



AI 사고 피해 구제의 쟁점과 과제 EU 사례를 중심으로

황현아, 손민숙

본 보고서에 수록된 내용은 집필자 개인의 의견이며 우리원의 공식 의견이 아님을
밝혀 둔다.



목 차

• 요약	1
I. 서론	2
1. 연구배경	2
2. 선행연구	3
3. 연구내용	4
II. AI 사고 대응 체계	5
1. 위험원과 사고 대응 체계	5
2. AI 사고	6
3. AI 사고 대응 체계	19
4. 소결	23
III. AI 사고 책임법제와 보험제도의 쟁점	25
1. 책임법제와 보험제도	25
2. 책임법제 및 보험제도 측면에서 본 AI 사고의 특성	29
3. AI 사고 책임법제의 쟁점	33
4. AI 사고 책임보험의 쟁점	36
IV. 주요국 동향: EU AI 법제를 중심으로	41
1. 주요국 동향	41
2. EU AI 관련 입법 현황	51
3. EU AI 규제법: AI Act	53
4. EU AI 책임법: PLD 및 AILD	59
5. 시사점	80

V. 결론	85
• 참고문헌	88



〈표 II-1〉 AI 개념 요소	11
〈표 II-2〉 OECD AI 사고 및 관련 개념	14
〈표 II-3〉 AI 사고 개념	18
〈표 II-4〉 AI 사고 대응 체계	24
〈표 III-1〉 EG-NTF(2019)가 제시한 AI의 특성 및 책임법제 이슈	30
〈표 III-2〉 AI 사고의 특성 및 책임법제 이슈	36
〈표 IV-1〉 주요국 AI 법제 동향(요약)	50
〈표 IV-2〉 AI Act의 구성	53
〈표 IV-3〉 AI Act의 주요 개념	55
〈표 IV-4〉 EG-NTF(2019)의 책임법제 관련 쟁점 사항	66



그림 차례

〈그림 IV-1〉 EU AI 관련 입법 현황	52
〈그림 IV-2〉 AI 시스템 위험도에 따른 차등 규제	58

Abstract

Issues and Challenges in Remedies for AI Incident Harm : Focusing on the EU Case

As the use of AI expands, interest in AI incidents is increasing. AI systems are characterized by autonomy, adaptability, and general-purpose functionality. These features make it difficult to identify the responsible party and to determine the cause of an incident. Consequently, significant questions arise: whether to adopt a comprehensive regulatory framework or a sector-specific approach, whether to impose strict liability on particular stakeholders, whether to ease the burden of proof, whether a dedicated AI insurance scheme is necessary, and whether compulsory insurance should be introduced.

The European Union's proactive legislative efforts offer valuable lessons. Initially, the EU considered imposing strict liability and compulsory insurance, but it later opted to focus on alleviating the victim's burden of proof, and ultimately withdrew the draft Directive in 2025. This legislative trajectory provides three implications for Korea. First, victim protection and the fair allocation of liability are essential in the event of AI-related harm. Second, a cautious approach should be taken rather than pursuing radical institutional changes. Third, an individualized response framework is more appropriate than a comprehensive one.

In South Korea, full-scale discussions on AI accident liability are also expected to take place. Policymakers should refer to the EU's example and consider a realistic and phased approach that aligns with the pace of technological and industrial development, rather than radical legal revisions.

요약

AI 활용이 확대됨에 따라 AI 사고에 대한 우려도 커지고 있다. AI는 자율성, 적응성, 범용성 등 기존의 위험원(危險原)과 구별되는 특성을 가진다. 이로 인해 AI 사고는 기존의 다른 사고들과 달리 책임주체를 특정하기 어렵고, 사고 원인을 규명하기 곤란하며, 사고 영역이 광범위하고 피해 범익도 다양하다. 이러한 AI 사고의 특성은 AI 사고 피해 구제 측면에서 체계상 쟁점과 요건상 쟁점을 야기한다.

체계상 쟁점은 AI 사고 자체를 대상으로 한 포괄적 대응 방안과 개별 AI 활용 영역별 대응 방안 중 어느 것이 더 적합한지의 문제이다. 요건상 쟁점은 AI 사고에 대해 특정 이해관계자에게 엄격책임을 지울 필요가 있는지, 고의·과실 및 인과관계에 대한 입증 책임을 완화할 필요가 있는지의 문제이다. 보험과 관련해서는 기존 보험제도를 확장하여 AI 사고에 대응할 수 있는지, 아니면 AI 사고에 대한 별도의 보험이 필요한지, AI 사고 관련 배상책임보험을 의무보험으로 해야 하는지, 한다면 어느 범위까지 의무화해야 하는지 등이 문제 된다.

EU는 이 문제에 대해 선제적인 입법에 착수하였다. 논의 초기에는 고위험 AI 운영자에게 엄격책임을 부과하고 보험가입을 의무화하는 방안이 검토되었으나, 공개된 AI 책임지침에서는 AI 사고 피해자의 입증책임을 완화하는 방안이 반영되었다. 동 지침 초안은 2025년 철회되었으나 초안 작성 및 철회에 이르는 과정에서의 논의는 우리나라에 세 가지 시사점을 준다. 첫째, AI 사고 발생에 대비하여 피해자 보호 및 공평한 책임 배분 방안에 대한 체계적인 검토가 필요하다는 점이다. 둘째, 논의 초기에는 급진적 제도 변화 필요성이 제기되는 경향이 있으나 점차 신중론으로 의견이 수렴된다는 점이다. 셋째, 당면한 AI 사고에 대한 대응을 위해서는 포괄적 대응 체계보다 개별적 대응 체계가 적합하다는 점이다.

우리나라도 향후 AI 사고 피해 구제 논의가 본격화될 것으로 예상되는바, EU의 사례를 참고하여 급진적인 법 개정보다는 기술 및 산업 발전 속도를 고려한 현실적이고 단계적인 접근 방안을 우선적으로 고려할 필요가 있을 것이다.

1. 연구배경

AI 활용이 확대됨에 따라 AI 사고에 대한 관심도 높아지고 있다. 특히 교통, 의료, 금융 등 우리의 생명, 신체, 재산에 직접적인 영향을 미치는 영역에서 AI를 활용하는 경우 더욱 우려가 크다. 그러나 전 세계가 AI 기술과 산업 발전을 위해 경쟁하고 있는 상황에서 사고에 대한 우려로 AI 활용을 무조건 금지하거나 제한할 수는 없다. AI를 효과적으로 활용하되 사고가 발생하지 않도록 적절한 위험관리 방안이 필요하다. 또한 만에 하나 사고가 발생했을 때를 대비한 신속한 피해 구제 방안도 함께 마련되어야 한다.

이는 AI만의 문제는 아니다. 자동차, 항공기, 원자력 등 인류에게 편익과 위험을 함께 가져다주는 위험원(危險源)이 등장할 때마다 우리는 그 위험원을 활용하면서도 그로 인한 사고를 방지하고, 사고가 발생했을 때 피해자를 구제하는 방안을 마련해 왔다. 새로운 위험원에 대한 대응 방식은 위험의 크기와 성격에 따라 달라진다. 일상적인 수준의 낮은 위험의 경우 관련된 개별 경제주체들이 직접 위험을 관리하고 피해 구제 방안도 스스로 마련할 수 있다. 위험의 규모와 불확실성이 클수록 개별적인 대응이 어렵고 제도를 통한 공동체 차원의 대응이 필요해진다. 보험제도는 공동체를 통한 위험 대응의 가장 대표적인 예이다.

보험산업은 새로운 위험원이 등장할 때마다 위험을 인수하고 관리하며 사고 발생 시 피해자를 구제하는 역할을 담당해왔다. 특히 위험의 분산 및 피해 구제라는 측면에서 보험은 다른 위험관리 방안과 구별되는 독자적인 역할을 담당한다. AI라는 새로운 위험원에 대해서도 보험산업은 이러한 역할을 수행해야 할 것이다. 이를 위해서는 ‘위험원’으로서의 ‘AI’, ‘보험사고’로서의 ‘AI 사고’의 특성을 살펴볼 필요가 있다. AI 및 AI 사고에 대해서는 기술적, 산업적, 사회적 측면의 다양한 고찰이 이루어지고 있으나, 위험의 인수와 관리, 사고 피해의 보상이라는 보험의 관점에서 이를 검토해 볼 필요가 있는 것이다. 본고는 AI와 AI 사고에 대한 검토를 토대로 AI 사고 보험제도의 방향성을 모색해 보고, 특히 AI 사고에 대응하기 위한 책임보험의 역할을 살펴보고자 한다.

2. 선행연구

AI 사고 피해 구제 및 보험제도에 관한 선행연구는 크게 책임법제 관련 연구¹⁾와 보험제도 관련 연구로 나누어 볼 수 있다.

AI 사고 책임법제 관련 선행연구는 주로 AI 사고에 대해 누가, 어떤 요건에 따라 책임을 질 것인지를 다룬다. 주된 쟁점은 AI 사고에 대해 기존 책임법제로 대응 가능한지, 아니면 새로운 책임법제가 필요한지, 새로운 책임법제가 필요하다면 어떠한 형식과 내용이어야 하는지이다. AI 사고에 대한 새로운 책임법제가 필요하다는 주장은 다시 3가지로 구분된다. 첫째, AI 자체에 법인격을 부여해서 AI가 일으킨 사고에 대해 AI 스스로 책임을 부담하도록 하는 방안이다. 전자인(電子人) 제도 도입에 관한 논의가 그것이다. 둘째, AI 제조업자에게 책임을 지우는 방안이다. 제조물책임 법리에 따라 AI 제조업자가 AI 사고에 대해 무과실책임을 지도록 하는 방안이다. 셋째, AI 사고에 대해 자동차 사고에 적용되는 ‘운행자 책임’과 유사한 엄격책임을 적용하는 방안이다.²⁾

AI 사고에 대한 보험제도 관련 선행연구는 상대적으로 많지 않다. 해외의 주요 선행연구로는 EG-NTF(2019), Faure and Li(2022), Soyer and Tettenborn(2022), Lior(2022)가 있다. EU의 AI 관련 법체계 수립 과정에서 수행된 연구인 EG-NTF(2019)는 EU AI 법제의 기본 원칙인 ‘위험 기반 접근법’³⁾을 보험에도 적용하여, ‘고위험 AI’에 대해서는 AI 운영자(Operator)에게 엄격책임을 부과하면서 동시에 보험가입도 의무화해야 한다는 견해를 제시한다. 이때 고위험 AI 운영자에게 적용되는 책임법제 및 의무보험은 자동차 사고에 적용되는 운행자책임 및 배상책임보험과 유사한 형태이다. 반면 Faure and Li(2022)는 AI 사고에 대한 의무보험 도입 여부는 AI 자체의 위험성을 기준으로 정할 것이 아니라 배상의 무자의 무자력 가능성을 기준으로 정해야 한다고 본다.

한편 Soyer and Tettenborn(2022)은 AI 사고에 대한 책임법제와 보험제도는 기존의 제도와 조화를 이루어야 하며, 이러한 전제에서 볼 때 AI 사고로 생명·신체 손해가 발생한 경우에는 보다 강화된 책임법제 및 보험제도가 요구되나, 명예훼손 및 인격권 침해, 기타 손해에 대해서는 일반적인 과실책임 체계를 통해 해결하는 것이 적합하다는 의견을 제시한다.

