datamaker

2025

데이터메이커 맞춤형 데이터 구축 컨설팅 소개서

Copyright © datamaker All rights reserved.

Contents

01 AI 개발의 한계, '데이터 절벽'을 넘어서기 위하여 3P

데이터 부족 현상 극복과 성능 향상을 위한 고품질 데이터 구축

02 데이터 구축에 컨설팅이 필요한 이유 4P

데이터메이커에서 제공하는 데이터 구축 컨설팅의 중요성

03 데이터메이커의 구축 서비스가 특별한 이유 6P

도메인 특화 데이터셋 구축 서비스

인하우스 & 글로벌 데이터랩

자체 개발 어노테이터의 기술력

04 전 산업에 적용되는 데이터메이커 컨설팅 & 구축 서비스 11p

Use Case

05 AI 도입을 위한 최적의 파트너 데이터메이커 13P

06 데이터 구축 상담 및 문의 안내 14P

AI 개발의 한계, '데이터 절벽'을 넘어서기 위하여

AI 기술이 빠르게 발전하면서, 단순한 범용모델을 넘어 특정 산업과 분야에 최적화된 AI 모델이 요구되고 있습니다. 의료, 금융, 법률, 제조 등 고도화된 도메인 지식이 필수적인 산업에 특화된 AI 개발을 위해서는 일반적인 수준의 데이터를 넘어 해당 도메인에 특화된 전문 데이터가 필요하며, AI가 올바르게 학습하고 신뢰할 수 있는 결과를 제공하려면, 해당 **산업의 전문성을 반영한 정확하고 정제된 고품질 데이터 구축**이 필수적입니다.

그러나 현재 AI 업계는 데이터 고갈 문제에 직면해 있습니다.
AI 학습 속도의 가속화가 학습용 데이터 증가 속도를 앞지르면서, 기존의 공개 데이터와 크롤링 데이터는 한계를 드러내고 있으며, 개인정보 보호와 규제 강화로 인해 새로운 데이터를 확보하는 것이 점점 더 어려워지고 있습니다.
특히, AI의 기능과 역할이 다양해지면서 데이터 작업에도 **도메인 지식이 요구되는 많아지고 있어, 일반 작업자가 아닌 특정 분야의 전문가가 직접 데이터 구축 과정에 참여해야 하는 경우가 많은 상황입니다.**
예를 들어, 의료 AI의 경우 의료 전문가가 직접 데이터 검수와 라벨링을 해야 하는 경우가 많아지는 것입니다.

이러한 상황에서 **기업 자체적으로 고품질 데이터를 구축하는 것은 쉽지 않습니다.**
방대한 양의 데이터 수집 뿐만 아니라, 정확한 가공, 구조화, 정제, 라벨링 과정이 필수적이며, 이를 위해서는 전문 기술과 인력이 필요합니다. 또한 효율적인 데이터 구축 전략 없이는 시간과 비용이 과도하게 소모될 뿐만 아니라, 최종 데이터 품질을 일정 수준 이상으로 개선하는 것도 어렵습니다.

결국, 어떤 데이터를 어떻게 구축할지에 대한 고민의 해답을 찾지 못하게 되면, 학습에 활용할 수 있는 데이터가 바닥나는 **데이터 절벽 문제**에 빠지게 됩니다.

기업이 인공지능 개발에 필요한 데이터를 구축할 때 생기는 다양한 고민들

지금 우리에게 어떤 데이터가 적합한지 모르겠어요

- AI 모델로 원하는 성능을 내기위해서 어떤 데이터가 필요한지 모르겠어요.
- 현재 우리에게 필요한 데이터의 우선 순위를 정하고 싶어요.

데이터 수집부터 정제, 라벨링까지… 이걸 어떻게 다 하죠?

- 특정 분야에 전문적인 지식과 경험 역량을 가진 작업자가 필요해요.
- 데이터 품질을 높고 일관되게 유지할 수 있도록 작업의 전문성이 보장되어야 해요.

이게 적절한 시간과 비용이 맞을까요?

