

118 AI 책임성

AI Accountability

AI의 판단 과정과 결과에 대한 책임 주체를 명확히 하는 원칙

- AI 개발·운영 전 단계에서 투명성, 설명 가능성, 감독 체계를 확보하려는 개념
- 규제 공백이나 책임 불명확성이 발생하지 않도록 제도적 기준을 마련하는 방향

AI 책임성의 개념

AI 책임성은 AI가 어떤 기준과 근거로 판단을 내렸는지 설명할 수 있도록 하고, 오류나 피해가 발생했을 때 책임의 주체를 명확히 하는 원칙을 의미합니다. 이는 모델 개발부터 서비스 운영 전 단계에서 투명성과 설명 가능성을 확보해 기술의 부정적 영향을 최소화하기 위한 핵심 개념으로 논의되고 있습니다. 최근 미국에서는 주(州)의 AI 규제 권한을 제한하는 연방 예산안이 추진되며 감독 체계가 약화될 수 있다는 우려가 제기되었습니다. 이러한 논쟁은 'AI 시스템이 잘못된 결정을 내렸을 때, 이를 누가 책임지고 어떻게 설명할 것인가'라는 문제를 다시 부각시키고 있습니다. AI 책임성은 결국 AI의 판단 과정이 사회적 기준과 규범 안에서 검증 가능해야 한다는 요구를 반영하며, 오늘날 글로벌 AI 정책 논의에서 핵심적인 화두로 자리 잡고 있습니다.

AI 책임성의 중요성

AI 책임성은 공공과 민간이 AI를 설계하고 운영하는 과정에서 책임 구조를 분명히 하기 위한 기준으로 활용됩니다. 기업은 알고리즘이 어떤 기준과 과정으로 결과를 도출했는지 설명할 수 있어야 하고, 편향·오류로 피해가 발생하면 시정 조치를 마련해야 합니다. 특히 자동화된 의사결정이 대출, 채용, 의료, 복지 등 민감한 분야에 적용될 경우, 결과에 대한 이유를 명확히 제시할 수 있어야 신뢰를 얻을 수 있습니다. 정부는 공공·민간서비스 전반에서 AI가 과도한 위험을 초래하지 않도록 감독 기준과 절차를 세우고, 책임 주체가 모호해지지 않도록 제도적 장치를 갖춰야 합니다. 행정·의료·복지 등 공공 영역에서 AI가 의사결정 보조로 활용될 경우, 국민이 그 과정과 근거를 확인할 수 있는 투명성 확보도 필요한 요소입니다. 이러한 정책적 관점은 AI 활용이 확대될수록 책임성 기반이 더욱 중요해짐을 보여줍니다.

AI 책임성을 위한 과제

AI 책임성을 확보하기 위해서는 기술·제도·감독 체계 전반에서 해결해야 할 과제가 존재합니다. 첫째, 복잡한 모델이라도 이용자와 감독 기관이 이해할 수 있는 수준의 설명 가능성과 투명성을 확보해야 합니다. 둘째, 개발사·운영사·정부 간 책임 범위를 명확히 규정하는 법적 기준이 필요합니다. 현재 AI 생태계는 참여 주체가 다양해 책임 소재가 쉽게 분산될 수 있기 때문입니다. 셋째, 규제 권한이 축소되거나 여러 수준으로 나뉘는 경우 책임 공백이 생길 수 있어, 감독 체계의 일관성·실효성을 유지하는 제도 설계가 요구됩니다. 또한 AI가 특정 집단에 불리한 결과를 낳지 않도록 편향 관리와 영향 평가, 사후 조치 체계도 함께 마련해야 합니다. 이러한 과제들은 기술 발전과 사회적 신뢰를 균형 있게 유지하기 위한 기본 조건으로 논의되고 있습니다.