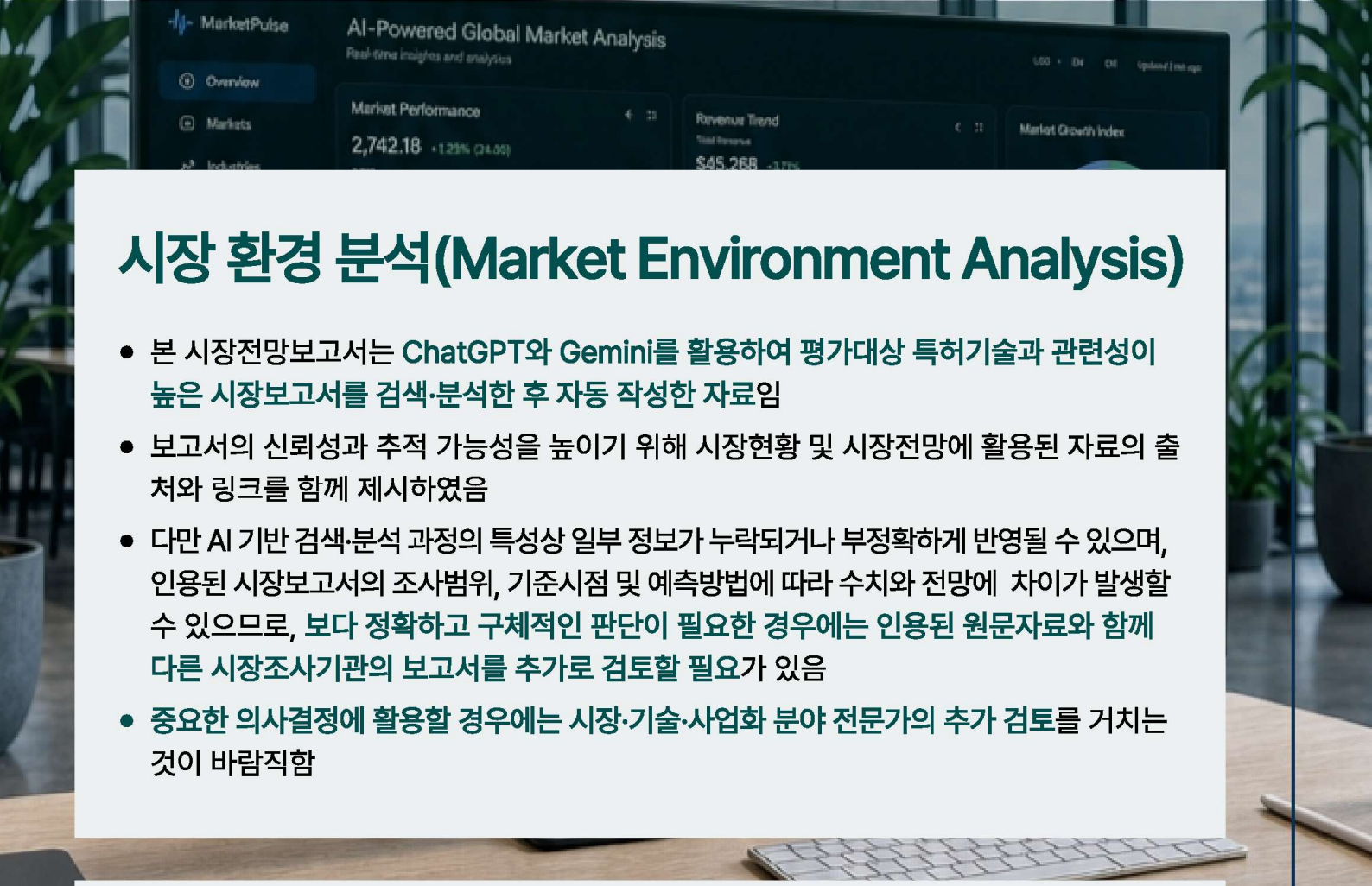


MARKET REPORT

AI-Powered Data Analytics and Comprehensive Market Evaluation

인공지능 데이터 학습·처리 기술



시장 환경 분석(Market Environment Analysis)

- 본 시장전망보고서는 **ChatGPT와 Gemini를 활용하여 평가대상 특허기술과 관련성이 높은 시장보고서를 검색·분석한 후 자동 작성한 자료**임
- 보고서의 신뢰성과 추적 가능성을 높이기 위해 시장현황 및 시장전망에 활용된 자료의 출처와 링크를 함께 제시하였음
- 다만 AI 기반 검색·분석 과정의 특성상 일부 정보가 누락되거나 부정확하게 반영될 수 있으며, 인용된 시장보고서의 조사범위, 기준시점 및 예측방법에 따라 수치와 전망에 차이가 발생할 수 있으므로, **보다 정확하고 구체적인 판단이 필요한 경우에는 인용된 원문자료와 함께 다른 시장조사기관의 보고서를 추가로 검토할 필요**가 있음
- **중요한 의사결정에 활용할 경우에는 시장·기술·사업화 분야 전문가의 추가 검토**를 거치는 것이 바람직함

시장보고서 전체 구성(Contents)

01. 시장보고서 작성기준

02. 시장정의 및 적용범위

03. 조사기관별 시장규모 및 CAGR

04. CAGR 시나리오(연평균 성장률)

05. 시장규모 시나리오

06. 국가별 분석(KR, US, JP, CN, EU)

07. 주요업체 및 경쟁환경

08. 성장요인 및 제약요인

09. SWOT 분석

10. PEST 분석

11. Porter's 5 Forces 모델

12. 시장성 시나리오 분석

13. 시장성 종합평가

14. 시장자료 출처 목록



1 시장보고서 작성기준

본 보고서는 시장성 평가를 목적으로 AI 학습용 데이터의 수집·정제·라벨링·주석·증강·큐레이션·검증 기술 가운데 텍스트 및 이미지·영상 기반 학습처리 시장을 조사하였다. 직접시장 수치가 확인되는 Grand View Research의 데이터 유형별 세부시장을 기준으로 하였고, 복수 조사기관의 전체 AI training dataset 시장 전망을 성장률 범위 검증에 활용하였다. 기관별 포함 서비스와 데이터 유형이 달라 수치 편차가 크므로 보수·기준·낙관의 3개 시나리오를 구성하였다. 공식 정량점수에는 기준 시나리오의 2030년 직접시장 규모와 CAGR만 반영하며, 국가별 정책·데이터 규제·경쟁강도·품질 및 저작권 리스크는 정성평가에만 활용하였다.