1) AI 사고에 대한 책임법제는 형사책임, 민사책임, 행정책임 등으로 구분되나, 여기서는 보험제도와 직접 관련되는 민사책임 관련 연구에 한정하여 살펴봄

2) 책임법제에 관해서는 방대한 선행연구가 이루어진바, 구체적인 내용은 관련 부분에서 소개하기로 함

3) AI의 위험도에 따라 규제 정도를 달리해야 한다는 원칙으로, 이에 대해서는 EU AI Act의 주요 내용 부분에서 후술함

Lior(2022)는 AI 사고 대응을 위해 새로운 보험제도를 마련하는 것보다 기존의 보험제도를 활용하는 방안을 우선 고려해야 한다고 본다.

국내 선행연구로는 지광운(2024), 이현균(2024), 황현아(2024a), 황현아(2024b) 등이 있다. 지광운(2024)은 AI 사고에 대해 보험을 통한 대응 필요성을 제기하였고, 이현균(2024)은 모빌리티의 한 분야인 UAM 관련 보험제도의 방향성을 검토하였다. 황현아(2024a)는 EU AI 책임 지침 초안 및 관련 보고서들에 대한 분석을 토대로 AI 보험제도의 방향을 모색하였고, 황현아(2024b)는 AI 시대 보험의 역할과 과제에 대한 개괄적 분석을 제시하였다.

본 보고서는 AI 사고 대응을 위한 피해 구제 체계를 마련해야 한다는 인식에서 출발하여, 먼저 그 대상이 되는 'AI' 및 'AI 사고'의 의미를 명확히 하고, 이를 전제로 AI 사고 피해 구제에 적합한 보험제도를 구상해 보는 한편, AI 발전 초기 단계라는 점을 고려할 때 유의해야 할 점들을 일별해 보고자 한다.

3. 연구내용

본 보고서는 세 부분으로 구성된다. 첫째, AI 사고 대응 체계의 한 구성 요소로서의 책임법제와 보험제도의 의의와 역할을 살펴본다. 검토의 전제로 AI 사고의 의의와 특성을 구체화 해본다. AI 및 AI 사고는 기술적, 산업적, 사회적, 윤리적·철학적 측면에서 다양한 특성을 가질 수 있으나, 본고에서는 '보험사고'로서의 AI 사고의 특성을 고찰해 본다.

둘째, AI 사고 관련 책임법제와 보험제도의 쟁점들을 살펴본다. 이 부분은 향후 보다 많은 논의를 거쳐 구체화되고 그에 대한 해결책이 모색되어야 할 것인바, 본 보고서에서는 향후 함께 고민해 보아야 할 사항들을 정리해 본다.

셋째, AI 사고 대응 체계를 구성하는 주요국의 법제 동향을 살펴본다. 특히 AI 관련 법제를 선제적으로 정비하고 있는 EU의 법제 동향을 중점적으로 검토해 본다. 사전적 위험관리 수단에 해당하는 규제법, 사후적 피해 구제 수단에 해당하는 책임법 관련 EU의 논의를 살펴보고, 특히 AI 사고에 대한 엄격책임 및 의무보험 도입 관련 논의의 전개 과정 및 현황을 통해 시사점을 도출한다.

II

AI 사고 대응 체계

1. 위험원과 사고 대응 체계

위험원에 대해서는 그 위험원의 특성을 반영한 사고 대응 체계가 마련되어 있다. 가장 일상적인 위험원인 자동차의 경우 누구나 자동차 사고의 가해자나 피해자가 될 수 있고, 사고 예방을 위해서는 차량 자체의 안전성, 운전자의 주의, 체계적인 인프라가 확보되어야 하며, 사전 예방을 철저히 하더라도 사고를 100% 방지할 수는 없다는 특성을 가진다. 이러한 자동차 사고의 특성을 반영하여 자동차의 안전한 제작·운영 및 피해자의 신속한 구제를 위한 제도가 마련되어 있다. 자동차 사고의 경우 사고의 일상성 및 신속한 피해자 구제 필요성으로 인해 자동차보험의 역할이 매우 크다. 반면 사고 시 막대한 피해가 발생하는 원자력과 같은 위험원의 경우 사후 구제보다는 사고 예방에 초점을 두고 대응 체계가 마련될 수밖에 없다. 원자력 사고 피해자 구제를 위한 원자력 사고 배상책임보험도 운영되고 있으나 자동차보험과는 그 역할이나 운영 방식에 큰 차이가 있다. 이처럼 서로 다른 특성을 갖는 위험원들에 대해서는 서로 다른 대응 체계가 필요하다. 따라서 AI 사고에 대해서도 자동차 사고나 원자력 사고와 구별되는 AI 사고의 고유한 특성을 고려한 대응 체계가 마련될 필요가 있다.

본고의 주된 검토 대상인 AI 사고 책임법제와 보험제도는 AI 사고 대응 체계의 한 구성요소인바, 그 내용은 위험원으로서의 AI의 특성을 고려하여 마련되어야 한다. 따라서 AI 사고 책임법제 및 보험제도에 관한 논의에 앞서 먼저 ‘AI 사고(AI Accident 또는 AI Incident)’의 의미와 범위를 살펴보아야 한다. ‘AI’와 ‘사고’ 모두 일상적으로 사용되는 용어이고 그 의미도 사뭇 자명해 보여 새삼 그 의미나 범위에 대한 검토가 필요한지 의문이 있을 수 있다. 그러나 본고에서 다루고자 하는 ‘AI 사고’는 일상적인 개념이 아닌, 법적 책임 발생의 근거이자 보험의 보상 대상을 확정하는 개념이므로, 그 의의를 명확히 할 필요가 있다. 이러한 AI 사고 개념을 토대로 AI 사고 대응 체계를 전체적으로 조망해 보고, 대응 체계 내에서 책임법제와 보험제도가 가지는 의의와 역할을 파악해 본다.

2. AI 사고

가. AI의 의미

1) 일반적 정의

일반적으로 ‘인공지능(Artificial Intelligence, 人工知能)’은 ‘인간지능(Human intelligence, 人間知能)’ 및 ‘자연지능(Natural intelligence, 自然知能)’에 대비되는 개념으로 이해된다. 인간, 동물 등의 자연 생명체가 가지는 지능과 구별되는, 인간이 인공적으로 만들어낸 지능이라는 의미로 ‘기계지능(Machine intelligence)’과 같은 의미로도 사용된다.

AI에 대한 체계화된 정의는 1956년 존 맥카시에 의해 제안되었다. 그는 AI가 대중의 주목을 받기 시작했던 다트머스 회의에서 AI를 지능적 기계, 특히 지능적 컴퓨터 프로그램을 만드는 과학과 공학 분야라고 정의하였다. 이 시기에는 주로 공학과 기술의 관점에서 AI가 정의되었다. 이후 AI는 공학적, 기술적, 산업적 관점에서 재정의되었고, 이제는 철학적, 법률적 의미까지 가지게 되었으며, 이미 일상적 용어로 널리 쓰이고 있다. AI에 대해 하나의 합의된 정의를 내리기 어려우며, 우리 삶의 모든 영역에 적용되고 있는 디지털 알고리즘 내지 기계지능을 ‘현대 AI’라고 부르자는 제안도 있다.⁴⁾

AI의 분류 기준도 다양하다. 역량, 작동방식, 위험도 등에 따라 AI 유형을 세분화하는 시도가 이어지고 있다. 우선 AI의 역량에 따라 초인공지능(Artificial Super Intelligence; ASI), 강인공지능(Artificial General Intelligence; AGI), 약인공지능(Artificial Narrow Intelligence; ANI)으로 분류한다. 작동방식에 따라서는 규칙 기반(Rule-based AI), 기계학습 기반(Machine learning-based AI), 딥러닝 기반(Deep learning-based AI)으로 분류한다. 최근에는 생성형 AI(Generative AI)를 별도로 구별하기도 하며, 물리적 실체가 있는지에 따라 디지털 AI와 피지컬 AI로 나누기도 한다.

이처럼 AI의 의미를 구체화하기 위한 시도는 공학, 기술, 산업, 철학, 법률 등 다양한 관점에서 이루어지고 있고, AI의 발전에 따라 AI의 의미는 계속 변화하고 있다. 그러나 AI를 둘러싼 법률상, 계약상 권리 의무 관계를 확정하기 위해서는 보다 명확하고 구체적인 정의가 요구된다.

4) 대런 아세모글루·존슨(2023), p. 432

2) 법률상 정의

AI 관련 법령을 제정한 국가들은 해당 법령에서 AI를 정의하고 있다. 우리나라의 경우 2025년 1월 제정되어 2026년 1월 22일부터 시행 중인 ‘인공지능 발전과 신뢰 기반 조성 등에 관한 기본법’(이하 ‘인공지능 기본법’이라 함)에 ‘인공지능’과 ‘인공지능 시스템’에 대한 정의 규정이 있다. 동 법은 ‘인공지능이란 학습, 추론, 지각, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현한 것’이라고 정의하였다. 또한 ‘인공지능 시스템이란 다양한 수준의 자율성과 적응성을 가지고 주어진 목표를 위하여 실제 및 가상환경에 영향을 미치는 예측, 추천, 결정 등의 결과물을 추론하는 인공지능 기반 시스템’이라고 정의하였다. ‘고영향 인공지능’과 ‘생성형 인공지능’에 대해서는 각각 별도 정의 규정을 마련하였다. 주된 규제 대상인 ‘고영향 인공지능’은 사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능 시스템으로 에너지, 식수, 보건의료, 의료기기, 원자력, 범죄수사, 채용, 대출심사, 교통, 공공서비스, 학생 평가 등의 분야에서 활용되는 것을 의미한다. ‘생성형 인공지능’은 입력한 데이터의 구조와 특성을 모방하여 콘텐츠를 생성하는 인공지능을 의미한다.)

5) 인공지능 기본법 제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

1. “인공지능”이란 학습, 추론, 지각, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현한 것을 말한다.
2. “인공지능 시스템”이란 다양한 수준의 자율성과 적응성을 가지고 주어진 목표를 위하여 실제 및 가상환경에 영향을 미치는 예측, 추천, 결정 등의 결과물을 추론하는 인공지능 기반 시스템을 말한다.
3. “인공지능기술”이란 인공지능을 구현하기 위하여 필요한 하드웨어·소프트웨어 기술 또는 그 활용 기술을 말한다.
4. “고영향 인공지능”이란 사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템으로서 다음 각 목의 어느 하나의 영역에서 활용되는 것을 말한다.
 - 가. 「에너지법」 제2조제1호에 따른 에너지의 공급
 - 나. 「먹는물관리법」 제3조제1호에 따른 먹는물의 생산 공정
 - 다. 「보건의료기본법」 제3조제1호에 따른 보건의료의 제공 및 이용체계의 구축·운영
 - 라. 「의료기기법」 제2조제1항에 따른 의료기기 및 「디지털의료제품법」 제2조제2호에 따른 디지털의료기기의 개발 및 이용
 - 마. 「원자력시설 등의 방호 및 방사능 방재 대책법」 제2조제1항제1호에 따른 핵물질과 같은 항 제2호에 따른 원자력시설의 안전한 관리 및 운영
 - 바. 범죄 수사나 체포 업무를 위한 생체인식정보(얼굴·지문·홍채 및 손바닥 정맥 등 개인을 식별할 수 있는 신체적·생리적·행동적 특징에 관한 개인정보를 말한다)의 분석·활용
 - 사. 채용, 대출 심사 등 개인의 권리·의무 관계에 중대한 영향을 미치는 판단 또는 평가
 - 아. 「교통안전법」 제2조 제1호부터 제3호까지에 따른 교통수단, 교통시설, 교통체계의 주요한 작동 및 운영
 - 자. 공공서비스 제공에 필요한 자격 확인 및 결정 또는 비용징수 등 국민에게 영향을 미치는 국가, 지방자치단체, 「공공기관의 운영에 관한 법률」 제4조에 따른 공공기관 등(이하 “국가기관 등”이라 한다)의 의사결정
 - 차. 「교육기본법」 제9조 제1항에 따른 유아교육·초등교육 및 중등교육에서의 학생 평가
 - 카. 그 밖에 사람의 생명·신체의 안전 및 기본권 보호에 중대한 영향을 미치는 영역으로서 대통령령으로 정하는 영역
5. “생성형 인공지능”이란 입력한 데이터(「데이터 산업 진흥 및 이용촉진에 관한 기본법」 제2조 제1호에 따른 데

