

스타트업 생태계 동향 리포트

데이터 활용을 둘러싼 제도적 제약 : AI 스타트업이 겪는 현실

스타트업 생태계 동향 리포트

데이터 활용을 둘러싼 제도적 제약 : AI 스타트업이 겪는 현실

Contents :

1. 들어가며
2. 산업 현장에서 직면한 제도적 제약
 - 2.1. 동의 기반 수집 원칙
 - 2.2. 동의 외 적법 근거
 - 2.3. AI 학습 데이터 특례
 - 2.4. 저작권자 동의
 - 2.5. 공정 이용 범위
3. 나가며

1. 들어가며

최근 전 세계적으로 인공지능(AI) 산업이 빠르게 성장하면서 국가 간 기술 경쟁은 그 어느 때보다 치열해지고 있다. AI 기술의 성능은 GPU와 같은 컴퓨팅 자원의 확보도 중요하지만, 궁극적으로는 학습에 투입되는 데이터의 규모·품질·다양성이 기업과 국가의 AI 경쟁력을 좌우하는 핵심 요소이다. 특히 생성형 AI와 대규모 언어모델(LLM)의 등장 이후, 데이터는 모델 성능을 결정짓는 필수 자원으로 자리 잡았으며, 국가 차원의 전략 자산으로서 그 중요성이 더욱 강조되고 있다. 그러나 국내 AI 기업들은 기술 개발 역량과는 별개로, 데이터 접근성과 활용 가능성에서 제도적 제약을 지속적으로 경험하고 있다. 이러한 제약은 기업의 모델 개발·검증·고도화 과정에 영향을 미치며, 결과적으로 AI 산업 전반의 성장 속도를 제한하는 요인으로 지적된다.

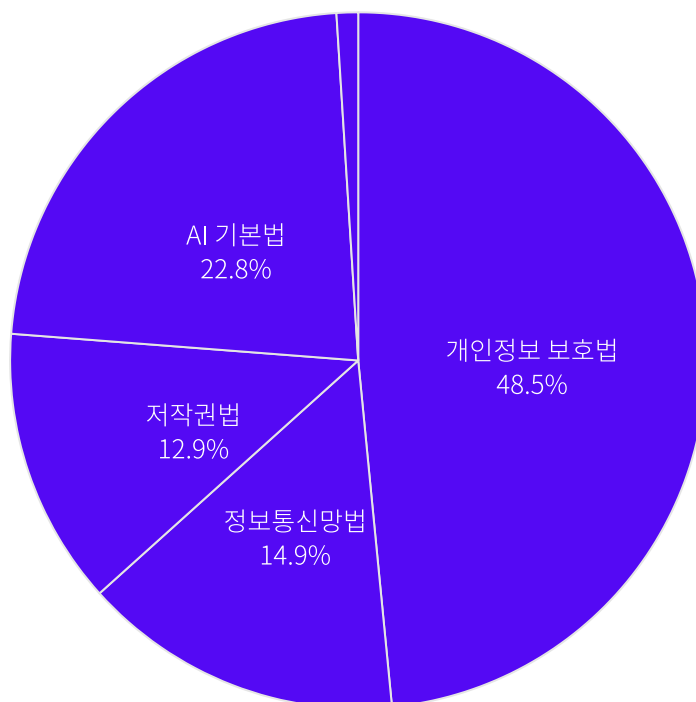
본 리포트는 국내 AI 스타트업을 대상으로 실시한 설문조사 결과를 바탕으로, 기업이 실제 데이터 확보·가공·학습 과정에서 마주하는 제도적 제약을 파악하고자 한다. 이를 통해 데이터 활용 실무와 현행 법·제도 체계 사이에 어떤 간극이 존재하는지를 살펴보고, 향후 정책 수립 과정에서 고려해야 할 주요 과제를 도출하는 데 목적이 있다. 궁극적으로 본 보고서는 한국 AI 산업이 보유한 잠재력이 제도적 한계로 제약받지 않도록, 정책적 개선 방향을 모색하는 데 기초 자료를 제공하고자 한다.