※ 주의 : 본 보고서는 시장자료의 편차를 인정하고, 직접시장 자료를 우선하되 공식 점수는 정량성이 높은 시장규모와 CAGR만으로 산정하는 구조임

2 시장 정의 및 적용 범위

본 시장은 인공지능 모델 학습을 위해 텍스트와 이미지·영상 데이터를 수집하고, 정제·비식별화·분류·라벨링·주석·세그멘테이션·객체탐지·캡셔닝·OCR·데이터 증강·합성데이터 생성·품질검증·편향관리·버전관리 형태로 가공하여 제공하는 데이터셋 및 관련 처리 서비스 시장으로 정의하였다. 텍스트 영역은 자연어처리, 대규모언어모델, 검색·요약·번역·질의응답·문서이해용 데이터를 포함하고, 이미지 영역은 컴퓨터비전, 객체탐지, 분류, 의료영상, 자율주행, 제조검사 및 멀티모달 학습용 이미지·영상 데이터를 포함한다. 음성 전용, 표형·시계열·센서 데이터, 모델 학습용 GPU·클라우드 인프라, 범용 AI 소프트웨어 및 최종 응용서비스 매출은 제외하였다.

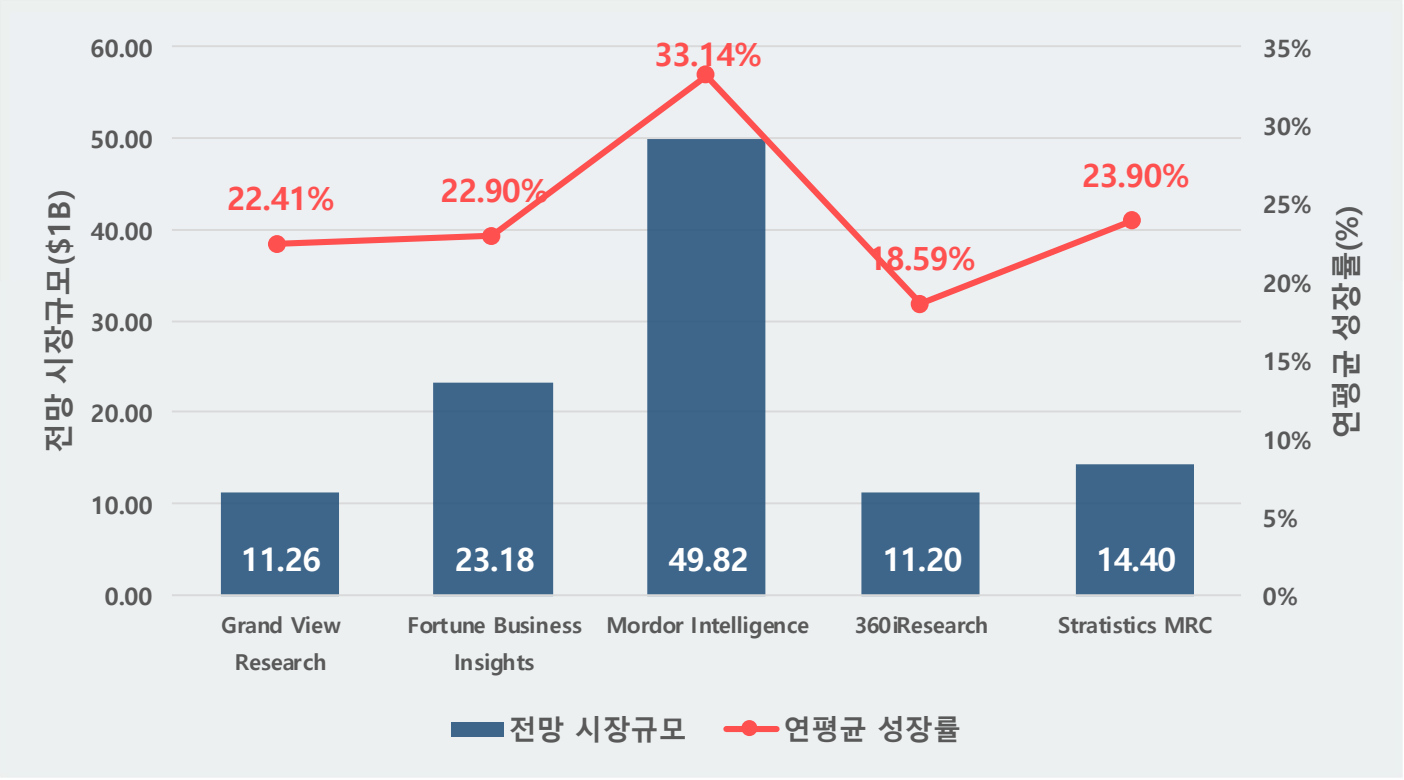
※ 주의 : 특허기술분야와 정확히 일치하는 시장보고서가 없는 경우는 기술의 상위 개념의 보고서를 분석하여 하위기술의 비중을 반영하여 분석함

3 조사기관별 시장규모 및 CAGR 비교

Grand View Research의 직접 세부시장 자료를 합산하면 2025년 텍스트 시장은 8.931억 달러, 이미지·영상 시장은 13.400억 달러로 합계 22.331억 달러이며, 2033년 합계는 112.575억 달러이다. 직접시장 합산 CAGR은 약 22.41%이다. 전체 AI 학습데이터 시장에 대한 다른 기관의 전망은 360iResearch 18.59%, Fortune Business Insights 22.90%, Stratistics MRC 23.9%, Mordor Intelligence 33.14%로 분포한다. 시장규모 차이는 데이터 유형, 라벨링·서비스 포함범위, 합성데이터와 플랫폼 매출 포함 여부 및 기준연도 차이에서 발생한다.