- 예산 내에서 최대한 효율적으로 구축하는 방법을 찾고 싶어요.
- AI 개발 일정에 최대한 문제가 되지 않도록 빠르게 데이터 구축이 필요해요.

데이터 구축에 컨설팅이 필요한 이유

데이터 구축 컨설팅은 앞서 언급한 기업들의 고민을 해소하기 위해 가장 필요한 솔루션입니다.
데이터 구축은 단순한 데이터 수집이 아니라, 기업이 필요로 하는 최적의 데이터 구조를 설계하는 것부터 시작됩니다. 데이터메이커는 다양한 산업에서 축적한 데이터 구축 경험을 바탕으로, 고객의 AI 모델이 최대의 성능을 발휘할 수 있도록 맞춤형 데이터의 구조와 구축 전반의 설계를 돕고 최적의 전략을 제공합니다. 이러한 데이터 구축 컨설팅이 선행되어야, 인공지능 개발의 성공 가능성을 높이고 데이터 구축에 소요되는 비용의 효율을 극대화할 수 있습니다.

요구사항 정의
인공지능 학습에 필요한
학습용 데이터의 세부 명세 정의

수집 방안 설계
필요한 데이터를 적법한 절차로
구축하기 위한 최적의 출처 및 방안

가공 절차 설계
데이터 정제, 라벨링, 검수
세부 절차 설계

작업 인력 배정
최적의 비용과 품질을 위한
전문 작업 및 검수 인력 설계

운영 인력 배정
유관 분야의 데이터 구축 경험을
보유한 운영 PM 배정

저작도구 세팅
라벨링에 필요한 저작도구
기능 커스텀 제공

납품 산출물 정의
최종 데이터의 포맷과
납품 형식 및 상세 정보 설계

데이터메이커를 통해 해결하는, 인공지능 학습용 데이터 구축 문제

AI에 적합한 데이터에 대한 정확한 정의

- 도메인 데이터셋 구축 노하우를 반영한 맞춤 제안 제공
- 업계별 최적의 데이터 유형과 수집 방안 컨설팅 지원

데이터 구축 전 과정을 체계적으로 지원

- 정밀한 가공이 가능한 전문 작업 인력 및 풍부한 경험의 운영 인력 투입
- 체계적인 데이터 처리 시스템으로 일정 내 안정적 구축

비용과 일정에 최적화된 데이터 구축 솔루션 제공

- 비용 대비 최적의 효과를 내는 최적의 데이터 구축 솔루션 제공
- 시스템 적용 일정과 예산 맞춘 효율적인 데이터 처리 및 납품 프로세스

datamaker

데이터메이커의 구축 서비스가 특별한 이유

국내 주요 대기업부터 기관 및 학계와 스타트업까지, 다양한 기업들이 선택한 데이터메이커의 데이터 구축 서비스는 오직 데이터메이커에서만 제공 가능한, 차별화된 **컨설팅 기반 고품질 데이터 구축 서비스**입니다.

다수의 협업으로 검증된 서비스

대기업 중견기업	SK telecom, LG AI Research, LG Energy Solution, LG전자, LG화학, LG생활건강, 삼성중공업, HYUNDAI, NAVER Cloud, NAVER LABS, posco, posco ict, 한화시스템, zum internet, kt cs, NH투자증권, LS ELECTRIC, ESTsoft
기관 및 지자체	KETI 한국전자기술연구원, NIA 한국지능정보사회진흥원, ETRI, K data, nipa 정보통신산업진흥원, KITECH 한국생산기술연구원, KIMM 한국기계연구원, 한국전력공사, 대전광역시, 대전교통공사, KICT 한국건설기술연구원, DiCIA 대전정보문화산업진흥원, 지능정보산업협회, 대전창조경제혁신센터, KATECH 한국자동차연구원, RIST, 대전테크노파크, 개인정보보호위원회, AI Hub, 국방과학연구소, 치안정책연구소
학계	KAIST, 서울대학교, 고려대학교, 동국대학교, 성균관대학교, 충남대학교, 서정대학교, DUT 두원공과대학교, 연세대학교, 한밭대학교, 한양대학교, SNUH 서울대학교병원, CNU 충남대학교병원, 고려대학교안산병원, 강원도립주의료원
중소기업 스타트업	Dable, banksalad, RideFlux, MetriX, 사이냅스프트, MOBILTECH, SOLUGATE, KRISTIN, 365mc, DRONEID, LOCS, 이야기연구소, ecoin energy, seabank, ENABLE FIND, SORIZAVA, peter pet, SEADRONIX, ClueMetic, GENESIS BLOCK, D.HIVE, PIRECO, GOPIZZA