2. 산업 현장에서 직면한 제도적 제약

스타트업얼라이언스는 국내 AI 스타트업 101개사를 대상으로 AI 관련 정책 인식조사를 실시하였다. 조사는 2025년 9월 10일부터 19일까지 총 10일간, AI를 활용한 제품이나 서비스를 개발·제공하는 스타트업을 대상으로 이루어졌다. 설문은 종합 비즈니스 네트워킹 플랫폼 ‘리멤버(Remember)’를 활용한 온라인 방식으로 진행되었다.

먼저 AI 비즈니스 성장에 가장 제약이 되는 법·제도를 질문한 결과, AI 스타트업들은 ▲개인정보 보호법(48.5%), ▲AI 기본법(22.8%), ▲정보통신망법(14.9%), ▲저작권법(12.9%) 순으로 제약 요인으로 꼽았다.

(n=101, 단위: %)



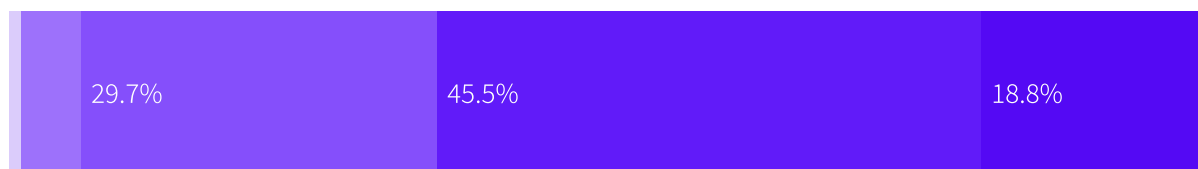
Q. 귀사의 AI 비즈니스 성장에 있어 가장 큰 법·제도적 걸림돌은 무엇이라고 생각하십니까?

이어서 개별 법령별로, AI 개발·운영 과정에서 데이터 수집·활용에 제약을 준다(그렇다+매우 그렇다)고 응답한 비율은 '개인정보 보호법'이 71.3%로 가장 높았고, 이어 정보통신망법 64.4%, 저작권법 62.4% 였다.



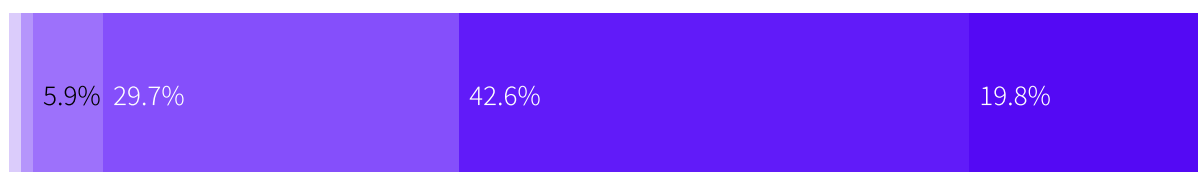
● 잘 모르겠다 ● 전혀 그렇지 않다 ● 그렇지 않다 ● 보통이다 ● 그렇다 ● 매우 그렇다