※ 주의 : 본 시장보고서 작성에 사용된 전체 시장보고서 목록은 시장보고서 파트의 후반부 「14. 시장자료 출처 목록」을 참조

조사기관별 시장규모 및 CAGR SUMMARY



단위 : USD billion

조사기관	보고서명	기준 연도	기준 시장규모	전망 연도	전망 시장규모	CAGR
Grand View Research	Text and Image-Video AI Training Dataset Segment Databooks	2025	2.23	2033	**11.26**	**22.41%**
Fortune Business Insights	AI Training Dataset Market, 2026-2034	2025	3.59	2034	**23.18**	**22.90%**
Mordor Intelligence	AI Training Dataset Market, 2026-2031	2025	8.74	2031	**49.82**	**33.14%**
360iResearch	AI Training Dataset Market Global Forecast, 2026-2032	2025	3.39	2032	**11.20**	**18.59%**
Stratistics MRC	AI Training Dataset Market Forecasts to 2032	2025	3.20	2032	**14.40**	**23.90%**

4 CAGR 시나리오(연평균 성장률)

시나리오	CAGR 범위	대표 CAGR	근거자료	설정근거
보 수	18~20%	**18.59%**	360iResearch	자동화·내재화와 범용 데이터 단가하락 반영
기 준	22~24%	**22.41%**	Grand View Research 직접 세부시장	텍스트·이미지 직접시장 합산 성장률 적용
낙 관	30~34%	**33.14%**	Mordor Intelligence	생성형 AI·전문 후학습·규제대응 데이터 급성장 반영

조사기관별 성장률은 18.59~33.14%로 분포한다. 보수 시나리오는 360iResearch의 18.59%를 적용하여 기업의 데이터 내재화, 자동 라벨링과 가격경쟁으로 성장속도가 둔화되는 경우를 반영하였다. 기준 시나리오는 텍스트와 이미지·영상 직접 세부시장을 합산한 22.41%를 적용하여 LLM·컴퓨터비전·멀티모달 AI의 일반적 확산을 반영하였다. 낙관 시나리오는 Mordor Intelligence의 33.14%를 적용하여 생성형 AI, 전문 후학습 데이터, 합성데이터 및 규제대응형 고품질 데이터 수요가 빠르게 확대되는 경우를 반영하였다.

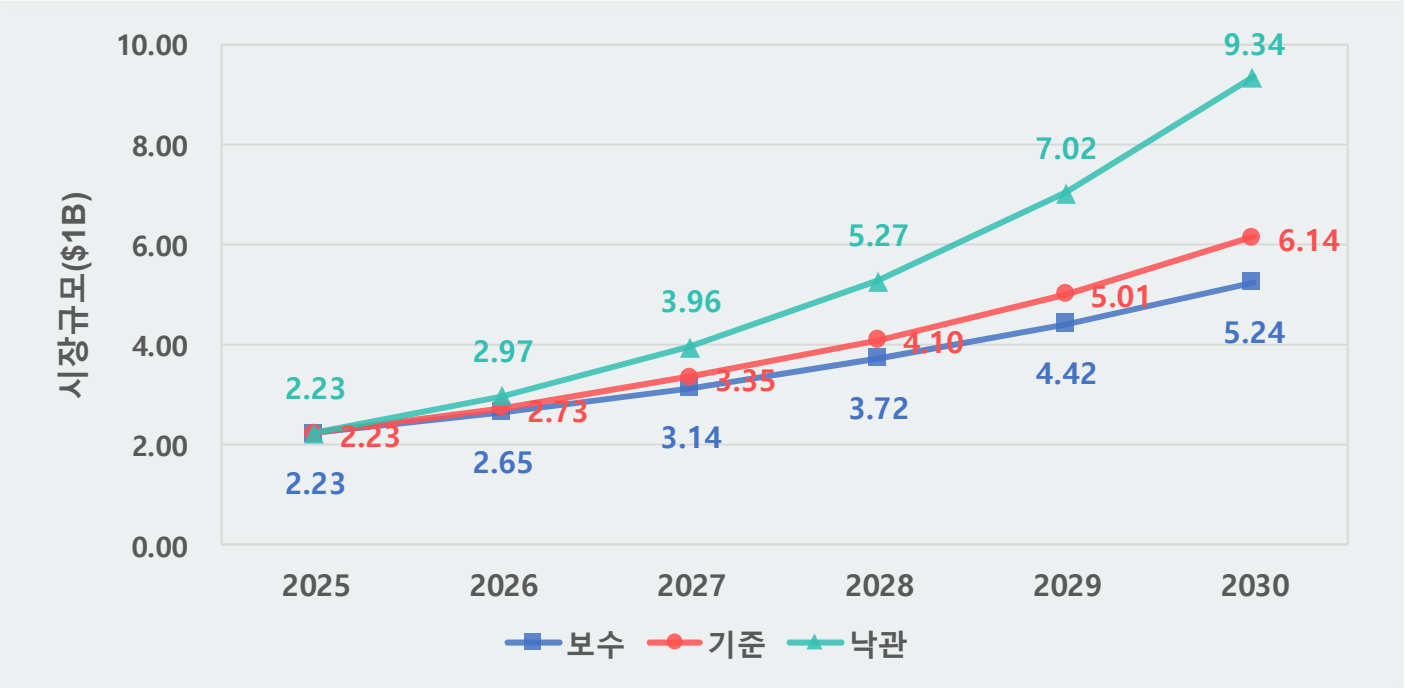
5 시장규모 시나리오

단위 : USD billion

시나리오	기준연도	기준 시장규모	전망연도	전망 시장규모	설정근거
보 수	2025	2.23	2030	**5.24**	직접시장 공통 기준값에 360iResearch 18.59% 적용
기 준	2025	2.23	2030	**6.14**	텍스트·이미지 직접시장 합산 CAGR 22.41% 적용
낙 관	2025	2.23	2030	**9.34**	직접시장 공통 기준값에 Mordor Intelligence 33.14% 적용