오직 데이터메이커만의 특장점

도메인 특화

전문성을 갖춘 작업자들이 데이터 구축에 참여하여, 산업별 특성을 반영한 정확한 데이터셋을 구축

인하우스 작업자

체계적인 교육과 최적화된 작업 환경을 제공하여 데이터 품질을 일관되게 유지

가나 데이터랩

효율적인 프로세스와 최적화된 운영 방식으로 고품질 데이터를 합리적인 비용으로 구축 가능

데이터 라벨링 저작도구

데이터 구축 전 과정 (수집, 가공, 검수, 품질관리)을 자동화하여 빠르고 정확한 데이터 제공

도메인 특화 데이터셋 구축 서비스

고품질 인공지능 학습용 데이터 확보를 위해서는, 인공지능이 적용되는 산업에 대한 도메인 지식이 반드시 필요하며 특히, LLM 모델에 대하여 작업자의 피드백을 기반으로 강화학습을 진행하는 RLHF(Reinforcement Learning from Human Feedback) 작업과 같은 최신 모델 학습에 있어서는 전문 분야의 지식과 경험을 가진 인력이 참여하는 것이 특히나 더 중요합니다.

데이터메이커는, 누적 400건 이상의 다양한 산업 별 프로젝트를 수행한 경험을 보유하고 있으며, 이를 토대로 각 산업 분야에 적합한 **도메인 특화 데이터셋**을 제공할 수 있습니다. 최근에는, LLM 모델의 적용 분야와 임무가 다양해지면서 코딩, 수학 문제 풀이, 작문, 창작 등 다양한 전문역량을 요구하는 데이터 구축 작업 또한 수요가 많아지고 있으며, 데이터메이커는 국내 주요 LLM 기업들에게 해당 데이터셋을 공급하는 주요 협력사로 성장하였습니다.

산업 별 누적 계약 건

410건 이상

다양한 산업별 경험 보유

교통 및 물류	엔터테인먼트	의료	에너지	제조
58	47	49	43	41

통신	기타	교육	안전	농축수산	국방	관제	스포츠
41	31	28	23	15	15	12	9

도메인 특화 데이터 for LLM

최근 데이터메이커를 통해 가장 많이 진행되는 데이터 구축 프로젝트는 역시나 LLM 모델 학습을 위한 도메인 특화 데이터 구축입니다.

수학 문제 풀이, 외국어 통번역, 전문 분야 지식기반 대화까지 정말 상세한 분야마다 LLM이 활용되면서 이를 위한 도메인 특화 데이터셋에 대한 구축 수요도 크게 증가하고 있습니다.

데이터메이커는 이러한 수요를 충족시킬 수 있는 폭넓은 산업 분야별 전문 인력을 바탕으로 RAG 구축, RLHF, Red Team 등 다양한 LLM 전용 데이터 솔루션을 제공하고 있습니다.

LLM 분야 주요 고객사

NAVER — 하이퍼클로바 X

추론, 외국어, 문제풀이, 안전성 평가 등 RLHF 및 검증 작업

LG AI Research — EXAONE

한국어, 영어 대화 데이터 구축 및 검증, 모델 안전성 평가

SK telecom — 에이닷

LLM 서비스를 위한 멀티모달 학습용 데이터 구축

인하우스 & 글로벌 데이터랩

고객사의 데이터 구축 수요를 맞춤형으로, 가장 적절한 비용과 기한 내에 구축하기 위해서는, 데이터 구축에 참여하는 인력을 어떻게 구성하는지가 매우 중요합니다. 데이터메이커는 인하우스 기반의 보안 관리와, 가나에 위치한 자회사를 활용하여 프로젝트의 목적, 데이터 유형, 예산 규모에 따라 최적의 구축 방식을 제공합니다.