Q. 귀사는 AI 개발·운영과정에서 「개인정보 보호법」 이 데이터 수집·활용에 제약을 준다고 생각하십니까?



● 잘 모르겠다 ● 전혀 그렇지 않다 ● 그렇지 않다 ● 보통이다 ● 그렇다 ● 매우 그렇다

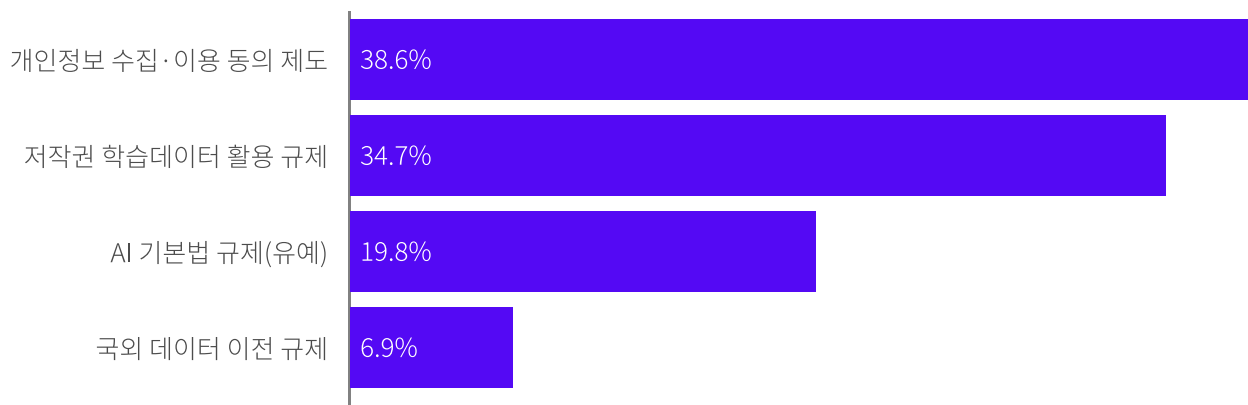
Q. 귀사는 AI 개발·운영과정에서 「정보통신망법」 이 데이터 수집·활용에 제약을 준다고 생각하십니까?



● 잘 모르겠다 ● 전혀 그렇지 않다 ● 그렇지 않다 ● 보통이다 ● 그렇다 ● 매우 그렇다

Q. 귀사는 AI 개발·운영과정에서 「저작권법」 이 데이터 수집·활용에 제약을 준다고 생각하십니까?

이어서 데이터 수집·활용과 관련해 정부가 가장 우선적으로 완화해야 하는 영역은 ▲개인정보 수집·이용 동의 제도(38.6%), ▲저작권 학습데이터 활용 규제(34.7%) 순으로 나타났다.