시나리오별 연도별 시장규모 추이

단위 : USD billion



6 국가별 분석(KR, US, JP, CN, EU)

국가별 분석은 한국, 미국, 일본, 중국, 유럽의 5개 권역으로 고정하였다. 국가별 공개수치는 대부분 텍스트·이미지 분리값이 아닌 전체 AI 학습데이터 시장이므로 공식 직접시장 점수에는 포함하지 않고 지역 수요의 참고지표로만 활용하였다. 공통적으로 생성형 AI와 컴퓨터비전 투자가 수요를 확대하지만, 저작권·개인정보·데이터 이전·모델 투명성 규제가 데이터 구축방식과 비용에 직접 영향을 준다.

국가 / 지역	종합해석
 한 국	한국의 전체 AI 학습데이터 시장은 Grand View Research 기준 2025년 0.607억 달러에서 2033년 3.734억 달러로 성장하며 CAGR은 25.4%로 제시된다. AI-Hub는 2026년 6월 LLM·LMM·합성·추론 데이터를 포함한 42개 학습용 데이터셋을 전면 개방하였다. 공공데이터 인프라와 한국어·산업특화 데이터가 기회이나, 공공사업 의존과 국내 단가경쟁은 제약이다.
 미 국	미국의 전체 AI 학습데이터 시장은 2025년 7.012억 달러에서 2033년 27.164억 달러로 성장하며 CAGR은 18.4%로 제시된다. 빅테크·AI 모델기업·자율주행·헬스케어 수요와 대규모 벤처투자가 강점이다. NIST의 생성형 AI 위험관리 프레임워크는 데이터 출처, 편향, 프라이버시와 검증가능성을 강조하여 고품질·추적가능 데이터 수요를 확대하는 동시에 준수비용을 높인다.
 일 본	일본의 전체 AI 학습데이터 시장은 2025년 1.824억 달러에서 2033년 10.250억 달러로 성장하며 CAGR은 24.1%이고, 이미지·영상이 2025년 52.85%로 가장 큰 유형이다. 제조검사·로봇·모빌리티·의료영상이 수요를 견인한다. 경제산업성의 AI 사업자 가이드라인은 전 생애주기 거버넌스와 안전성을 요구하여 품질관리형 데이터 사업에 기회를 제공한다.
 중 국	중국의 전체 AI 학습데이터 시장은 2025년 3.830억 달러에서 2033년 19.670억 달러로 성장하며 CAGR은 22.8%로 제시된다. 대규모 인터넷·전자상거래·스마트시티·제조영상 데이터와 정부 주도 AI 투자가 강점이다. 생성형 AI 서비스 잠정관리조치는 텍스트·이미지·음성·영상 생성서비스의 데이터 적법성과 보안을 요구하여 내수형 데이터 생태계를 강화하는 반면 국경간 데이터 활용을 제한할 수 있다.
 유 럽	유럽의 전체 AI 학습데이터 시장은 2025년 8.667억 달러에서 2033년 47.358억 달러로 성장하며 CAGR은 23.7%이고, 이미지·영상이 최대 유형이다. 자동차·산업·의료·공공부문의 고품질 데이터 수요가 강하다. EU AI Act의 범용 AI 의무는 저작권정책, 기술문서 및 학습콘텐츠 요약 공개를 요구하므로 데이터 출처관리와 라이선스 확인이 경쟁요소가 된다.

7 주요업체 및 경쟁환경

AI 학습데이터 시장은 Scale AI, Appen, Sama, iMerit, Labelbox, SuperAnnotate, AWS, Google/Kaggle, Microsoft, TELUS Digital, Snorkel AI 등 데이터 구축 전문기업과 클라우드·AI 플랫폼 기업이 경쟁하는 분산형 시장이다. 경쟁축은 단순 인력기반 라벨링에서 데이터 출처와 동의관리, 전문가 검수, 인간 피드백 기반 후학습, 합성데이터, 자동라벨링, 데이터 중심 AI 운영과 보안형 전달로 이동하고 있다. 대형 모델기업은 자체 데이터 파이프라인을 구축할 수 있어 공급자의 교섭력을 제한하지만, 의료·법률·제조·국방처럼 전문성과 규제가 높은 분야에서는 도메인 전문가와 추적가능한 데이터가 높은 진입장벽을 형성한다.

기업명	국가 / 지역	주요제품 및 기술	시장내 주요 역할	경쟁포인트	사업화 관점
Scale AI	미국	데이터 라벨링·생성형 AI 평가·후학습	전문 데이터 플랫폼 선도	대규모 운영·전문가·보안	정부·기업·후학습 데이터 확대
Appen	호주	텍스트·이미지 수집·주석·검색평가	글로벌 크라우드 데이터 기업	다국어 인력망·운영경험	품질·생성형 AI 데이터 전환
Sama	미국/케냐	미국/케냐	이미지·영상 전문기업	고정밀 주석·윤리적 고용	자율주행·산업영상 확대
iMerit	미국/인도	의료·자율주행·생성형 AI 데이터	전문가 데이터 기업	도메인 전문가·고난도 검수	규제산업 전문화
Labelbox	미국	라벨링·데이터 카탈로그·모델평가	기업용 데이터 중심 AI 플랫폼	자동화·협업 워크플로	소프트웨어·평가 데이터 확대
SuperAnnotate	미국	멀티모달 주석·생성형 AI 데이터	통합 데이터 플랫폼	프로젝트 관리·자동화	기업 후학습 데이터 확대
Amazon Web Services	미국	Ground Truth·클라우드·ML 파이프라인	클라우드 통합형 공급자	인프라·모델개발 결합	고객 락인과 자동화 확대
Google / Kaggle	미국	공개데이터·데이터 허브·Vertex AI	생태계·클라우드 플랫폼	개발자 커뮤니티·모델연계	공개·기업 데이터 생태계 확대