국내 인하우스 기반의 작업 운영

데이터 구축 프로젝트에서 일반적으로 크라우드워커를 활용한 방식이 많이 사용되지만, 이 방식은 작업자의 숙련도 편차, 비대면 관리로 인한 품질 관리의 어려움, 민감 데이터 보안 취약, 고난도 작업의 정밀성 부족 등 여러 한계를 갖고 있습니다. 특히 프로젝트별 도메인 이해도가 필요한 산업 데이터 구축의 경우, 재택근무가 많은 크라우드 소싱 방식으로는 한계가 발생할 수 밖에 없습니다.

데이터메이커는 이러한 문제를 해결하기 위하여, 인하우스 데이터랩의 상주 인력을 기반으로 운영하고 있습니다. 전문 작업자들은 프로젝트 별 체계적인 교육과 사전 테스트를 통해 작업 기준과 도메인에 대한 깊이 있는 이해를 갖춘 후 투입되며, 이후에도 실시간 피드백과 체계적인 검수 시스템을 통해 데이터의 정확성과 일관성을 유지합니다.

데이터메이커 인하우스의 보안 환경과 고성능 장비, 검증된 인력으로 민감한 데이터와 고난도 산업 프로젝트에도 유연하고 안정적인 구축 환경을 제공합니다.

인하우스

보안 강화
폐쇄망과 내부 통제 시스템을 통한 민감 데이터 보호

작업 환경 최적화
고성능 PC 기반의 전문 가공 작업

실시간 피드백 체계
상주 인력 간의 원활한 커뮤니케이션을 통한 작업 품질 향상

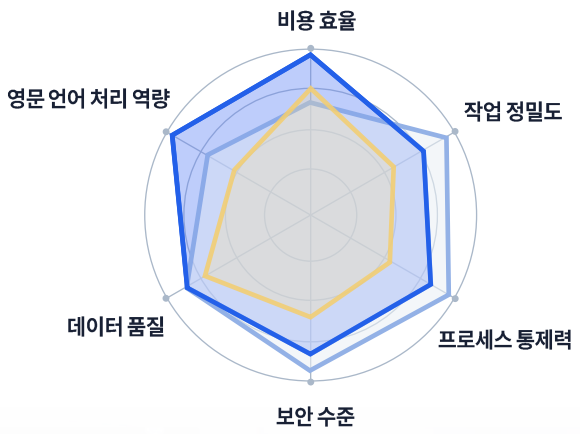
유연한 근무 배치
파견 근무 등 프로젝트 상황에 맞춘 근무 유형 설정 가능

가나 데이터랩

아프리카 가나에 위치한 Ghana Data Lab에서는 130명 이상의 전문 인력인 비용 효율성과 고정밀 작업을 동시에 만족하는 글로벌 대응형 데이터 구축 솔루션을 제공합니다.

가나는 영어를 공용어로 사용하는 국가로, 영문 기반 데이터 구축에 최적화되어 있으며, 풍부한 인재풀과 합리적인 인건비 바탕으로 높은 수준의 작업 품질을 안정적으로 확보하고 있습니다.
올해로 7년째 다양한 데이터 구축 프로젝트를 경험한 데이터메이커의 가나 데이터탭에서는 철저한 품질 관리 및 운영 체계를 통해 영문 기반 AI 학습용 데이터 구축, 복잡한 이미지 데이터 라벨링 등 다양한 기업의 고난도 니즈를 충족시키고 있습니다.