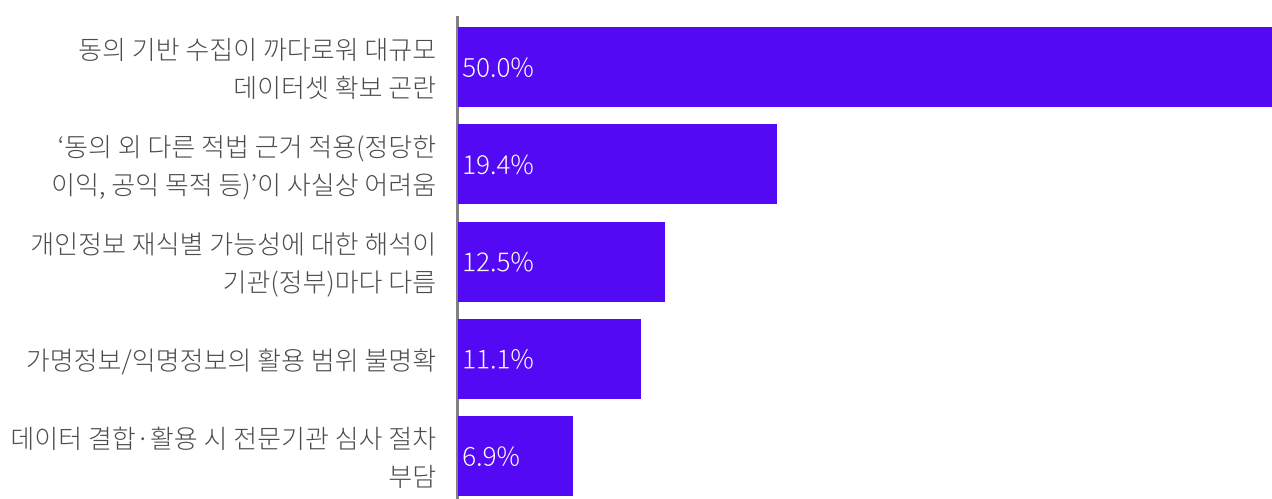


Q. 데이터 수집·활용과 관련해 정부가 가장 우선적으로 완화해야 한다고 생각하는 영역은 무엇입니까?

이처럼 AI 스타트업이 데이터 확보·가공·학습 과정에서 경험하는 제약은 개인정보보호법, 저작권법 등 여러 법령이 복합적으로 작용한 결과로 나타난다. 아래에서는 설문조사를 통해 도출된 주요 제약 요인을 정리했다.

2.1. 동의 기반 수집 원칙

AI 스타트업에게 개인정보 보호법 상 가장 큰 제약이 되는 부분은 ‘동의 기반 수집 원칙(50%)’으로 나타났다.



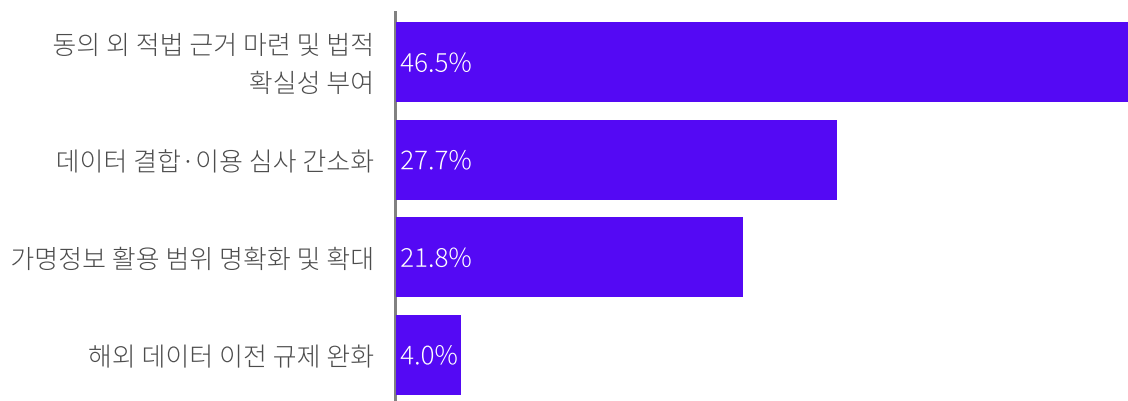
Q. 「개인정보 보호법」 상 어떤 부분이 가장 큰 제약이라고 생각하십니까?

현행 개인정보보호법은 원칙적으로 ‘정보주체의 동의’를 중심으로 개인정보의 수집·이용을 규율하고 있다. 그러나 AI 학습용 데이터셋은 대규모·비정형 데이터가 대부분을 차지해, 개별 정보주체의 동의를 건별로 확보하는 것이 사실상 불가능하다. 예를 들어 웹 크롤링 데이터나 서비스 이용 로그와 같이 정보주체를 식별하거나 동의를 재확보하기 어려운 형태의 데이터가 많아 동의 중심 구조는 실무적 한계를 드러낸다.

한편 글로벌 기업들은 비식별 처리나 정당한 근거 기반의 활용이 확산되고 있어 대규모 데이터 확보가 상대적으로 용이한 반면, 국내 기업은 동의 중심 구조로 인해 데이터 규모 경쟁에서 불리한 위치에 놓이고 있다. 특히 스타트업의 경우 인력·비용 측면에서 동의 확보에 필요한 행정적 부담을 감당하기 어려워, 데이터 확보에 대한 부담이 크게 작용하게 된다.

2.2. 동의 외 적법 근거

AI 스타트업은 「개인정보 보호법」에서 가장 완화가 필요한 부분으로 ‘동의 외 적법 근거 마련 및 법적 확실성 부여(46.5%)’를 꼽았다.