8 성장요인 및 제약요인

성장요인은 LLM·멀티모달 모델과 컴퓨터비전 확산, 모델성능을 좌우하는 고품질 데이터의 중요성, 합성데이터·자동주석·증강기술의 발전, 의료·법률·제조 등 산업특화 AI의 확대, 규제에 대응하는 출처·동의·편향관리 수요이다. 제약요인은 저작권·개인정보·데이터 이전 규제, 라벨 오류와 편향, 전문가 검수비용, 합성데이터의 품질저하와 오염 위험, 빅테크의 내재화와 자동화에 따른 범용 라벨링 단가하락이다. 따라서 단순 물량형 데이터보다 추적가능성·전문성·보안·평가 데이터를 제공하는 사업의 가치가 높아질 가능성이 크다.

성장요인



성장요인	내 용	시장성 평가관점 중요성
LLM·멀티모달 AI 확산	대규모언어모델과 비전-언어 모델은 방대한 텍스트·이미지 쌍, 지시응답, 선호도 및 안전성 평가 데이터를 필요로 한다. 모델 경쟁이 사전학습 규모에서 후학습 품질로 이동하면서 고품질 데이터 수요가 지속된다.	모든 생성형·멀티모달 모델의 반복적 학습·평가 수요를 창출
데이터 품질 중심의 성능개선	중복·오류·편향을 제거하고 도메인에 맞게 큐레이션한 데이터는 모델 정확도와 일반화에 직접 영향을 준다. 학습데이터 품질관리, 데이터 계보와 검증도구는 모델개발 비용과 재학습 위험을 줄인다.	모델성능과 안전성 개선을 통해 데이터 투자수익을 높임
합성데이터·자동주석 발전	생성모델, 액티브러닝, 자동세그멘테이션과 약지도는 데이터 구축속도와 비용효율을 높인다. 희귀사례·위험상황·개인정보 민감영역에서 실제데이터를 보완하여 학습범위를 확대한다.	단위비용을 낮추고 희귀·민감 데이터의 공급범위를 확대
산업특화·전문가 데이터 확대	의료영상, 법률문서, 제조검사, 자율주행과 로봇은 일반 크라우드 인력이 처리하기 어려운 전문 라벨과 검증을 요구한다. Vertical AI 확산은 단가와 진입장벽이 높은 전문가 데이터 시장을 확대한다.	고단가·고진입장벽 세부시장과 장기 고객관계를 형성
규제대응형 데이터 수요	EU AI Act, NIST 위험관리 및 각국 AI 가이드라인은 데이터 출처, 저작권, 편향, 개인정보와 문서화를 강조한다. 추적가능하고 라이선스가 명확한 데이터셋과 감사 가능한 처리기술의 수요가 증가한다.	출처·권리·감사가 가능한 데이터에 프리미엄을 부여

제약요인



제약요인	내용	시장성 평가관점 중요성
저작권·개인정보·데이터 적법성	웹 수집 텍스트와 이미지에는 저작권, 초상권, 개인정보 및 이용약관 문제가 존재한다. 적법한 수집·동의·비식별화와 삭제요청 대응이 미흡하면 소송·규제·모델재학습 비용이 발생할 수 있다.	데이터 사용중단·소송·재학습과 시장접근 제한 위험
라벨 오류·편향과 품질변동	크라우드 작업자의 이해도와 지침 해석 차이는 라벨 불일치와 편향을 유발한다. 오류가 대규모 학습에 누적되면 모델성능·공정성·안전성이 저하되므로 다단계 검수와 통계적 품질관리가 필요하다.	모델 오류·차별·안전성 문제로 고객 신뢰와 비용에 영향
전문가 데이터의 높은 비용	의료·법률·공학·안전성 평가 데이터는 자격을 갖춘 전문가와 보안환경을 요구한다. 희귀사례 확보, 다중 검수와 지속적 업데이트 비용이 높아 프로젝트 수익성과 확장속도를 제한할 수 있다.	고품질 프로젝트의 원가와 확장속도를 제한
합성데이터 오염과 모델붕괴 위험	합성데이터는 희귀상황과 개인정보 문제를 보완하지만 원 모델의 오류·편향을 증폭하거나 실제 분포를 왜곡할 수 있다. 합성데이터 비중과 품질검증이 부적절하면 데이터 다양성과 모델 신뢰성이 저하될 수 있다.	실제분포 왜곡과 반복학습에 따른 성능저하 가능성
자동화·내재화와 단가하락	대형 AI 기업은 자동라벨링과 자체 데이터팀을 확대하고, 범용 플랫폼은 가격경쟁을 심화한다. 단순 분류·바운딩박스 작업은 상품화될 가능성이 높아 공급자는 전문성·보안·출처관리와 평가 데이터로 차별화해야 한다.	범용 작업의 가격·마진을 낮추고 사업모델 전환을 요구