워크포스 유형별 품질·효율성 분석



가나

비용 효율성과 영문 처리에 강점을 가진 글로벌 대응형 고정밀 인력

인하우스

품질, 보안, 통제력 등 전반적인 균형과 안정성이 뛰어난 전담 전문 인력

크라우드워커

유연성과 접근성은 높지만 정밀도와 보안 면에서 한계가 있는 인력

국내와 해외에서 지원하는 폭넓은 Workforce를 제공하여, 고객의 다양한 수요를 충족합니다

크라우드 워커 + 데이터랩 인하우스
국내 전문 인력 + 가나 현지 인력
각 workforce 별 특장점

고객의 수요 및 상황을 고려하여
가장 적합한 작업자 투입 옵션과
효율적인 데이터 구축 견적 산정

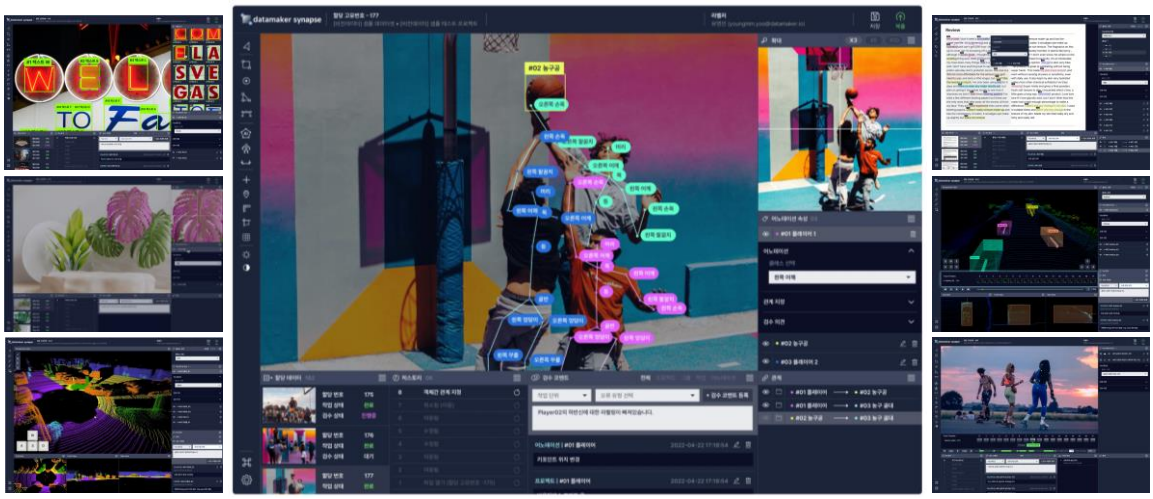
Man Hour $3 으로
현지 임금 대비 높은 소득 제공과
고품질 데이터를 저렴한 가격에 확보

자체 개발 어노테이터의 기술력

데이터 구축의 핵심 Key 어노테이터

데이터 라벨링을 위한 필수 도구이자 SW인 어노테이터(annotator)는 다양한 기능 지원을 통해 작업 효율을 극대화 시켜, 데이터 구축에 소요되는 비용을 절감시키는 측면에서 매우 중요한 역할을 합니다.

특히, 3D 센서에서 취득된 PCD(Point Cloud Data)데이터, LLM 학습을 위한 Prompt 데이터 등은 애초에 해당 데이터를 다룰 수 있는 어노테이터를 보유하고 있지 않은 곳도 많기 때문에, 어노테이터의 기능 및 기술 수준이 데이터 구축 과정에 있어 중요한 요소가 됩니다.



80개 이상의 기능을 보유한 datamaker synapse

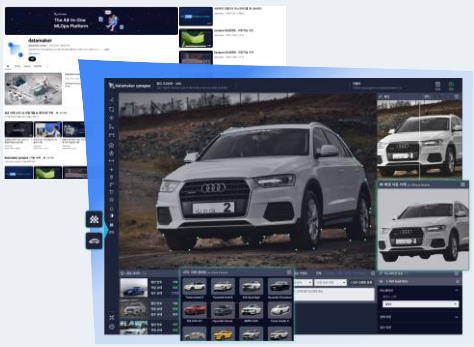
데이터메이커의 자체 개발 어노테이터 datamaker synapse는 80가지 이상의 기능을 보유한 국내 최고 수준의 어노테이터로, GS(Good Software) 인증 1등급을 획득하면서 성능을 입증하였습니다.

특히, AI 모델 기반의 Auto 라벨링 기능은 인력의존도를 낮추는 전처리 프로세스를 통해 라벨링 작업 속도를 향상시키고 비용을 절감시키는 핵심 기능입니다.