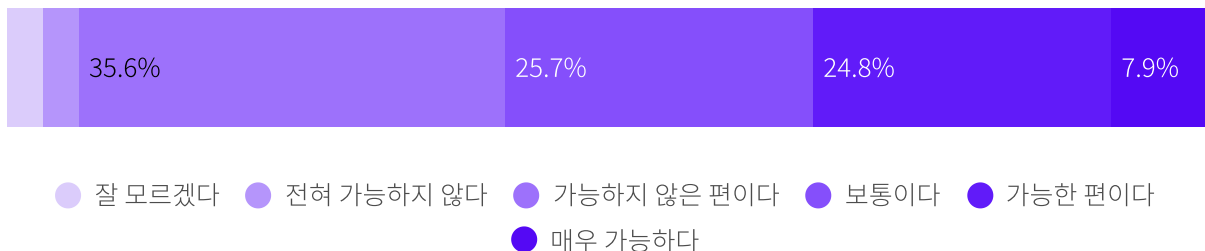
Q. 「개인정보 보호법」에서 가장 완화가 필요하다고 생각하는 영역은 무엇입니까?

개인정보 처리에는 동의 외에도 여러 법적 근거가 존재하지만, 실제 기업이 활용할 수 있는 범위는 매우 제한적이라는 지적이 많다. '정당한 이익(legitimate interest)' 근거가 부재하거나 제한적으로 해석되어, 기업들이 이를 실질적인 대안으로 사용하기 어렵다는 것이다. 또한 동의 외 처리 근거에 대한 규제기관의 해석 기준이 명확하지 않아, 기업 입장에서는 법적 리스크를 우려하여 보수적으로 대응할 수밖에 없다. 이로 인해 동의 외 근거를 활용하는 사례는 극히 제한적으로 나타나며, 이는 데이터 활용 위축으로 이어지게 된다.

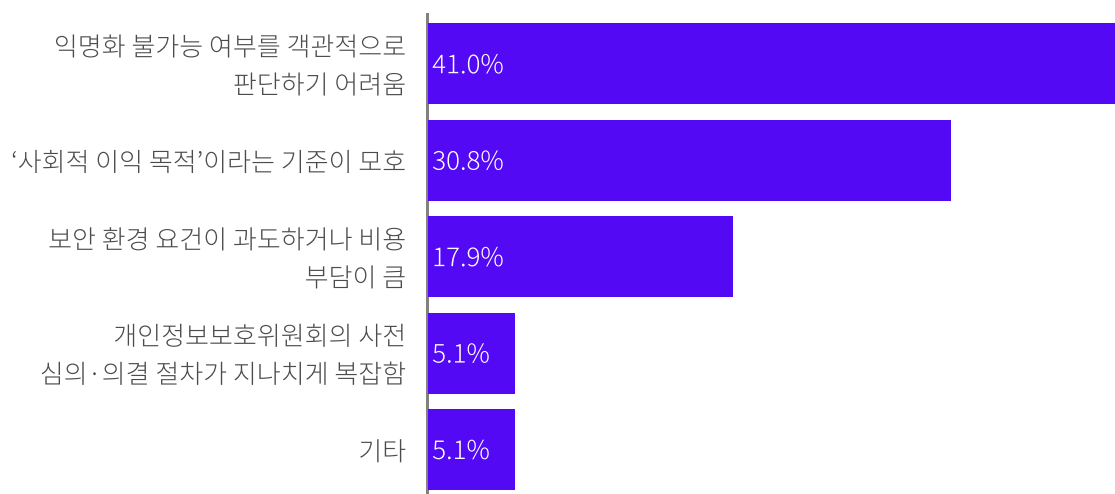
2.3. AI 학습 데이터 특례

최근 국회에서는 민병덕·고동진 의원안 등 AI 학습 목적의 데이터 활용을 위한 특례가 논의되고 있으나, 현장에서는 제도의 실효성에 대한 우려가 여전히 크다. 개인정보 규제 특례가 실무적으로 가능하지 않다(전혀 가능하지 않다+가능하지 않은 편이다)고 응답한 비율은 39.6%로, 가능하다(가능한 편이다+매우 가능하다)는 응답(32.7%)보다 높아, 특례임에도 현실적으로 활용하기 어려운 제도라는 인식이 확인되었다.