9 SWOT 분석



구 분		분석 내용
S	강점 Strengths	텍스트와 이미지·영상 데이터는 LLM, 컴퓨터비전 및 멀티모달 AI의 공통 핵심투입물이며 적용산업이 광범위하다. 데이터 수집부터 정제·주석·증강·평가·계보관리까지 가치사슬이 길고, 도메인 전문가·보안·규제준수와 결합하면 높은 전환비용과 반복매출을 형성할 수 있다.
W	약점 Weaknesses	범용 라벨링은 인력집약적이고 공급자가 많아 가격경쟁이 심하며, 프로젝트별 수요변동과 작업자 품질편차가 크다. 대형 모델기업의 내재화와 자동주석 확산으로 단순 작업의 마진이 낮아질 수 있고, 직접 세부시장 통계도 기관별 정의 차이가 크다.
O	기회 Opportunities	기업용 생성형 AI, 멀티모달 에이전트, 의료·제조·법률 등 Vertical AI, 인간 피드백 기반 후학습과 모델평가가 고품질 데이터 수요를 확대한다. 저작권과 개인정보가 명확한 라이선스 데이터, 국가·언어 특화 데이터, 합성데이터 검증과 데이터 계보 기술은 신규시장 기회이다.
T	위협 Threats	저작권·개인정보 분쟁, AI 규제 강화, 데이터 국외이전 제한과 보안사고는 데이터 사용가능성을 축소할 수 있다. 빅테크의 자체 데이터 구축, 오픈소스 데이터셋, 자동라벨링과 합성데이터가 기존 인력형 서비스의 단가와 점유율을 위협할 수 있다.

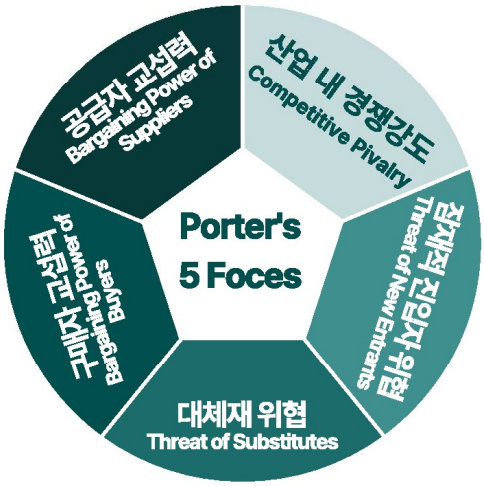
10 PEST 분석



구 분		분석 내용
P	정책·규제 Political / Policy	각국의 개인정보·저작권·AI 규제, 공공데이터 개방과 데이터 주권정책이 시장구조에 직접 영향을 준다. EU AI Act의 학습콘텐츠 요약과 저작권정책 의무, 미국 NIST의 위험관리, 한국 AI-Hub의 공공데이터 공급, 일본·중국의 AI 거버넌스는 규제준수형 데이터 처리와 현지화 수요를 확대한다.
E	경제·투자 Economic	AI 투자와 클라우드·반도체 지출 확대는 데이터 구축예산을 늘리지만, 경기둔화 시 데이터 라벨링은 비용절감 대상이 될 수 있다. 자동화는 단위비용을 낮추는 동시에 처리량을 확대하며, 범용 데이터의 가격하락과 전문가·독점 데이터의 프리미엄이 병존하는 이중구조를 형성한다.
S	사회·수요 Social	다국어·문화권별 표현, 유해콘텐츠, 공정성 및 개인정보에 대한 사회적 요구가 데이터 설계에 반영되어야 한다. 작업자의 노동조건과 심리적 안전, 데이터 제공자의 동의, 소수집단 대표성은 기업평판과 모델 신뢰성에 영향을 주며 책임있는 데이터 공급이 구매기준으로 강화되고 있다.
T	기술·특허 Technology	생성형 AI, 비전기반 파운데이션 모델, 액티브러닝, 자동주석, 합성데이터와 데이터 중심 AI 도구가 처리효율을 높인다. 반면 모델이 생성한 데이터를 다시 학습하는 과정의 오류증폭을 방지하려면 데이터 계보, 실데이터 검증, 품질지표와 지속적 모니터링 기술이 필요하다.