우리나라 인공지능 기본법상 AI의 정의는 EU의 입법례를 따른 것으로 보인다. 2024년 8월 발효된 EU AI Act는 ‘AI 시스템’을 ‘다양한 수준의 자율성(Autonomy)을 가지고 작동하도록 설계된 기계 기반 시스템으로, 배포 후 적응성(Adaptiveness)을 나타낼 수 있고, 명시적·묵시적 목표를 달성하기 위해 입력된 정보로부터 예측, 콘텐츠, 추천 또는 결정을 생성하는 방법을 추론하는 것’이라고 정의하고 있다.⁶⁾ EU AI Act도 우리나라 인공지능 기본법과 마찬가지로 주된 규제 대상으로서의 ‘고위험 AI 시스템(High-risk AI System)’의 해당 기준을 별도로 정하고 있는데, 기존에 안전 관리가 필요한 위험한 제품에 AI 시스템이 탑재되는 경우와 AI 시스템 활용 자체가 고위험을 창출하는 경우가 이에 해당한다.⁷⁾ 시스템적 위험을 야기할 수 있는 ‘범용 AI 모델’은 별도로 정의하고 있다.⁸⁾

미국의 ‘AI 이니셔티브법(National Artificial Intelligence Initiative Act of 2020)’은 AI 시스템을 ‘인간이 정한 일련의 목표에 따라 실제 또는 가상환경에 영향을 주는 예측, 추천 또는 결정을 내릴 수 있는 기계 기반 시스템(Machine-based system)’이라고 정의하고, “AI 시스템은 기계 및 인간에 의한 입력 자료를 사용하여 실제 및 가상환경을 인지하고, 이러한 인지를 자동화된 방식의 분석을 통해 모델로 추상화하며, 모델의 추론(Model inference)을 통해 정보 또는 행동에 대한 선택지를 만들어낸다”고 규정하고 있다.⁹⁾¹⁰⁾

이터를 말한다. 이하 같다)의 구조와 특성을 모방하여 글, 소리, 그림, 영상, 그 밖의 다양한 결과물을 생성하는 인공지능시스템을 말한다.

6) EU AI Act Article 3

(1) ‘AI system’ means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;

7) Article 6 Classification rules for high-risk AI systems

EU AI Act는 고위험 AI 시스템 해당 요건 및 범위를 부속서에서 구체적으로 정하고 있는바, 해당 내용은 제IV장에서 상세히 살펴보기로 함

8) EU AI Act Article 3

(63) ‘general-purpose AI model’ means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are released on the market;

9) 15 USC § 9401(3) artificial intelligence

(3) Artificial intelligence The term “artificial intelligence” means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to- (A) perceive real and virtual environments; (B) abstract such

일본의 ‘인공지능 관련 기술의 연구개발 및 활용 추진에 관한 법률(人工知能関連技術の研究開発及び活用の推進に関する法律)’은 ‘인공지능 기술’에 대한 정의 규정을 두고 있는바, 동 규정은 ‘인공지능 관련 기술이란 인공적인 방법으로 인간의 인지, 추론 및 판단에 관한 지적 능력을 대체하는 기능을 실현하는데 필요한 기술과 입력된 정보를 해당 기술을 이용하여 처리하고 그 결과를 출력하는 기능을 실현하기 위한 정보처리시스템에 관한 기술을 말한다’고 정의하고 있다.¹¹⁾¹²⁾

다른 국가들도 법령이나 가이드라인에서 AI 내지 AI 시스템의 정의 규정을 마련하였다. OECD가 49개국을 대상으로 조사한 결과 26개국에서 AI 개념에 대한 법적 정의 규정을 마련하였고, 14개국은 가이드라인 등 연성규범에서 AI를 정의하고 있었으며, 9개국은 AI에 대한 별도의 정의를 하지 않고 있다고 한다.¹³⁾

3) AI의 정의에 나타난 AI의 개념 요소

국가별, 법령별로 AI의 정의에 차이가 있으나 공통적인 개념 요소가 있다. AI는 인지, 학습, 판단, 추론 능력을 가진 기계 기반 시스템이라는 점, 일정 수준의 자율성과 적응성을 가진다는 점, 입력된 정보에 기초하여 예측, 콘텐츠, 추천, 결정 등의 출력물을 제공한다는 점, 그 출력물은 광범위한 영역에서 활용이 가능하다는 점이다. 이하에서는 항목별로 살펴보기로 한다.

첫째, AI는 인지, 학습, 판단, 추론 능력을 가진 기계 기반 시스템이다. 우리나라 인공지능 기본법과 일본의 인공지능법은 AI와 인간지능의 유사성에 주목하여 ‘인간이 가진 지적 능력과 유사한 능력을 가진 기계 기반 시스템’이라는 의인화 방식으로 서술하고 있는 반면, 미국의 AI 이니셔티브법은 AI가 수행하는 작업 자체에 주목하여 ‘인지, 추상화, 추론, 출력물 생성을 수행하는 기계 기반 시스템’이라고 서술하고 있다. 표현 방식에 차이가 있으나

perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action.

10) 번역은 박상철(2024), p. 32 및 국회도서관의 번역본(미국 국가 인공지능 계획법 2020 번역본)을 참고함

11) 第二条(定義) この法律において、「人工知能関連技術」とは、人工的な方法により人間の認知、推論及び判断に係る知的な能力を代替する機能を実現するために必要な技術並びに入力された情報を当該技術を利用して処理し、その結果を出力する機能を実現するための情報処理システムに関する技術をいう。

12) 번역은 국회도서관의 번역본을 참고함

13) OECD(2024), pp. 11~12

취지는 동일한 것으로 보인다.¹⁴⁾ 핵심은 '지능'을 활용하여 일정한 작업을 수행하는 '기계' 기반 시스템이라는 점이다.

둘째, AI는 인간의 개입으로부터 어느 정도 독립적으로 업무를 수행할 수 있는 특성인 '자율성(Autonomy)'을 갖는다. 이는 '지능'이 있기 때문에 가능한 것이기는 하지만 지능이 있다고 하여 곧바로 자율성이 인정되는 것은 아니라는 점에서 구별되는 개념 요소로 볼 수 있다. 우리나라 인공지능 기본법과 EU AI Act는 자율성을 AI의 개념 요소로 명시하고 있다. 미국은 '인지를 자동화된 방식의 분석을 통해 모델로 추상화하고, 모델 추론을 통해 선택지를 만들어 낸다'고 표현하고 있다. 표현에 차이는 있으나 양자 모두 AI에 의한 인지-판단-출력의 일련의 과정이 일정 정도 인간의 개입 없이 이루어진다고 보는 점에서 '자율성'이 AI의 특성이라고 전제하고 있다고 볼 수 있다.

셋째, AI는 일단 출시된 이후에도 스스로 학습을 통해 성능을 향상시키는 '적응성(Adaptiveness)'을 갖는다.¹⁵⁾ 이러한 적응성은 머신러닝 및 딥러닝 기반 AI가 갖는 특성이라는 점에서 규칙 기반 AI를 포괄하지 못한다. 다만 우리나라와 EU가 적응성을 AI의 개념 요소로 명시하고 있는 점, 향후 AI의 발전 방향은 적응성을 가진 AI 중심일 것으로 예상되는 점을 고려할 때, 적응성을 AI의 개념 요소로 보아도 무방할 것으로 생각된다.

넷째, AI는 입력된 정보에 기초하여 의도된 바에 따라 예측, 콘텐츠, 추천, 결정 등의 결과물(Out-put)을 만들어 낸다. 결과물의 형태는 다양할 수 있는데, 글, 소리, 그림, 영상 등의 콘텐츠를 만들어 내는 AI는 '생성형 AI'로 분류되고, 결과물이 로봇과 결합하여 현실 세계에 물리력의 형태로 나타나는 AI는 '피지컬 AI'로 분류된다.

다섯째, AI는 특정 분야에 제한되지 않고 인간의 삶 전반과 산업의 모든 영역에서 활용될 수 있는 '범용성'을 가진다. 범용성을 갖는 기술의 대표적인 예로 증기기관, 전기, 인터넷 등이 있다. 범용성은 AI 관련 법령상 정의 규정에 직접적으로 명시되어 있는 개념 요소는 아니다. 그러나 AI는 지능에 기반하여 추론을 통해 인간이 설정한 목표를 달성하는 것이고, 이때 그 목표에는 분야의 제한이 없으며, 실제로도 교통, 의료, 통신, 제조, 교육 등 우리가 생각할 수 있는 거의 모든 분야에서 AI가 활용되고 있다. 따라서 범용성은 AI의 특성

14) 이러한 미국의 정의가 OECD, EU AI Act 및 우리나라 인공지능 기본법상 정의보다 더 합리적이라고 보는 견해가 있음; 박상철(2024), pp. 28-31

15) EU AI Act 전문은 AI의 적응성에 관해 "The adaptiveness that an AI system could exhibit after deployment, refers to self-learning capabilities, allowing the system to change while in use."라고 정의하고 있음

중 하나라고 볼 수 있다.

이상에서 본 바와 같이 AI의 정의가 통일되어 있지는 않으나, 적어도 법적 규율 대상으로서의 AI는 인지, 학습, 판단, 추론 능력 등 지능을 가진 기계 기반 시스템으로, 일정 수준의 자율성과 적응성을 가지고, 입력된 정보에 기반하여 예측, 콘텐츠, 추천, 결정 등 일정한 출력물을 제공하는 것으로서, 범용성을 갖는 기술이라고 정의해 볼 수 있을 것이다.¹⁶⁾

〈표 II-1〉 AI 개념 요소

항목	내용
기계지능	인지, 학습, 판단, 추론 능력을 가진 기계 기반 시스템임
자율성	인간의 개입으로부터 어느 정도 독립적으로 업무를 수행할 수 있음
적응성	일단 출시된 이후에도 스스로 학습을 통해 성능을 향상시킬 수 있음
출력물 생성	입력물에 기반하여 예측, 콘텐츠, 추천, 결정 등 출력물 생성함
범용성	분야에 제한되지 않고 인간의 삶 전반 및 전 산업 영역에서 활용될 수 있음

주: EU AI Act 및 우리나라 인공지능 기본법 등의 AI 개념 정의에서 도출한 개념 요소임

나. AI 사고의 의미

1) ‘사고’의 의미

‘사고(事故)’의 사전적 정의는 ‘뜻밖에 일어난 불행한 일’ 또는 ‘사람에게 해를 입혔거나 말썽을 일으키는 나쁜 짓’이다. 이러한 사고 개념은 매우 광범위한 의미를 갖는다. 고의적으로 타인에게 피해를 끼친 경우부터 인간의 힘으로 어찌할 수 없는 불가항력적인 상황까지 모두 ‘사고’의 개념에 포함된다. 사고는 원하지 않는 나쁜 결과, 즉 ‘손해’ 내지 ‘해악’을 필수적 요소로 한다. 즉, 사고는 누군가에게 손해가 되는 결과를 야기한 사건이다.