다양한 유형의 데이터 지원

LLM
센서
이미지
비디오
오디오
텍스트

자체 개발 Auto Labeling 툴



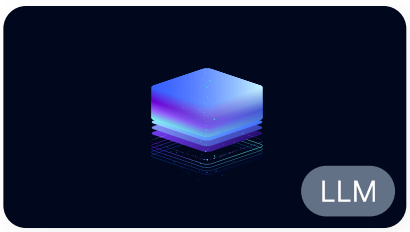
기능 영상 보러가기

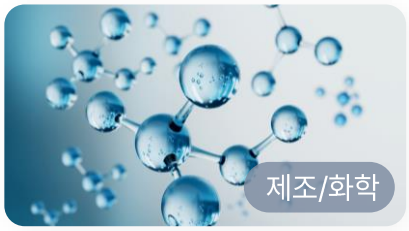
전 산업에 적용되는
데이터메이커 컨설팅 & 구축 서비스

AI 학습용 데이터 구축, 산업별 특화 데이터셋, 글로벌 프로젝트까지
데이터메이커는 다양한 기업들과 함께 성공적인 데이터 구축 프로젝트를 수행해왔습니다.
실제 적용 사례를 통해 프로젝트의 완성도와 효과를 직접 확인해보세요.

→ 데이터 구축 사례 바로가기

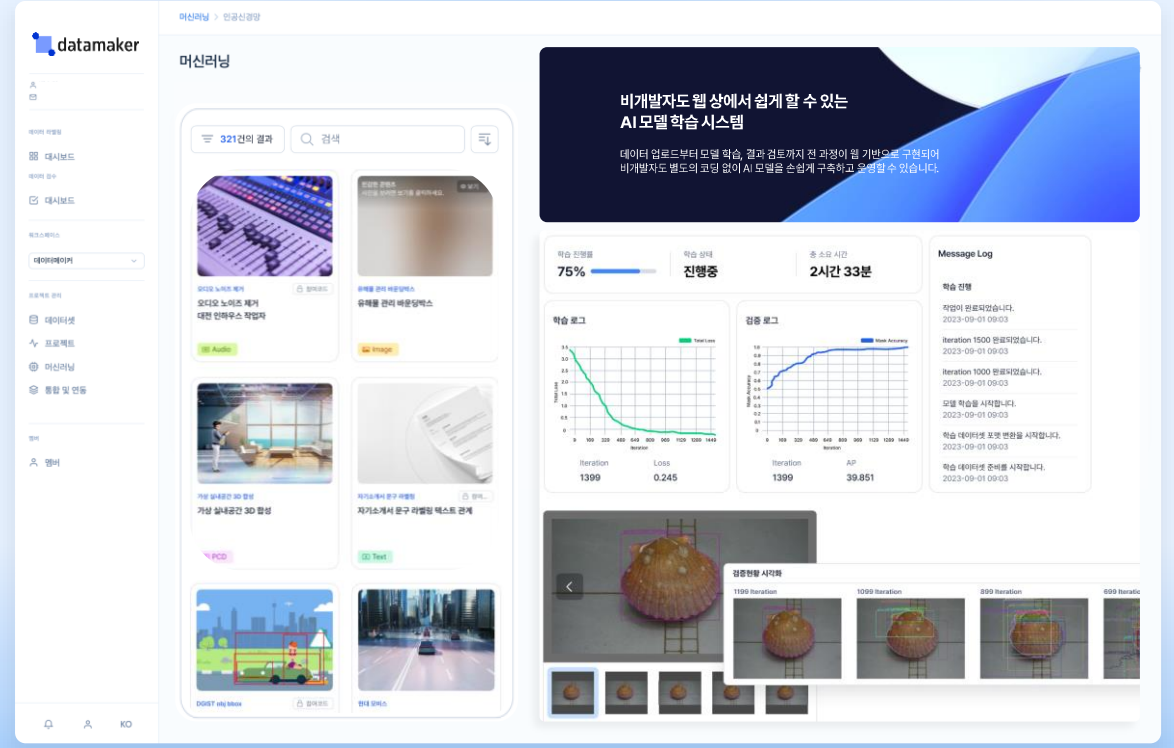
Use Case





카드를 선택해 다양한 Use Case 별 성공 사례를 확인해보세요.