실무적으로 가능하지 않다고 응답한 이유로는 ▲익명화 가능 여부를 객관적으로 판단하기 어렵다는 지적이 41%로 가장 많았고, ▲‘사회적 이익 목적’이라는 기준이 모호(30.8%), ▲보안 환경 요건이 과도하거나 비용 부담이 큼(17.9%) 순으로 나타났다.



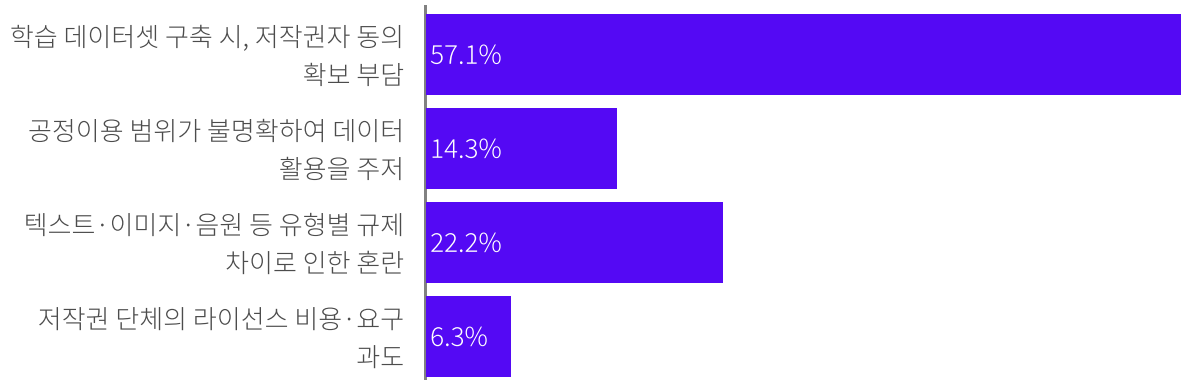
Q. 개인정보 규제 특례(예: 민병덕·고동진 의원 발의안)와 같이, 정부의 사전 승인 등 3가지 요건을 요구하는 절차가 실무적으로 가능하다고 보십니까?