법률상 ‘사고’는 법적 권리·의무 발생의 요건 사실에 해당하는 경우가 많다. ‘사고’ 자체가 법적 개념이라고 보기는 어려우나, 자동차 사고, 의료 사고, 항공기 사고, 원자력 사고, 해양 사고 등 ‘사고’를 다루는 법령들은 대부분 사고 관련 당사자들의 권리와 의무에 대해 정하고 있다. 사고는 민사법(民事法) 측면에서는 주로 ‘불법행위’로, 형사법(刑事法) 측면에서

16) 황현아(2026), p. 29

는 주로 '범죄'로 규율되며, 행정법(行政法) 측면에서는 '예방' 및 '규제'의 대상으로 규율된다. 책임법제의 종류에 따라 책임의 형태도 달라지는데, 민사법에서는 가해자의 피해자에 대한 손해배상책임, 형사법에서는 행위자에게 형벌의 형태로 부과되는 형사책임, 행정법에서는 사고가 발생한 위험원의 관리책임자에 대한 제재 형태로 부과되는 행정책임이 부과된다.

보험에서 '사고'는 특수한 의미를 갖는다. '보험사고'는 보험자의 보험금 지급 책임을 발생시키는 사건을 의미한다. 보험계약에서 보장하기로 한 위험이 실제 발생하는 것이 보험사고이다. 따라서 보험사고는 보험계약의 내용에 따라 결정된다. 자동차보험의 보험사고는 교통사고로 인한 생명, 신체, 재산상 손해 발생이다. 의료배상책임보험의 보험사고는 의료사고로 인한 환자의 생명, 신체상 손해 발생이다. 실손의료보험의 보험사고는 상해, 질병으로 인한 치료비의 지출이다. 보험사고는 법적 책임이 전제되는 경우도 있고 그렇지 않은 경우도 있다는 점에서 법률상 '사고' 개념과 반드시 일치하지는 않고, 불확정성과 우연성을 요구한다는 점이 가장 큰 특징이다.

요컨대, 일반적인 사고 개념은 해악·손해를 야기하는 사건을 의미하고 반드시 책임이나 보상과 결부되는 것은 아니다. 반면, 법률적 의미의 사고는 법적 책임 발생의 전제가 되는 요건 사실에 해당하고, 보험사고는 보험계약에서 정한 보험금 지급 책임을 발생시키는 사건을 의미한다. 이러한 사고 개념과 마찬가지로 'AI 사고'도 일반적인 개념과 책임법제 및 보험에서의 개념에 차이가 있을 수 있다.

2) 'AI 사고'에 대한 일반적 정의

가) AI와 관련된 손해의 발생(AIID)

일반적인 '사고' 개념에 따를 때 'AI 사고'는 AI와 관련되어 뜻밖에 일어난 일이나 불행한 일 또는 AI가 해를 끼치거나 말썽을 일으키는 일을 의미한다. 일반적인 의미의 AI 사고 정의의 대표적인 예는 'AI 사고 데이터베이스(AI Incident Database: AIID)¹⁷⁾의 정의이다. AIID는 AI 사고 사례를 수집·관리하는 플랫폼으로 2020년부터 현재까지 AI 관련 사고 사례를 수집하여 공개하고 있다. 이때 사례 수집 대상인 'AI 사고(AI Incident)'는 'AI 시스템

17) AIID 운영 주체인 'Partnership on AI'는 애플, 아마존, 딥마인드, 구글, 페이스북, IBM, 마이크로소프트 등의 연구자들이 설립한 조직으로, AI 활용의 베스트 프랙티스를 연구·구축하기 위한 협의체임

과 관련하여 사람, 재산, 또는 환경에 현실적 피해가 발생하였거나 피해 가능성이 매우 높은 상황이 발생한 경우'라고 정의하고 있다.¹⁸⁾

AIID의 'AI 사고' 개념은 AI와 관련해서 발생한 모든 직·간접적 손해(Harm) 및 손해 발생 가능성까지 포괄한다. 자율주행차가 보행자를 충돌하거나, 트레이딩 알고리즘 오류로 '플래시 크래시'¹⁹⁾를 유발하거나, 안면인식 시스템 오류로 무고한 사람이 체포되는 사례 등이 주요 사고 사례로 제시된다. 최근에는 음성 복제, 신원 조작에 기반한 사칭 사기 유형, 딥페이크 및 허위정보를 이용한 선거 캠페인, 착취적인 유해 콘텐츠 생성, 의료·법률 등 전문 분야에서의 AI 오남용으로 인한 지식 검증 체계의 붕괴, AI의 허위 결과물에 의한 지식 자체의 왜곡 현상 등의 사고 사례도 나타나고 있다.²⁰⁾

이와 같이 손해 발생 원인을 제한하지 않고 손해의 범위도 생명, 신체, 재산 등에 국한하지 않고 선거제도, 민주주의, 지식 및 사상, 시장의 건전성에 미치는 부정적 영향까지 포함하는 의미로 넓게 설정하면 그만큼 AI 사고의 범위는 넓어지게 된다. 광범위한 AI 사고 개념은 AI의 다양한 부정적 영향을 종합적으로 파악하는 데는 적합할 수 있으나, 책임 귀속이나 보험 보상의 대상으로 삼기에는 지나치게 광범위하고 불명확하다.

나) AI의 개발·사용·오작동에 의한 법익 침해(OECD)

OECD는 2024년 AI 사고 모니터링 시스템(AI Incident Monitor: AIM)을 도입하며 모니터링 대상인 'AI 사고(AI Incident)'의 개념을 정의하였다. AIM에서 정의한 AI 사고란 'AI 시스템의 개발, 사용, 오작동으로 인해 직간접적으로 손해(Harm)가 발생한 사건, 상황 또는 일련의 사건을 의미하며, 이때 손해는 (i) 사람의 생명, 신체, 건강을 침해하는 경우, (ii) 중요 기반시설의 관리 및 운영에 장애를 초래한 경우, (iii) 인권을 침해하거나 기본권, 노동권, 지적재산권 보호를 목적으로 하는 법령을 위반한 경우, (iv) 재산권, 공동체 및 환경에 피해를 끼친 경우를 의미함'이라고 정의된다.²¹⁾ OECD는 AI 사고의 관련 개념으로 AI 위

18) 'AI incident: an alleged harm or near harm event to people, property, or the environment where an AI system is implicated'(AIID Editor's Guide: Definition); AI Incident Database, Editors' guide (<https://incidentdatabase.ai/editors-guide/>) (2026. 2. 4. 최종 접속)

19) 금융시장에서 주식, 채권, 선물 등의 자산 가격이 아주 짧은 시간 안에 급격하게 폭락했다가 다시 급등하는 현상으로, 자동화된 알고리즘 거래 시스템의 이상 반응이나 오류에 의해 발생할 수 있음

20) AI Incident Database, AIID Blog(<https://incidentdatabase.ai/blog/incident-report-2025-april-may/>) (2026. 2. 4. 최종 접속)

험, AI 재난 등을 제시하며, 손해의 현실화 여부, 피해 법익의 종류 및 규모 등에 따라 AI로 인한 위험을 구분하고 있다.²²⁾

〈표 II-2〉 OECD AI 사고 및 관련 개념

구분 기준	손해의 현실화 여부	피해의 규모
AI 해악 (AI Harm)	실질적 해악 (Actual Harm)	AI 사고(AI Incident)
		중대한 AI 사고(Serious AI Incident)
	잠재적 해악 (Potential Harm)	AI 재난(AI Disasters)
		AI 위험(AI Hazards)
		중대한 AI 위험(Serious AI Hazards)

주: OECD의 AI 사고 관련 개념의 의미는 아래와 같음

- 1) AI 위험(AI Hazard)이란 시스템의 개발, 사용 또는 오작동으로 인해 AI 사고를 유발할 수 있는 사고, 상황 또는 일련의 사건을 의미함
- 2) AI 재난(AI Disaster)이란 공동체나 사회의 기능을 방해하고 자체 자원을 사용하여 대처할 수 있는 능력을 시험하거나 초과할 수 있는 심각한 AI 사고를 의미함. AI 재난의 영향은 즉각적이고 국지적일 수도 있으며, 광범위한 영역에서 장기간 지속될 수도 있음
- 3) 중대한 AI 사고·위험이란 사망이나 중상해를 야기하거나, 중요 기반 시설에 회복할 수 없는 위험을 야기하거나, 기본권, 재산권 등 중요 법익을 중대하게 침해하는 경우(EU AI Act 규정상 고위험 AI 관련 정의에 따라 정의함)

자료: OECD(2024)

OECD의 정의는 사고의 원인을 AI 시스템의 개발, 사용, 오작동으로 구체화하고, 손해의 유형을 4가지로 제한하였다는 점에서 AIID의 AI 사고 개념보다 범위가 좁고 구체적이나, 법적 책임 및 보험 보상의 대상으로 삼을 정도로 구체화되었다고는 보기 어렵다.

21) OECD(2024), p11; An AI incident is an event, circumstance or series of events where the development, use or malfunction of one or more AI systems directly or indirectly leads to any of the following harms:

- (a) injury or harm to the health of a person or groups of people;
- (b) disruption of the management and operation of critical infrastructure;
- (c) violations of human rights or a breach of obligations under the applicable law intended to protect fundamental, labour and intellectual property rights;
- (d) harm to property, communities or the environment.

22) OECD(2024), p. 10

3) 'AI 사고'에 대한 법률적 정의

가) AI의 개발·사용·오작동에 의한 중대한 법적 침해(AI Act)

EU AI Act는 고위험 AI 시스템 운영자에게 중대한 사고 발생 시 보고의무를 부과하고 있다.²³⁾ 이때 보고 대상인 '중대한 사고(Serious incident)'는 'AI 시스템의 오작동 등으로 (i) 사람이 사망하거나 사람의 건강에 대한 심각한 피해가 발생하거나, (ii) 핵심 기반시설의 관리나 운용에 중대하고 되돌릴 수 없는 장애가 발생하거나, (iii) 기본권 보호를 위한 EU 법령상의 의무를 위반하거나, (iv) 재산 또는 환경에 대한 심각한 피해가 발생하는 것을 의미한다'고 정하고 있다.²⁴⁾ 사고의 유형은 OECD AIM에서 제시한 것과 유사하고, 사고 범위는 피해의 종류와 크기에 따라 정하고 있다.

EU AI Act상 중대한 사고에 대한 보고의무는 고위험 AI 시스템 운영 과정에서 발생한 사고를 관리하기 위한 것으로 AI 사고에 대한 법적 책임이나 피해 구제와 곧바로 연결되지는 않으며, 보고 대상 사고에 대한 책임 문제는 별도로 다루어지게 된다.

