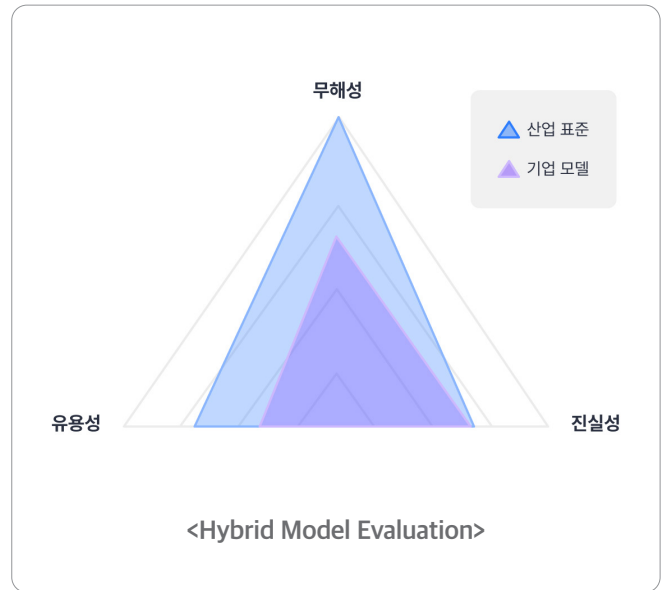
특히 이커머스는 많게는 억 단위의 SKU(Stock Keeping Unit, 상품 재고 관리 코드)가 있고 제품 카테고리 또한 세분화 되면서 수 백 개에서 수만 개 단위로 추산됩니다. 매일 수만 개의 신규 상품이 수동 등록되는 경우도 많은데 복잡한 카테고리 구성 때문에 오분류(혹은 미분류) 문제가 자주 발생하고요. 따라서 오분류를 최소화하여 고객 경험을 개선하고 데이터의 활용 가치를 높이기 위해 제품 카테고리를 자동 분류하고 검증하는 AI 모델을 개발하는 케이스가 늘고 있습니다. 클라우드웍스는 이를 위해 제품의 메타 정보, 이미지, 리뷰 속성 등을 태깅하거나 제품명, 파일명 등을 수정하고 검수하는 등 기본적인 데이터 프로젝트 외에도 LLM을 활용해 분류를 자동화하는 서비스를 개발하는 등 기업의 다양한 요구에 대응해 나가고 있습니다.



만든 '데이터·모델'도 다시 보자, 품질 검증 프로젝트

산업군에 상관없이 많은 기업들이 이미 구축한 데이터 중에서 정밀함, 정확도 등의 측면에서 데이터의 품질이 떨어지는 것으로 파악되어 수정이 불가피한 경우가 있습니다. 클라우드웍스는 이러한 데이터 품질 검증 프로젝트를 진행해 약 37,000개의 관련 데이터를 구축했습니다. 특히 공공기관 주관 아래 기업이 AI 공공데이터 구축 사업을 추진할 경우 반드시 별도 기관을 통해 품질 검증 과정을 거쳐야 하는데, 검증을 통과하지 못한 데이터의 경우 보완 조치가 시행되어야 합니다. 클라우드웍스는 한국정보통신기술협회(TTA)와 함께 인공지능 학습용 데이터 유형별 의미적 정확성 검사를 다년간 수행해 왔는데요. 특히 명확한 품질 기준과 데이터 유형 및 난이도 등에 따라 최적화된 검사 절차(OBR, Readiness Check), 일반 검사/자동 검사(ML assisted) 도구와 전문 검수 인력을 갖추고 있어 데이터의 정확도, 정밀도, 재현율 등 검사 결과에 있어 높은 신뢰도를 확보하고 있습니다.





데이터 품질 검증과는 별개로 LLM이나 관련 AI 서비스 구축 이후 모델의 답변 품질과 성능을 검증하고, 모델의 고도화를 위한 데이터셋을 구축하는 경우도 많았습니다. 먼저, 전문 인력을 구성하여 답변의 유용성, 진실성, 무해성, 편향성 등 다양한 측면을 검토해 모델 및 서비스가 최대한 많은 사람의 니즈를 충족하는지, 부족한 부분은 없는지 체계적으로 평가를 진행합니다. 이후 평가 결과에 따라 부족한 학습 데이터를 추가 구축하고 있습니다. 기업이 사용하는 LLM은 테스트와 평가를 통한 지속적인 관리가 필수인 만큼 클라우드웍스는 Hybrid Red-Teaming이나 Hybrid Model Evaluation 등 다양한 검증 방법을 제공하고 있습니다. E



www.crowdworks.ai
sales@crowdworks.kr