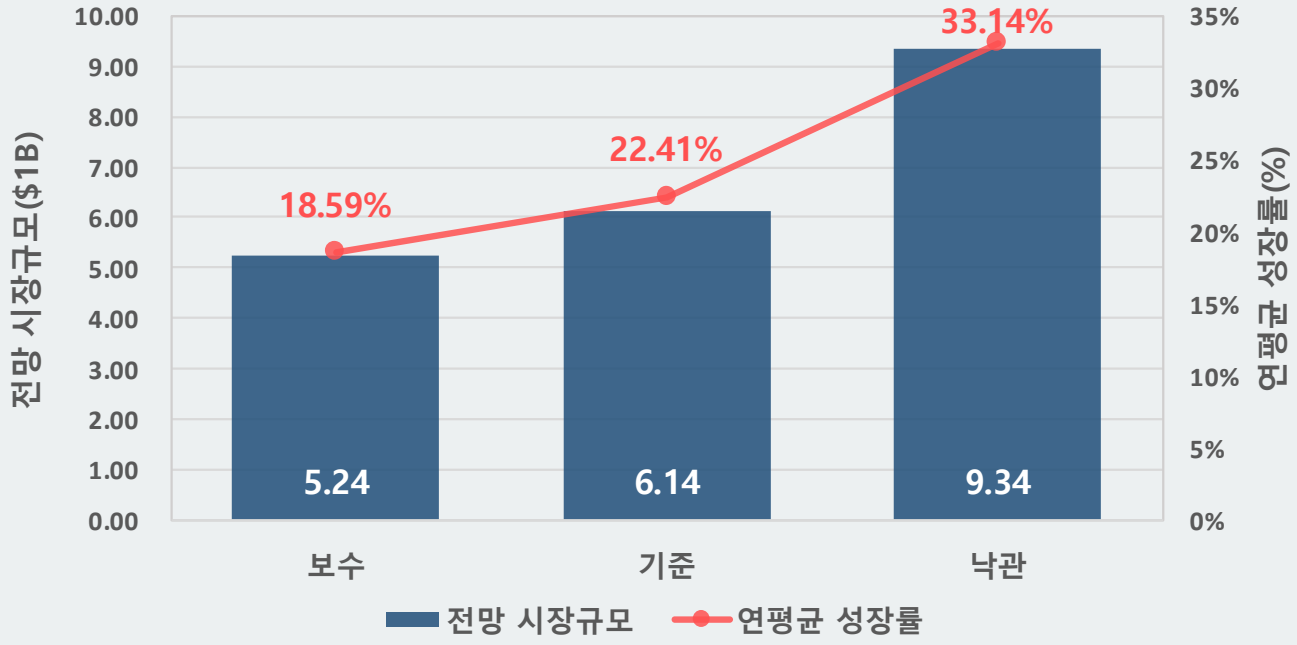
11 Porter's 5 Forces 분석

- 마이클 포터(Michael Porter)의 5 Forces는 특정 산업의 구조와 매력도(수익성)를 분석하는 경영 전략 프레임 워크임
- 5개의 경쟁(산업 내 경쟁강도, 잠재적 진입자 위협, 대체재 위협, 구매자 교섭력, 공급자 교섭력)을 점검하여 전략 수립 시 내부 역량과 결합하여 의사결정을 지원함



구 분	강 도	분석내용
산업 내 경쟁강도 Competitive Rivalry	**High**	강도: High. 데이터 전문기업, 크라우드 기반 공급자, 클라우드 플랫폼과 모델기업이 동시에 경쟁한다. 범용 작업은 가격과 속도 경쟁이 심하고, 고난도 분야는 전문가 품질·보안·출처관리·모델평가 역량이 구매결정을 좌우하여 세부시장별 경쟁구조가 빠르게 변한다.
잠재적 진입자 위협 Threat of New Entrants	**Medium**	강도: Medium. 기본 라벨링 도구와 인력조직만으로 소규모 진입은 가능하지만 대형고객 확보에는 보안인증, 글로벌 인력관리, 품질통계, 데이터 계보, 전문도메인 인력과 플랫폼 연동이 필요하다. 독점 데이터와 장기 고객관계는 실질적 진입장벽이다.
대체재 위협 Threat of Substitutes	**Medium~ High**	강도: Medium~High. 고객의 자체 데이터팀, 자동라벨링, 공개데이터셋, 약지도·자기지도 학습 및 합성데이터가 외부 데이터 서비스의 대체재이다. 다만 규제산업과 고위험 모델은 독립검수·전문가 데이터·실제 희귀사례가 필요해 완전 대체는 어렵다.
구매자 교섭력 Bargaining Power of Buyers	**High**	강도: High. 빅테크와 대형 AI 개발사는 대규모 물량, 자체 인력과 플랫폼 선택권을 바탕으로 단가와 품질조건을 요구할 수 있다. 반면 독점적 언어·산업 데이터, 전문검수와 보안환경을 보유한 공급자는 높은 전환비용으로 협상력을 확보할 수 있다.
공급자 교섭력 Bargaining Power of Suppliers	**Medium~ High**	강도: Medium~High. 핵심 공급요소는 데이터 권리자, 전문 작업자, 도메인 전문가와 안전한 클라우드 인프라이다. 희귀 의료영상·법률코퍼스·지역언어처럼 공급이 제한된 데이터의 권리자와 전문가는 높은 교섭력을 가지며 인건비와 라이선스 비용에 영향을 준다.

12 시장성 시나리오 분석

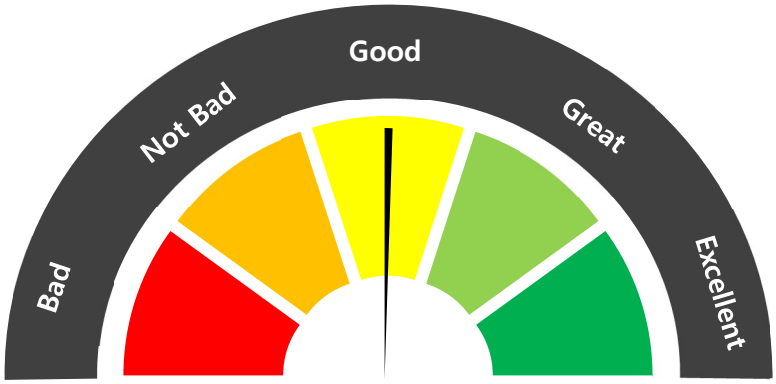


시나리오	정량 가정	해석 내용
보 수	CAGR 18.59% 2030년 약 52.38억 달러	자동화·고객 내재화·범용 라벨링 단가하락이 크게 작용
기 준	CAGR 22.41% 2030년 약 61.38억 달러	LLM·컴퓨터비전·멀티모달 및 산업특화 수요가 지속
낙 관	CAGR 33.14% 2030년 약 93.42억 달러	전문 후학습·평가·합성·규제대응 데이터가 급속 확대

AI 학습데이터 시장은 Scale AI, Appen, Sama, iMerit, Labelbox, SuperAnnotate, AWS, Google/Kaggle, Microsoft, TELUS Digital, Snorkel AI 등 데이터 구축 전문기업과 클라우드·AI 플랫폼 기업이 경쟁하는 분산형 시장이다. 경쟁축은 단순 인력기반 라벨링에서 데이터 출처와 동의관리, 전문가 검수, 인간 피드백 기반 후학습, 합성데이터, 자동라벨링, 데이터 중심 AI 운영과 보안형 전달로 이동하고 있다. 대형 모델기업은 자체 데이터 파이프라인을 구축할 수 있어 공급자의 교섭력을 제한하지만, 의료·법률·제조·국방처럼 전문성과 규제가 높은 분야에서는 도메인 전문가와 추적가능한 데이터가 높은 진입장벽을 형성한다.