나) AI 출력물(Out-put)에 의한 손해 또는 출력 실패로 인한 손해(AILD 초안)

EU의 AI 책임 지침(AI Liability Directive: AILD) 초안은 그 적용 대상을 'AI 시스템의 출력물(Out-put)로 인해 발생한 손해, 또는 출력이 기대되는 상황에서 출력물 생성 실패(Failure of such a system to produce an out-put)로 인해 손해가 발생한 경우 비계약상 과실 기반 민사 손해배상청구(Non-contractual fault-based civil law claim for compensation of the damage)'라고 정하고 있다.²⁵⁾ 이는 AI의 출력물 생성 또는 생성 실패로 손

23) Section 2 Sharing of information on serious incidents

Article 73 Reporting of serious incidents

1. Providers of high-risk AI systems placed on the Union market shall report any serious incident to the market surveillance authorities of the Member States where that incident occurred.

24) AI Act Article 3(49)

(49) 'serious incident' means an incident or malfunctioning of an AI system that directly or indirectly leads to any of the following:

(a) the death of a person, or serious harm to a person's health;
(b) a serious and irreversible disruption of the management or operation of critical infrastructure.
(c) the infringement of obligations under Union law intended to protect fundamental rights;
(d) serious harm to property or the environment;

해가 발생한 경우를 동 법의 적용 대상인 AI 사고로 보되, 당사자 간의 계약에 따라 책임을 물을 수 있는 경우에는 그 계약에서 정한 바에 따르고, 특별법에 의해 무과실책임을 물을 수 있는 경우²⁶⁾라면 그 법령에서 정한 바에 따르도록 하며, 그 밖의 경우로서 주로 민법상 과실책임이 적용되던 영역에 대해 AILD가 적용된다는 의미이다. 이상과 같은 AILD 적용 대상에 비추어 볼 때, AILD가 상정하고 있는 AI 사고 개념은 'AI 시스템의 출력물(Out-put)로 인해 발생한 손해, 또는 출력이 기대되는 상황에서 출력물 생성 실패로 인해 민사 손해배상청구의 대상이 되는 불법행위 상황이 발생한 경우'라고 볼 수 있다.²⁷⁾ 이때 'AI 시스템의 출력'은 기존의 인간의 행위 중 '작위'에, '출력이 기대되는 상황에서 출력물 생성 실패'는 '부작위'에 대응되는 개념인 것으로 보인다.

다) 고위험 인공지능에 의한 손해 발생(한국 인공지능 책임법안)

우리나라 국회에서 발의된 AI 관련 일부 법안들은 '인공지능 사업자의 책임'에 대해 정하고 있는바, 동 법안의 AI 사업자의 책임 발생 요건이 'AI 사고'에 해당한다고 볼 수 있다. 22대 국회에서 발의되었던 '인공지능책임법(안)'은 '고위험 인공지능 사업자는 이 법에 따른 의무 위반으로 고위험 인공지능 제품 또는 서비스 이용자에게 손해가 발생한 경우 그 손해를 배상할 책임이 있다'고 정하고 있다.²⁸⁾ 위 규정에서 상정하고 있는 'AI 사고'는

25) AILD Article 2 Definitions

For the purposes of this Directive, the following definitions shall apply:

(5) 'claim for damages' means a non-contractual fault-based civil law claim for compensation of the damage caused by an output of an AI system or the failure of such a system to produce an output where such an output should have been produced;

26) 자동차손해배상보장법(이하 '자배법'이라 함)상 운전자책임이나 제조물책임법상 제조물책임이 성립하는 경우가 이에 해당함

27) AILD 초안은 현재 철회된 상태임. AILD의 구체적인 내용은 후술함

28) 인공지능책임법(황희 의원 대표발의, 의안번호 2203235); 이는 인공지능기본법 제정으로 폐기되었음 제22조(고위험인공지능사업자 책임의 일반원칙)

- ① 고위험인공지능사업자는 이 법에 따른 의무 위반으로 고위험인공지능 제품 또는 서비스 이용자에게 손해가 발생한 경우 그 손해를 배상할 책임이 있다.
- ② 제1항에 따라 손해배상책임을 지게 되는 고위험인공지능사업자가 다음 각 호의 어느 하나에 해당하는 사실을 입증하는 경우에는 손해배상책임을 면제 또는 감경할 수 있다.
 1. 이용자의 손해가 해당 고위험인공지능으로 인한 것이 아니라는 사실
 2. 해당 고위험인공지능사업자가 고위험인공지능 기술을 사용하여 제품 또는 서비스를 공급하지 아니하였다는 사실
 3. 해당 고위험인공지능사업자에게 고의 또는 과실이 없었다는 사실
 4. 해당 고위험인공지능 제품 또는 서비스를 공급한 당시의 과학·기술 수준으로는 결함의 존재를 발견할 수 없었다는 사실
 5. 해당 고위험인공지능 제품 또는 서비스의 결함이 고위험인공지능사업자가 해당 고위험인공지능서비스를 공

‘고위험 인공지능 사업자가 인공지능책임법(안)에 따른 의무²⁹⁾를 위반하여 인공지능 제품 또는 서비스 이용자에게 손해를 끼치는 것’이라고 볼 수 있다. ‘인공지능책임법 위반’으로 사고 발생 원인을 제한한 것이 특징이다. 한편 21대 국회에서 발의되었던 ‘인공지능 책임 및 규제법(안)³⁰⁾도 ‘고위험 인공지능사업자가 인공지능 책임 및 규제법을 위반하여 고위험 인공지능 이용자에게 손해가 발생한 경우 그 손해를 배상할 책임이 있다’고 정하고 있는바, 인공지능 책임법(안)과 같은 취지라고 볼 수 있다.³¹⁾

이상의 규정들은 인공지능 책임법이 적용되는 사고 범위를 고위험 인공지능에 의해 발생한 사고에 한정하고, 책임주체를 고위험 인공지능 사업자로 제한하되, 손해의 규모나 종류는 별도로 제한하지 않았고, 책임 발생의 요건으로 인과관계나 고의·과실의 입증을 요구하지 않으며, 인공지능 사업자가 책임을 면하기 위해 인과관계나 고의·과실의 부존재를 입증하도록 하고 있다는 점이 특징이다.

4) 소결

AIID 및 OECD의 AIM과 같이 AI로 인한 각종 해악(Harm)의 발생 상황을 모니터링하는 것이 목적인 경우에는 가능한 다양한 사례를 수집할 수 있도록 AI 사고의 의미를 넓게 설정

급한 당시의 법령에서 정하는 기준을 준수함으로써 발생하였다는 사실

- ③ 정부는 고위험인공지능사업자에게 고위험인공지능 제품 또는 서비스로 인한 손해를 담보하기 위한 보험에 가입하도록 권고할 수 있다. 이 경우 정부는 보험가입을 위하여 재정적 지원을 할 수 있다.
 - ④ 정부는 고위험인공지능 제품 또는 서비스의 활용으로 인한 손해에 대응하기 위한 재정 또는 보험상품 개발 지원방안 등을 마련하여야 한다.
 - ⑤ 제3항에 따른 보험가입 권고·지원 및 제4항에 따른 재정 등의 지원에 필요한 사항은 대통령령으로 정한다.
- 29) 인공지능책임법(안)은 고위험 인공지능 개발사업자 및 이용사업자의 책무 규정(제19조, 제20조)을 두고 있었고, 동 규정은 현행 인공지능 기본법상 교영형 인공지능 사업자의 의무와 유사함
- 30) 인공지능 책임 및 규제법안(안철수 의원 대표발의, 의안번호 2123709 제11조); 21대 국회 임기 만료로 폐기됨
- 31) 인공지능 책임 및 규제법안 제11조(고위험인공지능사업자 책임의 일반원칙)
- ① 고위험인공지능개발사업자 또는 고위험인공지능이용사업자(이하 “고위험인공지능사업자”라 한다)는 이 법에 따른 의무 위반으로 고위험 인공지능 이용자에게 손해가 발생한 경우 그 손해를 배상할 책임이 있다.
 - ② 제1항에 따라 손해배상책임을 지는 고위험인공지능사업자가 다음 각 호의 어느 하나에 해당하는 사실을 입증하는 경우에는 손해배상책임을 면제 또는 감경할 수 있다.
 - 1. 고위험 인공지능 이용자의 손해가 해당 고위험 인공지능으로 인한 것이 아니라는 사실
 - 2. 해당 고위험인공지능사업자가 고위험 인공지능을 이용한 제품 또는 서비스를 제공하지 아니하였다는 사실
 - 3. 해당 고위험인공지능사업자에게 고의 또는 과실이 없었다는 사실
 - 4. 해당 고위험 인공지능을 이용한 제품 또는 서비스를 제공한 당시의 과학·기술 수준으로는 결함의 존재를 발견할 수 없었다는 사실
 - 5. 해당 고위험 인공지능을 이용한 제품 또는 서비스의 결함이 고위험인공지능사업자가 해당 고위험인공지능 서비스를 제공한 당시의 법령에서 정하는 기준을 준수함으로써 발생하였다는 사실

하는 것이 타당하다. EU AI Act의 ‘중대한 사고’ 개념 역시 고위험 AI 운영자가 정부 당국에 보고해야 하는 사고의 범위를 확정하기 위해 마련된 것인바 침해되는 법익을 중대한 경우로 제한하기는 하였으나 AI 사고 자체의 범위는 넓게 설정되어 있다.

반면 책임법제나 보험제도의 대상이 되는 ‘AI 사고’ 개념은 권리와 의무를 발생시키는 요건이 된다는 점에서 그 의미를 보다 구체적이고 명확하게 정할 필요가 있다. EU AILD 초안 및 우리나라 인공지능 책임법안에서 전제하고 있는 AI 사고 개념이 이러한 목적에 부합하는 개념이라고 볼 수 있다. 두 법안 모두 아직 제안 단계에 그치고 있고, 관련 논의도 충분히 성숙되었다고 보기 어려우나, AI 사고의 의미를 보다 구체화할 수 있는 단초를 마련하였다는 점에서 의의가 있다고 볼 수 있다.

〈표 II-3〉 AI 사고 개념

항목	목적	내용
AIID	사고 사례 수집	AI 시스템과 관련하여 사람, 재산, 또는 환경에 현실적 피해가 발생하였거나 피해 가능성이 매우 높은 상황이 발생한 경우를 말함
OECD AIM	사고 사례 수집	AI 시스템의 개발, 사용, 오작동으로 인해 직간접적으로 손해(Harm)가 발생한 사건, 상황 또는 일련의 사건을 의미하며, 이때 손해는 (i) 사람의 생명, 신체, 건강을 침해하는 경우, (ii) 중요 기반시설의 관리 및 운영에 장애를 초래한 경우, (iii) 인권을 침해하거나 기본권, 노동권, 지적재산권 보호를 목적으로 하는 법령을 위반한 경우, (iv) 재산권, 공동체 및 환경에 피해를 끼친 경우를 말함
AI Act	사업자 보고의무 부과	AI 시스템의 오작동 등으로 (i) 사람이 사망하거나 사람의 건강에 대한 심각한 피해가 발생하거나, (ii) 핵심 기반시설의 관리나 운용에 중대하고 되돌릴 수 없는 장애가 발생하거나, (iii) 기본권 보호를 위한 EU 법령상의 의무를 위반하거나, (iv) 재산 또는 환경에 대한 심각한 피해가 발생하는 경우를 말함
AILD	불법행위 특칙 적용 대상 확정	AI 시스템의 출력물(Output)로 인한 손해 발생, 또는 출력이 기대되는 상황에서 출력물 생성 실패(Failure of such a system to produce an out-put)로 인한 손해 발생 상황으로, 불법행위 손해배상청구의 대상이 되는 경우를 말함