AI 도입을 위한 최적의 파트너 데이터메이커



AI 개발 · MLOps

데이터메이커는 All-in-one MLOps 플랫폼 datamaker synapse를 기반으로 AI 모델 개발 서비스와 MLOps 플랫폼 납품, AI Agent 솔루션 개발까지 모든 과정을 지원합니다.
맞춤형 AI 개발과 내재화, 향후 관리를 위한 플랫폼 도입까지. AI 도입의 모든 여정을 데이터메이커와 함께 하세요.

AI 개발 서비스

- **SYNAPSE-Vision** 비전 분야 AI 모델 개발
- **SYNAPSE-Prediction** 예측 분야 AI 모델 개발
- **SYNAPSE-Language** 자연어 분야 AI 모델 개발

고객사 맞춤형 AI 모델 개발 서비스로 컨설팅을 통해 AI 도입 목적부터 함께 정립하며, 이에 부합하는 모델 학습용 데이터 구축과 AI 모델 개발까지의 모든 과정을 지원 대표적인 예시로, 충남대학교병원, 강원도원주의료원 등에 납품한 '의료 특화 AI 모델'과 119-112 통합 신고 시스템 전용 'LLM&자연어 분야 모델', CCTV 관제 센터-스마트양식장 등에 도입된 '예측 분야 모델 개발' 사례가 있음.

MLOps 납품

- **No Code** 비개발자도 사용 가능한 플랫폼
- **On-Premise** 폐쇄망 환경에 도입 가능한 플랫폼
- **Annotator** 라벨링 작업부터 모델 개발까지 한번에

datamaker synapse는 데이터 관리부터, 모델 학습, 배포까지 모든 과정을 단일 플랫폼 내에서 진행할 수 있도록 함.
특히, 모든 절차를 별도의 개발 없이도 UI 상의 기능을 활용하여 진행할 수 있으며, 데이터의 외부 반출이 어려운 경우에도 On-premise 솔루션 도입을 통하여 활용 가능하여 주로 국방, 의료, 제조 등의 산업 분야 내 기업 및 기관에서 도입

데이터메이커와 함께 AI 도입 프로젝트를 완성하세요

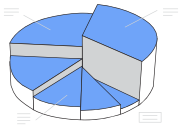
AI 도입 전담 파트너, 데이터메이커

오늘도 다수의 기업 및 기관에서 데이터메이커와 함께 AI 도입 프로젝트를 진행하고 있습니다.
데이터메이커는 데이터 구축부터 AI 모델 개발, 고도화까지 모든 과정을 고객 중심에서 맞춤형으로 협력하여 최고의 시너지를 창출해내는 AI 도입 전담 파트너입니다.

효과적인 컨설팅을 위해, 문의하기 시 아래 정보를 확인할 수 있어야 합니다.

AI 활동 목적 및 과제 정의

- 어떤 비지니스 문제를 해결하기 위한 AI가 필요한지
- 필요한 데이터 유형 및 관련 데이터 보유 현황



가용 예산

- 예산 규모에 따라 최적의 구축 범위 및 방식 제안 가능
- 제한된 예산에서도 효율적인 방식으로 구성해 드립니다.

희망 일정 및 구축 기간

- 전체 프로젝트 일정과 데이터 납기 희망 일정
- 단계별 산출물 및 검수 일정 조율을 위해 중요합니다.

꼭 모든 항목이 완벽히 정리되어 있지 않아도 괜찮습니다.

초기 아이디어 단계라도, 데이터메이커가 함께 구조화해 드릴 수 있습니다.
AI 도입을 위한 첫 시작부터 함께 설계하는 신뢰 파트너가 되겠습니다.

도입 문의하기 →

datamaker

AI&데이터 혁신 기업
데이터메이커

sales@datamaker.io
070-7139-9784
www.datamaker.io
대전 유성구 유성대로 871 4F, 5F

데이터메이커가 설계하는 맞춤형 데이터 구축 전략
AI & 데이터 혁신 기업, 데이터메이커