13 시장성 종합평가

시장성 종합평가 등급



평가항목	평가결과	판단근거
시장규모	B	기준 시나리오 2030년 61.38억 달러로 50억~200억 달러 구간
성장성	A	기준 CAGR 22.41%로 20~40% 구간
기술확장성	High	텍스트·이미지 수집부터 증강·평가·계보·산업특화까지 확장 가능
경쟁강도	High	전문기업·크라우드·클라우드·모델기업이 전 가치사슬에서 경쟁
진입장벽	Medium~High	규제산업은 권리·보안·품질·전문인력·고객신뢰의 복합 장벽 존재
사업화가능성	High	필수 반복수요가 있으나 범용서비스보다 전문·권리확보형 모델이 유리
종합결론	**B**	데이터 품질·권리·보안 차별화가 핵심

텍스트 및 이미지·영상 기반 AI 학습처리 직접시장은 2025년 약 22.33억 달러로 산정되며, 기준 CAGR 22.41%를 적용하면 2030년 약 61.38억 달러에 도달한다. 시장규모 환산점수는 80점, CAGR 환산점수는 90점이며 가중합 정량점수는 84.0점으로 B등급이다. LLM·컴퓨터비전·멀티모달 및 산업특화 AI의 확산은 높은 성장성을 지지한다. 반면 저작권·개인정보, 라벨편향, 전문가 비용, 합성데이터 품질과 고객 내재화는 리스크이다. 따라서 특허가 단순 라벨링 자동화에 그치지 않고 데이터 출처·권리·품질·편향·보안·전문검수 또는 멀티모달 정합성을 개선하는 경우 사업가치가 높아질 가능성이 크다.

14 시장자료 출처 목록

전체 출처 목록에는 직접 텍스트·이미지 세부시장과 전체 AI 학습데이터 시장을 제공한 전문 조사기관 5개, 5개 국가·권역 시장자료, 한국 AI-Hub, 미국 NIST, EU 집행위원회 및 일본·중국의 공식 AI 거버넌스 자료를 수록하였다.

연 번	1	출처기관	Grand View Research
보고서명	AI Training Dataset Market and Text / Image-Video Segment Databooks		
활용항목	직접 텍스트·이미지 시장규모, CAGR, 기준 시나리오, 기술·기업동향		
URL	https://www.grandviewresearch.com/industry-analysis/ai-training-dataset-market https://www.grandviewresearch.com/horizon/statistics/ai-training-dataset-market/type/text/global https://www.grandviewresearch.com/horizon/statistics/ai-training-dataset-market/type/image-video/global		

연 번	2	출처기관	Fortune Business Insights
보고서명	AI Training Dataset Market Size, Share & Industry Analysis, 2026-2034		
활용항목	전체시장 규모·CAGR 비교, 기준 성장률 보조		
URL	https://www.fortunebusinessinsights.com/ai-training-dataset-market-109241		

연 번	3	출처기관	Mordor Intelligence
보고서명	AI Training Dataset Market Size & Share Analysis, 2026-2031		
활용항목	전체시장 고성장 전망, 낙관 시나리오, 경쟁환경		
URL	https://www.mordorintelligence.com/industry-reports/ai-training-dataset-market		

연 번	4	출처기관	360iResearch
보고서명	AI Training Dataset Market Global Forecast, 2026-2032		
활용항목	전체시장 저성장 전망, 보수 시나리오, 자동화 리스크		
URL	https://www.360iresearch.com/library/intelligence/ai-training-dataset		

연 번	5	출처기관	Stratistics Market Research Consulting
보고서명	AI Training Dataset Market Forecasts to 2032 - Global Analysis		
활용항목	전체시장 규모·CAGR 비교자료		
URL	https://www.marketresearch.com/Stratistics-Market-Research-Consulting-v4058/AI-Training-Dataset-Forecasts-Global-40922172/		

연 번	6	출처기관	Grand View Research
보고서명	AI Training Dataset Market Regional Databooks: South Korea, United States, Japan, China and Europe		
활용항목	한국·미국·일본·중국·유럽 지역시장 참고수치		
URL	https://www.grandviewresearch.com/horizon/outlook/ai-training-dataset-market/south-korea https://www.grandviewresearch.com/horizon/outlook/ai-training-dataset-market/united-states https://www.grandviewresearch.com/horizon/outlook/ai-training-dataset-market/japan https://www.grandviewresearch.com/horizon/outlook/ai-training-dataset-market/china https://www.grandviewresearch.com/horizon/outlook/ai-training-dataset-market/europe		

연 번	7	출처기관	과학기술정보통신부·한국지능정보사회진흥원 AI-Hub
보고서명	2025년 구축 인공지능 학습용 데이터 전면 개방 안내		
활용항목	한국 AI-Hub 학습용 데이터 공급과 정책환경		
URL	https://aihub.or.kr/aihubnews/notice/view.do?currMenu=131&nttSn=10549&topMenu=103		

연 번	8	출처기관	U.S. National Institute of Standards and Technology
보고서명	Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile		
활용항목	미국 AI 위험관리, 품질·편향·프라이버시 분석		
URL	https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence		

연 번	9	출처기관	European Commission
보고서명	General-Purpose AI Obligations under the AI Act / Training Content Summary Template		
활용항목	EU AI Act, 저작권·학습콘텐츠·규제 분석		
URL	https://digital-strategy.ec.europa.eu/en/factpages/general-purpose-ai-obligations-under-ai-act https://digital-strategy.ec.europa.eu/en/faqs/template-general-purpose-ai-model-providers-summarise-their-training-content		

연 번	10	출처기관	일본 경제산업성·중국 국가인터넷정보판공실
보고서명	AI Guidelines for Business Ver.1.2 / Interim Measures for Generative AI Services		
활용항목	일본·중국 AI 거버넌스와 국가별 분석		
URL	https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20260331_report.html https://cqca.miit.gov.cn/zwgk/zcwj/flfg/art/2026/art_27b7613aa2b8419384caa6c31fbdd7f0.html		