이상과 같은 AI 사고 개념 논의의 현황을 고려할 때, 권리 의무 발생의 근거 사실이자 보험 보상 대상이 되는 ‘AI 사고’의 의미를 현재 단계에서 확정하기는 어려운 것으로 보인다. 다만 EU AILD 및 우리나라 인공지능 책임법안들이 다루고자 한 ‘AI 사고’가 주로 기존의 불법행위 영역에 속하던 사고로서 당사자 간의 계약에 의해 손해가 전보되기 어려운 경우인

점, 이러한 불특정 다수의 피해에 대해 보험에서 우선적인 보상 방안을 마련할 필요가 있다는 점을 고려할 때, AI 사고의 의미는 일응 'AI의 개발 및 활용으로 인하여 손해가 발생한 경우로서 기존 불법행위법의 적용 대상인 사고'라고 파악해 볼 수 있을 것이다. 이때 불법행위법 적용 대상인 사고란 고의·과실에 의한 위법행위 또는 이에 준하는 것으로 평가되는 위법행위로서 타인에게 손해를 끼친 경우를 의미한다.³²⁾

3. AI 사고 대응 체계

가. 사전 예방과 사후 구제

1) 일반적인 사고 대응 체계

일반적인 사고 대응 체계는 크게 사전 예방 체계와 사후 구제 체계로 구분해 볼 수 있다. 사전 예방 체계는 사고가 발생하지 않도록 사전에 위험을 관리하는 역할을 한다. 법제도 측면에서는 각종 규제법이 사고 예방 기능을 담당한다. 위험원의 제조, 운영, 이용에 관한 각종 규제들이 이에 해당한다. 사전 예방 조치에도 불구하고 사고는 발생할 수 있다. 이 경우 사고 피해자를 신속하게 구제하고 책임자에게 합당한 제재를 부과해야 한다. 법제도 측면에서는 책임법이 사후 구제 기능을 담당한다. 각종 민사책임 및 형사책임법제가 책임법에 해당한다. 보상을 통한 피해 회복을 담당하는 것은 민사책임법제이며, 보험제도도 피해 구제 체계의 한 축을 담당한다. 사후 구제를 위한 보험제도는 민사책임법제와 연계된 책임보험의 형태도 가능하고 책임법제와는 별개로 피해를 보상하는 손해보험 형태도 가능하다. 책임보험과 손해보험을 병용할 수도 있다.

사전 예방 체계와 사후 구제 체계는 그 목적과 기능에 차이가 있고 구체적인 적용 방식도 상이하나, 동일한 위험을 대상으로 한다는 점에서 상호 긴밀한 관련이 있다. 사고 예방 체계가 작동하지 않는 상황에서 사후 구제 체계를 마련할 수는 없고, 공정하게 책임을 배분할 수 있는 사후 구제 체계가 마련되어 있지 않을 경우 사전 예방 체계도 잘 작동하지 않는다. AI 사고에 대해서도 사전 예방 체계 및 사후 구제 체계가 함께 검토될 필요가 있다.

32) 이에 관한 구체적인 논의는 III장에서 서술함

2) AI 사고 사전 예방 체계와 사후 구제 체계

AI 사고에 대한 사전 예방 체계는 AI 규제법이 담당하고 있다. AI 규제법은 AI의 개발 및 활용 시 준수해야 할 사항을 정하는 일련의 법 제도를 의미한다고 볼 수 있다. 우리나라의 인공지능 기본법, EU의 AI Act가 이에 해당한다. 우리나라의 인공지능 기본법은 인공지능 기술 및 산업 발전의 기반을 마련하기 위한 목적이 주가 된다는 점에서 진흥법으로서의 성격을 가지고 있으나, AI의 개발과 활용에 대해 일정한 규제를 부과한다는 점에서 규제법의 성격도 함께 가지고 있다. 인공지능 기본법 중 AI 사고의 사전 예방 체계에 해당하는 사항은 인공지능 사업자에게 부과되는 투명성 확보 의무³³⁾와 안전성 확보 의무,³⁴⁾ 고영향 인공지능 관련 사업자의 책무³⁵⁾라고 볼 수 있다. 이러한 의무 이행 여부가 문제 되는 경우 사실조사 및 시정명령의 대상이 되고, 과태료가 부과될 수 있다.³⁶⁾

AI 사고에 대한 사후 구제 체계는 책임법의 영역이다. AI에 대한 별도의 책임법은 아직 입법이 되지 않은 상태이다. 우리나라의 경우 21대 국회 및 22대 국회에서 인공지능 책임법안이 발의되었으나 임기 만료 등으로 모두 폐기되었다. EU의 AILD 역시 상당 기간 논의가 진행되었으나 현재는 법안 자체가 철회된 상태이다. 현재 AI 사고에 대한 사후 구제는 일반불법행위법, 특수불법행위법, 계약법 등 기존의 책임법제에 맡겨져 있는 상황이다.

나. 포괄적 대응과 개별적 대응

1) AI의 범용성과 사고 대응 체계

과학기술의 발전으로 새롭게 등장한 위험원들은 대부분 위험원의 종류에 따라 사고 발생 영역, 피해자 및 피해 범익이 제한된다. 예컨대 자동차, 항공기, 의료기기가 야기할 수 있는 사고는 교통사고, 항공기사고, 의료사고로 그 영역이 제한된다. 원자력의 경우 광범위한 피해를 야기할 수 있으나, 사고로 인한 피해 범익은 생명, 신체, 재산, 환경 등에 한정된다.

반면 범용기술인 AI는 산업 및 일상생활 전반에서 활용될 수 있어, 사고가 발생하는 영역

33) 인공지능 기본법 제31조(인공지능 투명성 확보 의무)

34) 인공지능 기본법 제32조(인공지능 안전성 확보 의무)

35) 인공지능 기본법 제34조(고영향 인공지능과 관련한 사업자의 책무)

36) 인공지능 기본법 제40조(사실조사 등); 제43조(과태료)

이 광범위하고 피해자나 피해 법익에 제한이 없다. 모빌리티에 탑재된 AI가 오작동하면 교통사고를, 의료기기에 탑재된 AI가 오작동하면 의료사고를 야기하게 된다. 금융분야에서 활용되는 AI의 경우 활용 영역이 트레이딩인지 마케팅인지에 따라 거래오류를 야기할 수도 있고 불완전판매나 부당권유 같은 금융소비자 권익 침해를 야기할 수 있다. 생성형 AI가 만들어 내는 각종 콘텐츠는 청소년의 정신건강에 부정적인 영향을 미칠 수도 있고, 저작권이나 초상권을 침해할 수도 있으며, 성적 수치심을 야기할 수도 있고, 타인의 명예를 훼손시킬 수도 있으며, 선거의 공정성을 저해할 수도 있다. 전기, 에너지, 통신 같은 망산업(Network industries)으로 분류되는 기술도 범용기술이지만 사고 영역의 광범위성 및 피해 법익의 다양성 측면에서 볼 때 AI 사고가 훨씬 더 광범위한 영역에서 다양한 피해를 야기한다.

이처럼 광범위한 영역에서 발생하는 다양한 법익 침해 상황에 대해 'AI 사고'라는 단일한 범주로 묶어서 대응하는 것도 가능하고, AI 활용 영역별로 개별적으로 대응하는 것도 가능하다. 전자를 포괄적 대응, 후자를 개별적 대응이라고 할 수 있다. 이하에서는 AI 사고 위험에 대한 대응을 포괄적 대응과 개별적 대응으로 구분하여 각각의 장단점을 검토해 보기로 한다.

2) 포괄적 대응

포괄적 대응은 사전 예방 및 사후 구제의 대상을 'AI 사고'로 정하는 것이다. 'AI 윤리규범', 'AI 가이드라인', 'AI 규제법', 'AI 책임법', 'AI 보험제도' 등 활용 영역이나 맥락을 불문하고 AI 자체를 규율 대상으로 삼는 것은 모두 포괄적 대응에 해당한다.

포괄적 사전 예방 방안의 대표적인 사례가 EU AI Act와 우리나라의 인공지능 기본법이다.³⁷⁾ EU AI Act와 우리나라 인공지능 기본법은 AI에 대한 정의 규정을 마련하고, 규제 대상 AI에 해당하는 경우 단일한 주무관청이 동일한 규제들의 묶음을 집행하도록 하는 수평적 규제 체계를 마련하고 있다는 점에서 포괄적 사전 예방 체계에 해당한다.³⁸⁾ 이러한 포괄적 규제는 AI의 고유한 특성을 고려한 일관성 있는 규제가 가능하다는 장점이 있다. 반

37) 주요국 중 우리나라와 EU 외에 아직 AI에 대한 포괄적 규제법을 도입한 나라는 없는 것으로 보임. 캐나다의 경우 2022년 인공지능 및 데이터법(the Artificial Intelligence and Data Act; AIDA)이 발의되었으나, 2025년 1월 캐나다 의회의 회기 종료로 자동폐기되었음

38) 박상철(2024), p. 7

면 백화점식·저인망식 과잉규제를 야기할 수 있고 기존 개별 규제와 충돌할 수도 있다는 우려도 제기되고 있다.³⁹⁾

책임법 측면에서의 포괄적 대응 방안으로 AI 사고에 대해 AI 자체에 책임을 귀속시킬 수 있도록 하는 전자인(電子人) 제도 도입 방안, AI 사고 전체에 적용되는 일반적 책임법제 도입 방안 등이 논의되고 있다. 규제법과 달리 책임법 영역에서는 포괄적 대응에 해당하는 입법적 조치는 아직 이루어지지 않은 상태이고, ‘AI 사고 보험’과 같은 포괄적 보험 논의도 아직까지는 이루어지지 않고 있다.

3) 개별적 대응

개별적 대응은 AI가 활용되는 영역과 맥락을 고려하여 개별적으로 사고에 대한 사전 예방 및 사후 구제 방안을 마련하는 것이다. 이때 규율 대상은 개별 영역 및 맥락에서 규정되는 사고이다. AI 활용 영역별 사고의 대표적인 예로 자율주행차에 의한 교통사고, 수술 로봇이나 진단 로봇과 같은 의료 AI에 의한 의료사고, 실외이동로봇에 의한 충돌사고, 로봇어드바이저나 AI 설계사 등 금융 AI에 의한 금융소비자 권익 침해 사고 등이 있다. 한편 같은 분야라도 AI 활용 맥락이 다를 수 있다. AI에 의한 대출심사가 자동화된 의사결정 방식으로 이루어지는 경우와 AI 챗봇이 고객 상담 업무를 수행하는 경우는 같은 금융분야의 AI 활용이라도 그 맥락에 차이가 있는바 대응 방식도 달라야 할 것이다.