Q. 개인정보 특례규제 절차가 실무적으로 가능하지 않다고 생각하는 이유는 무엇입니까?

2.4. 저작권자 동의 확보

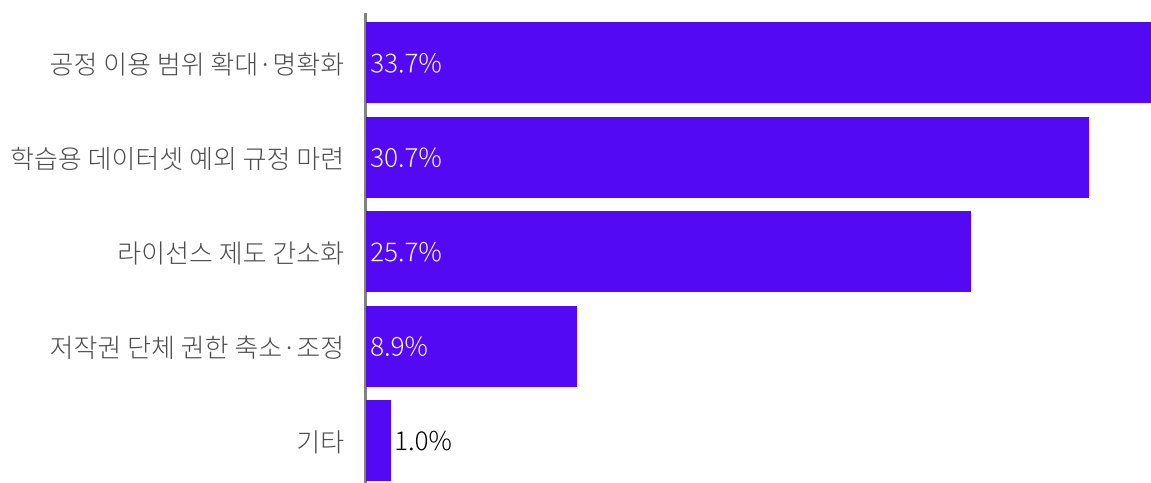
AI 스타트업에게 「저작권법」과 관련해 가장 큰 제약으로 지목된 항목은 ‘학습 데이터셋 구축 시 저작권자 동의 확보 부담’(57.1%)이었다. AI 학습 데이터는 이미지·텍스트·음성 등 다양한 형태로 구성되며, 이 중 상당수가 저작물에 해당한다. 그러나 대규모 데이터셋에서 개별 저작권자를 식별하고 동의를 확보하는 것은 현실적으로 매우 어렵다. 웹 기반 콘텐츠는 출처가 불명확하거나 원저작자를 확인할 수 없는 경우가 많아, 사실상 동의 확보가 불가능한 한계가 존재한다.



Q. 「저작권법」 상 어떤 부분이 제약이 된다고 생각하십니까?

2.5. 공정 이용 범위

AI 스타트업은 「저작권법」에서 가장 완화가 필요한 부분으로 ‘공정 이용 범위 확대 및 명확화’(33.7%)를 꼽았다. 이는 「저작권법」에 공정 이용 조항은 존재하지만, AI 학습에 적용할 수 있는지 여부가 명확하지 않기 때문이다. 공정 이용 판단 기준이 추상적이고, 관련 판례 및 가이드라인 부족하다는 점에서 기업은 저작권 분쟁 리스크를 안을 수 밖에 없다.



Q. 「저작권법」에서 가장 완화가 필요하다고 생각하는 영역은 무엇입니까?

3. 나가며

3.1. 정책적 시사점

앞서 살펴본 바와 같이 산업 현장에서 발생하는 데이터 활용 제약은 단순한 법 조항의 문제가 아니라, 현장의 데이터 처리 방식과 현행 규제 체계가 서로 맞지 않는 구조적 문제가 존재함을 보여준다. 이를 바탕으로 다음과 같은 정책적 시사점이 도출된다.

• 데이터 활용 중심 패러다임으로의 전환

기존 규제 체계는 개인정보 보호를 최우선 가치로 설정해 왔으며, 데이터 활용은 부차적 요소로 다뤄져 왔다. 그러나 AI 학습의 기술적 특성상 대규모·비정형 데이터의 활용이 필수적이며, 보호 일변도의 접근은 산업 성장을 제약할 수 있다. 따라서 데이터 활용 중심의 규율 체계로의 전환이 필요하다.

• 동의 중심 구조의 한계 보완과 동의 외 적법 근거 확충

동의 중심 구조는 AI 학습처럼 대규모 데이터를 처리해야 하는 산업 현장에서는 실무적으로 적용이 어렵다. 웹 데이터·로그 데이터 등은 정보주체 식별 및 재동의 확보가 사실상 불가능하기 때문이다. 현실적으로 동의 중심 구조를 완전히 대체하기는 어려운 만큼, ‘동의 외 적법 근거’의 도입과 명확화가 업계의 핵심 요구로 나타났다. 국제적으로는 GDPR의 ‘정당한 이익(legitimate interest)’이 기업의 합리적 데이터 활용 근거로 기능하고 있으며, 국내에서도 이와 같은 유연한 법적 근거 도입이 필요하다는 지적이 증가하고 있다. 이를 통해 기업은 적법하고 예측 가능한 방식으로 데이터 활용을 확대할 수 있다.

• 실효성 있는 AI 데이터 특례 설계

최근 발의된 AI 학습 데이터 특례는 산업계의 요구를 반영한 긍정적 시도이지만, 정부 사전 승인 등 절차적 요건이 과도하여 실제 활용 가능성이 낮다는 평가가 있다. 빠르게 개발·검증·배포가 이루어지는 AI 산업 특성상 사전규제 방식은 기업에 시간·비용 부담을 가중시키는 요소로 작용한다. 이에 사후책임·투명성 기반의 특례 구조가 현실적 대안이 될 수 있다.

• AI 학습용 데이터셋에 대한 별도 예외 규정 마련

유럽과 일본 등 주요 국가에서는 텍스트·데이터 마이닝(TDM) 예외 규정을 도입해 학습 목적의 데이터 이용을 명확히 허용하고 있다. 반면 국내 저작권법에는 학습 목적 이용에 대한 명시적 근거가 없어, 기업들이 대규모 데이터 구축을 주저하는 요인으로 작용한다. AI 학습의 기술적 특성과 국제적 규제 흐름을 고려할 때, 학습 목적에 한정된 예외 규정을 마련해 적법한 데이터 활용의 통로를 확보할 필요가 있다.

• 공정 이용 범위의 구체적 가이드라인 제시

현행 저작권법의 공정 이용 조항은 추상적으로 규정되어 있어 AI 학습에 적용 가능한지 여부가 명확하지 않다. 판례나 행정해석도 충분히 축적되지 않아, 기업들은 법적 불확실성을 피하기 위해 데이터 구축을 보수적으로 검토할 수 밖에 없다. 이에 따라 공정 이용 판단 기준을 명확히 하는 가이드라인 등 마련이 요구된다.

3.2. 결론

AI 경쟁력의 핵심은 더 이상 GPU와 같은 하드웨어 자원의 절대량이 아니라, 데이터를 얼마나 효율적으로 활용할 수 있는지에 달려 있다. 한국은 다양한 공공 및 민간 데이터 자산을 보유하고 있음에도 불구하고, 법·제도적 제약으로 인해 이러한 자산이 충분히 활용되지 못하고 있다. 이는 기술 역량의 문제가 아니라 데이터 활용을 가능하게 하는 제도적 기반이 미비한 데에서 비롯된 구조적 한계로 볼 수 있다.

특히 AI 기술이 빠른 속도로 발전하는 반면, 이를 규율하는 법적 근거와 규제체계는 그 속도를 따라가지 못하고 있어 활용 단계에서 지속적인 불확실성이 발생하고 있다. 이러한 불확실성은 기업의 데이터 확보·가공·학습 전략뿐 아니라, 투자 결정과 사업 확장, 기술 개발과 같은 핵심 경영 활동에도 영향을 미치고 있다.

따라서 AI 스타트업이 혁신 역량을 충분히 발휘하기 위해서는 데이터 활용을 둘러싼 제도적 불확실성을 완화하고, 책임 있는 활용을 가능하게 하는 명확한 기준과 절차가 마련될 필요가 있다. 현재 개인정보 보호법, 저작권법, AI 기본법 등 데이터 관련 법·제도 개편 논의가 활발히 진행되고 있는 만큼, 이러한 논의가 산업 현장의 요구와 제약을 균형 있게 반영하는 것이 중요하다. 향후 제도 설계 과정에서 본 보고서가 제시한 주요 제약 요인과 현장의 목소리가 충분히 고려된다면, 한국은 보유한 데이터 자산을 보다 효과적으로 활용하고, 글로벌 AI 경쟁에서도 실질적인 경쟁력을 확보할 수 있을 것이다.