규제법 측면의 개별적 대응은 영역별로 상당히 구체화되어 있다. 자율주행차의 경우 자율주행차에 대한 각종 규제 특례를 정한 자율주행차 상용화 촉진 및 지원에 관한 법률(이하 ‘자율주행차법’이라 함)이 2019년에 제정되었고, 자율주행 단계별로 제작 안전기준도 마련되었으며, 최근에는 자율주행차 성능인증제도 및 적합성 승인제도가 시행되었다. 지능형 로봇의 경우 지능형 로봇 관련 규제 특례에 해당하는 지능형 로봇 개발 및 보급 촉진법(이하 ‘지능형로봇법’이라 함)이 2008년 제정되었고, 최근에는 실외이동로봇에 대한 운행 안전 인증 제도도 마련되었다. 금융분야의 AI 활용에 대해서는 ‘금융분야 AI 활용 가이드라인’이 2021년부터 시행 중이며, 최근에는 금융분야 AI 위험관리 프레임워크(AI RMF)도 마련되었다.⁴⁰⁾ 개인정보 보호법은 AI를 이용한 자동화된 의사결정에 의해 정보 주체의 권

39) 이에 대한 본격적인 논의는 박상철(2024), pp. 8-27를 참고 바람

40) 금융감독원 보도자료(2026. 1. 16.), “금융분야 AI 위험관리 프레임워크(AI RMF) 도입”

리가 침해되는 것을 방지하고 정보 주체의 권익을 보호하기 위해 정보 주체에게 거부권, 설명 요구권을 부여하고 있다.⁴¹⁾ 신용정보의 보호 및 이용에 관한 법률(이하 '신용정보법'이라 함)도 같은 맥락에서 사람의 관여 없이 정보처리장치만으로 개인신용정보 등을 처리하여 신용을 평가하는 '자동화 평가'와 관련하여 해당 정보 주체에게 설명 요구권, 정보제출권, 이의제기권을 부여하고 있다.⁴²⁾ 해외 주요국들도 대부분 영역 및 맥락을 고려한 개별적 규제를 채택하고 있는 것으로 보인다.⁴³⁾

책임법제 및 보험제도의 경우 사고 발생 시 불특정 다수의 생명·신체에 중대한 해를 끼칠 수 있는 사고를 중심으로 우선적인 대응 체계가 마련되고 있다. 자율주행차 사고 대응을 위해 2020년 자배법을 개정하여 자율주행차 사고에 대해서도 기존 자동차 사고 책임법제 및 보험제도가 적용되도록 하여 기존 자동차보험에서 자율주행차 사고도 보상하도록 하되, 사고 원인에 따라 보험회사가 책임자에게 구상을 청구할 수 있도록 하였다. 또한 특약을 통해 시스템 오류 및 해킹 등에 의해 발생한 사고도 자동차보험에서 보상할 수 있도록 하였다. 해외 주요국에서도 유사한 방식으로 자율주행차 사고에 대한 책임법제와 보험제도를 마련하였다. 실외이동로봇의 경우 보행자 충돌사고 발생 우려가 있다는 점을 고려하여 실외이동로봇 운영자에게 보험가입 의무를 부과하고 있다. EU는 최근 제조물책임지침(Product Liability Directive; PLD)을 개정하여 AI를 비롯한 디지털 신기술 관련 제조물사고의 특성을 고려하여 입증책임을 완화하고 제조업자의 책임을 확대하였다.

4. 소결

이상에서 본 바와 같이 AI 사고에 대한 대응은 사전 예방 체계와 사후 구제 체계, 포괄적 대응과 개별적 대응으로 구분해 볼 수 있다. 사고 대응 관련 법제도 정비 상황을 보면 사후 구제보다는 사전 예방 관련 제도가 우선적으로 마련되고 있는 것으로 보인다. 한편 기존에는 개별적 대응 위주로 제도 정비가 진행되었으나, 최근에는 AI 기본법, AI 규제법, AI 책임법, AI 보험제도 같은 포괄적 대응 방안에 대한 관심도 높아지고 있다.

41) 개인정보 보호법 제37조의2(자동화된 결정에 대한 정보 주체의 권리)

42) 신용정보법 제36조의2(자동화평가 결과에 대한 설명 및 이의제기 등)

43) 박상철(2024)은 미국, 중국, 영국, 일본, 이스라엘, 인도, 싱가포르, 스위스 등 AI 혁신 주도 국가들은 대부분 부문·맥락 특유적, 혁신친화적 규제 체계를 채택하고 있다고 설명하고 있음; 박상철(2024), pp. 14~16

〈표 II-4〉 AI 사고 대응 체계

구분	법제	포괄적 대응	개별적 대응
사전 예방	규제법제	<ul style="list-style-type: none"> • 인공지능 기본법(2026) • (EU) AI Act(2025) 	<ul style="list-style-type: none"> • 레벨3 자율차 안전기준(2019) • 자율차 성능인증 및 적합성승인(2025) • 실외이동로봇 운행안전인증(2024) • 금융분야 AI 활용 가이드라인(2021) • 금융분야 AI 위험관리 프레임워크(2026) • 자동화결정(신용평가) 관련 규제(2020) • 자동화된 결정(개인정보) 관련 규제(2024)
사후 구제	책임법제	<ul style="list-style-type: none"> • 인공지능책임법(안)(폐기) • (EU) AILD(초안)(철회) 	<ul style="list-style-type: none"> • 자배법(자율주행차 사고책임)(2020) • (EU) 제조물책임지침 개정(2025)
	보험제도	<ul style="list-style-type: none"> • 인공지능책임법(안)(폐기) • (EU) AILD(초안) 작성 과정에서 논의되었으나 법안에는 포함되지 않음 	<ul style="list-style-type: none"> • 자동차보험 자율주행차 특약(2020) • 실외이동로봇 보험(2024)

주: 괄호 안 연도는 관련 법령 시행 연도임

본고에서 주로 검토하고자 하는 것은 AI 사고에 대한 사후 구제 체계를 구성하는 책임법제와 보험제도이다. AI 사고 피해 구제를 위한 포괄적 대응방안과 개별적 대응방안의 도입 가능성을 검토해 보고, 당면한 AI 사고 위협에 대한 대응방안의 필요성 및 장기적인 제도 정비의 방향성을 모색해 보고자 한다.

1. 책임법제와 보험제도

가. AI 사고 책임법제와 보험제도의 역할

AI 사고 책임법제와 보험제도는 AI 사고 대응 체계 중 사후 피해 구제 체계의 중요한 축이다. 따라서 AI 책임법제와 보험제도는 AI 사고로 인한 피해를 효과적으로 구제할 수 있도록 마련되어야 한다. 효과적인 피해 구제의 핵심 요소는 책임의 공평한 배분과 피해자에 대한 신속하고 적절한 보상이다. AI 사고로 인한 책임을 공평하게 배분하면서도 AI 사고 피해자에게 신속하고 적절한 보상이 이루어질 수 있도록 하는 것이 AI 책임법제와 보험제도에 기대되는 역할일 것이다. 문제는 이러한 역할을 가장 잘 수행할 수 있는 구체적 방안을 마련하는 것이다.

현재 AI 사고 책임법제 및 보험제도와 관련된 논의의 핵심 쟁점은 기존의 책임법제와 보험제도로 AI 사고에 대응이 가능한지, 아니면 새로운 제도가 필요한지, 새로운 제도가 필요하다면 그 내용은 구체적으로 어떻게 마련해야 하는지 하는 것이다. 이하에서는 우선 현행 책임법제와 보험제도의 내용을 살펴보기로 한다.

나. 현행 책임법제

1) 일반법

현행 책임법제는 크게 일반법과 특별법으로 구분된다. 일반법에 해당하는 것은 민법 제5장의 불법행위 관련 규정으로, 사고의 유형을 불문하고 모든 경우에 적용된다. 민법상 불법행위 책임은 크게 3가지 유형으로 나누어 볼 수 있다. 첫째, 고의·과실로 위법행위를 하여 타인에게 손해를 가한 자, 즉 불법행위자에게 인정되는 책임이다(민법 제750조). 이는 가장 일반적인 형태의 불법행위 책임으로, '행위자 책임'에 해당한다.

둘째, 자신의 감독하에 있는 사람이 타인에게 손해를 가한 경우에 인정되는 책임이다. 미성년자나 심신상실자와 같이 타인에게 손해를 끼친 자가 스스로 그 책임을 부담하지 못하는 경우에는 그를 감독할 법정 의무가 있는 자가 손해배상책임을 부담한다(민법 제755조). 피용자가 사무집행과 관련하여 제3자에게 손해를 끼친 경우에는 사용자가 손해배상 책임을 부담한다(민법 제756조). 도급인의 지시에 따라 수급인이 업무를 처리하던 중 제3자에게 손해를 끼친 경우 도급인이 책임을 부담한다(민법 제757조). 이러한 감독자 책임, 사용자책임, 도급인 책임의 요건에는 차이가 있으나, 자신과 일정한 관계에 있는 타인의 행위로 제3자에게 손해가 발생한 경우 그에 대해 책임을 부담한다는 점에서 ‘감독자 책임’에 해당한다.

셋째, 자신이 점유하거나 소유하는 물건에 의해 손해가 발생한 경우, 그 물건의 점유자나 소유자가 책임을 부담한다(민법 제758조). 동물에 의해 피해가 발생한 경우 동물 점유자가 책임을 부담한다(민법 제759조). 점유자가 손해 방지를 위해 주의를 다한 경우에는 책임이 면제되기도 하지만, 원칙적으로 손해를 야기한 물건에 대한 권리를 기반으로 책임이 부과된다는 점에서 ‘권리자 책임’으로 볼 수 있다. 책임 귀속의 근거가 되는 가장 대표적인 권리는 소유권인바, 이하 편의상 ‘소유자 책임’이라 한다.

이처럼 우리 민법은 행위자 책임, 감독자 책임, 소유자 책임에 대해 정하고 있다. 즉, 내가 고의·과실로 타인에게 손해를 끼친 경우, 나의 감독 하에 있는 책임무능력자나 내가 고용한 피용자, 또는 내가 도급을 준 수급인이 타인에게 손해를 끼친 경우, 내가 점유하거나 소유한 물건이나 동물이 타인에게 손해를 끼친 경우, 나의 불법행위 책임이 인정되는 것이다.

이러한 책임은 AI 사고에 대해서도 적용된다. 내가 AI를 개발·활용하는 과정에서 고의·과실로 타인에게 손해를 끼친 경우, 내가 고용한 피용자가 AI를 개발·활용하는 과정에서 타인에게 손해를 끼친 경우, 내가 점유하거나 소유하고 있는 AI가 타인에게 손해를 끼친 경우, 나는 행위자, 사용자, 소유자·점유자로서 사고에 대해 책임을 부담한다.

2) 특별법

특별법은 특정 유형의 사고에 대해 적용하기 위해 별도로 마련된 책임법이다. 민법상 행위자, 감독자·사용자, 소유자·점유자가 아닌 자에게 책임을 귀속시켜야 하는 경우나, 민법에서 정한 요건에 따라 책임을 묻는 것이 적절하지 않고 별도의 요건을 마련할 필요가

있는 경우에 특별법이 마련된다. 대표적인 특별법으로 제조물책임법과 자배법이 있다.

제조물책임법은 제조물에 결함이 있는 경우 제조업자에게 책임을 귀속시킨다. 제조업자가 고의·과실로 결함 있는 제조물을 공급하여 손해를 야기했다면 민법상 행위자책임이 적용될 수 있으나, 제조업자의 고의·과실을 입증하는 것은 매우 어렵다. 이런 점을 고려하여 제조물책임법은 제조물에 결함이 존재하고 결함과 손해 사이에 인과관계가 존재하면 원칙적으로 제조업자의 고의·과실이 입증되지 않더라도 책임을 부담하도록 하고 있다(제조물책임법 제3조).

자배법은 자동차 사고에 대해 자동차 보유자에게 책임을 귀속시킨다. 자동차 보유자는 자신이 보유한 자동차가 사고를 일으킨 경우 자신이 운전을 하였는지 여부와 관계없이 원칙적으로 피해자에 대한 손해배상책임을 부담한다. 이를 '운행자책임'이라 한다. 자동차의 보유권 소유권 기타의 적법한 권한에 근거하여 자동차를 운행하는 것을 의미한다. '운전' 행위가 아닌 자동차 '보유'라는 권리에 기반한 책임이라는 점에서 민법상 소유자·점유자 책임과 유사하다.

3) AI 사고와 현행 책임법제

민법상 불법행위책임과 각종 특별법상 책임은 원칙적으로 AI 사고에 대해서도 동일하게 적용된다. 따라서 AI 사고가 발생했을 때 고의·과실로 위법행위를 하여 손해를 끼친 자가 있으면 그 행위자가 책임을 부담한다. 마찬가지로 감독자 책임, 소유자 책임, 제조물책임, 운행자책임 등 현행 책임법제에서 정한 요건에 따라 AI 사고에 대한 책임 귀속 여부를 판단하게 된다.

그런데 기존의 책임법제는 AI 기술을 고려하지 않은 것인바, 기존 책임법제를 AI 사고에 그대로 적용할 경우 책임의 공평한 배분과 신속하고 적절한 피해자 구제라는 책임법제 본래의 역할이 원활하게 수행될 수 있을 것인지가 문제 된다. 만약 기존의 책임법제가 AI 사고로 인한 손해를 공정하고 효과적으로 배분하지 못하거나, 법체계의 일관성이 훼손되거나, 사법 접근성에 제약이 발생하는 경우, AI 사고에 대해 기존 책임법제를 적용하는 것이 부적절할 수 있다.⁴⁴⁾

44) EG-NTF(2019), p. 34

다. 현행 책임보험제도

1) 책임보험의 의의

책임보험은 책임의 이행을 담보하는 보험이다. 책임에는 다양한 종류가 있으나 책임보험의 담보 대상이 될 수 있는 책임은 가해자가 피해자에게 부담하는 민사 손해배상책임이다. 형사책임과 같이 타인이 대신 이행하는 것이 허용되지 않는 책임은 보험의 담보 대상이 되지 않는다.

책임법제와 책임보험은 불가분의 관계에 있다. 책임법제에서 책임의 주체, 요건, 범위를 정하면, 책임보험은 그 책임주체를 피보험자로, 책임 요건을 보험사고 발생 요건으로, 책임 범위를 보험금 지급 범위로 하여 마련된다. 이처럼 책임법제는 책임보험제도의 전제가 되기 때문에 책임법제가 존재하지 않거나 불명확한 상황에서는 해당 책임을 담보하는 책임보험이 마련되기 어렵다. 앞서 본 일반법상 책임과 특별법상 책임에 대해서는 각각의 특성을 고려한 책임보험이 마련되어 있다. 자동차 사고에 대해 보유자가 피해자에게 부담하는 책임을 담보하기 위해 자동차보험의 대인 및 대물 담보가 마련되어 있다. 제조물책임법상 제조업자의 책임을 담보하기 위해 생산물배상책임보험이 제공되고 있다. 일반 불법행위책임 중 일상생활에서 발생할 수 있는 사고로 인한 책임은 일상생활배상책임보험으로 담보된다. 의료사고, 환경오염 사고, 개인정보침해 사고, 원자력 사고 등 특수한 사고에 대해서도 각각 관련 책임보험이 마련되어 있다. 위험성이 크거나 피해 구제 필요성이 큰 경우에는 보험가입이 의무화되기도 한다.

2) 책임보험의 역할

책임보험제도는 피해자의 실질적 구제에 기여한다. 책임법제가 마련되어 있더라도 책임보험제도가 뒷받침되어 있지 않으면 가해자의 배상자력 여부에 따라 피해자가 구제받지 못하거나, 가해자의 책임 여부를 다투는 동안 피해자 구제가 지연될 수 있는바, 이러한 피해 구제의 공백이나 지연을 방지함으로써 실질적으로 피해자가 보상을 받을 수 있도록 하는 것이 책임보험의 역할이다. 다른 한편으로, 책임보험은 가해자가 뜻하지 않은 사고로 인해 과도한 배상책임을 부담하여 경제적 파탄에 빠질 위험에서 벗어날 수 있도록 함으로써 가해자의 경제적 안정 도모에도 기여한다. 이처럼 책임보험의 핵심적 역할은 피해자를 구제하고 가해자를 보호하는 것이다.

3) AI 사고와 책임보험

AI 사고에 대해서도 책임보험제도는 사고 피해자를 신속하게 구제하고 AI 활용 주체가 과도한 배상책임 부담으로 경제적 파탄에 빠지는 것을 방지하는 역할을 수행해야 할 것이다. 이와 관련하여 기존의 책임보험제도를 통해 이러한 목적을 달성하는 것이 가능한지, AI 사고에 대응하기 위한 별도의 보험이 필요한지, 필요하다면 그러한 보험을 의무보험으로 운영할 필요가 있는지가 문제 된다.

2. 책임법제 및 보험제도 측면에서 본 AI 사고의 특성

가. 기존의 논의⁴⁵⁾

AI의 고유한 특성들은 책임법제 및 보험제도에 다양한 이슈를 야기한다. 이러한 특성은 AI 및 AI 사고의 의의를 정의할 때도 문제 되지만, AI 사고 책임법제 및 보험제도를 설계할 때에는 더욱 구체적인 검토를 요한다.

AI 사고 책임법제 및 보험제도에 관한 대표적인 연구인 EG-NTF(2019)는 책임법제 측면에 영향을 미치는 AI의 특성으로 ① 복잡성(Complexity), ② 불투명성(Opacity), ③ 개방성(Openness), ④ 자율성(Autonomy), ⑤ 예측가능성(Predictability)의 부족, ⑥ 데이터 주도성(Data-driven-ness), ⑦ 취약성(Vulnerability)을 제시하였다. 상술하면 다음과 같다.

기술 자체가 복잡해지고 기술 활용 과정에서 다양한 요인들이 복잡하게 상호작용함에 따라 사고 원인 및 사고 책임자 확정이 곤란해진다(복잡성). AI의 특성인 자기 학습과 블랙박스 효과는 AI 추론의 과정을 불투명하게 하여 고의·과실이나 인과관계 같은 손해배상책임 요건의 입증을 어렵게 만든다(불투명성). 기존의 제조물책임의 경우 제조물 공급 시점을 기준으로 결함 여부를 판단하는데, AI는 일단 시장에 공급된 이후에도 시스템 업데이트나 딥러닝을 통해 변경될 수 있어 제조물책임의 요건 및 판단 기준도 달라질 필요가 있다(개방성). 기존의 불법행위는 대부분 인간의 행위를 전제로 하는데, 인간의 개입 없이 자율적으로 작동하는 AI의 경우 책임 귀속 주체가 불분명해질 수 있다(자율성). 또한 AI의 자율성은 AI에 의한 작업 결과가 인간에 의해 100% 통제되지 않는다는 것을 의미하는바

45) AI 사고 책임에 관해 국내외 다수의 논문이 있으나 이를 모두 인용하기는 어려우나, 기존의 논의를 포괄적으로 검토한 EG-NTF(2019)와 이해원(2023)의 내용을 소개하는 것으로 같음함

예측가능성을 떨어트릴 수 있다(예측가능성). 데이터 자체의 오류나 편향에 기인해 인공지능이 산출한 결과에 문제가 발생할 수 있다(데이터 의존성). 상시 업데이트 및 외부 환경과의 상호작용은 사이버 보안 문제와 같은 취약성을 야기할 수 있다(취약성).⁴⁶⁾

EG-NTF(2019)는 이상과 같은 7가지 특성으로 인해 발생하는 책임법제 및 보험제도 이슈에 관해 18가지 분야에서 34가지 핵심 요소(Key findings)를 제시하고 있다.⁴⁷⁾ 동 보고서가 제시한 AI의 특성 중에는 자율성, 예측가능성 부족과 같은 AI 고유의 특성도 있고, 복잡성, 불투명성과 같이 첨단기술에 공통되는 특성도 있으며, 개방성, 데이터 주도성, 취약성처럼 첨단기술에서 공통적으로 나타나는 특성이나 AI 활용과 관련해서 특히 더 문제가 되는 특성들도 있다.

〈표 III-1〉 EG-NTF(2019)가 제시한 AI의 특성 및 책임법제 이슈

AI의 특성	책임법제 이슈
복잡성	기술 자체가 복잡해지고 기술 활용 과정에서 다양한 요인들이 복잡하게 상호작용함에 따라 사고 원인 및 사고 책임자 확정이 곤란해짐
불투명성	AI의 특성인 자기 학습과 블랙박스 효과는 AI 추론의 과정을 불투명하게 하여 고의·과실이나 인과관계 같은 손해배상책임 요건의 입증이 어려워 짐
개방성	AI는 일단 시장에 공급된 이후에도 시스템 업데이트나 업데이트를 통해 변경될 수 있어 제조물책임의 요건 및 판단 기준 시점(기존 책임: 제조물공급 시)이 달라질 필요가 있음
자율성	기존의 불법행위는 대부분 인간의 행위를 전제로 하는데, 인간의 개입 없이 자율적으로 작동하는 AI의 경우 책임 귀속 주체가 불분명해질 수 있음
예측가능성 부족	AI의 자율성은 AI에 의한 작업 결과가 인간에 의해 100% 통제되지 않는다는 것을 의미하는바 예측가능성을 떨어트릴 수 있음
데이터 주도성	데이터 자체의 오류나 편향에 기인해 인공지능이 산출한 결과에 문제가 발생할 수 있음
취약성	상시 업데이트 및 외부 환경과의 상호작용은 사이버 보안 문제와 같은 취약성을 야기할 수 있음

자료: EG-NTF(2019) 내용을 요약하여 정리함

46) EG-NTF(2019), pp. 32-34

47) 이에 관한 상세한 내용은 IV장에서 후술함

국내의 대표적인 연구로는 이해원(2023)이 있다. 이해원(2023)은 AI의 특성을 ① 자율성, ② 설명불가능성, ③ 예측불가능성, ④ 물리적 표출성으로 구분한다. 그중 설명불가능성은 규칙에 기반하지 않은 알고리즘의 특성, 데이터 의존성, 불투명성에서 야기된다고 보고 있다. 이러한 인공지능의 특성들이 인공지능에 의한 불법행위 상황을 해결하는데 야기하는 쟁점을 책임의 주체, 책임의 근거, 책임의 효과 측면으로 나누어 검토하고 있다.

나. AI 사고의 특성

피해 구제, 특히 책임법제와 책임보험의 대상이라는 측면에서 볼 때 AI 사고의 특성은 크게 4가지로 구분해 볼 수 있다. 첫째, 책임주체 특성의 곤란성, 둘째, 사고 원인 규명의 곤란성, 셋째, 사고 영역의 광범위성, 넷째, 피해 법익의 다양성이다. 이러한 특성은 AI 기술 자체의 특성에 기인하는 것도 있고, AI 기술의 활용 방식에 기인하는 것도 있다. AI 사고 피해 구제 방안을 마련하기 위해서는 이러한 특성에 대한 분석이 요구된다. 이하 항목별로 나누어 살펴본다.

1) 책임주체의 특정 곤란성

AI 사고는 사고를 유발한 '불법행위자'를 특정하기 어렵다. 일반적 책임법제인 민법상 불법행위책임은 고의·과실로 위법한 행위를 하여 타인에게 손해를 끼친 자인 불법행위자에게 책임을 부과한다. 이때 불법행위자에 해당하려면, 고의·과실이나 위법성 여부는 별론으로 하더라도, 일단 '행위'가 존재해야 한다. 그런데 AI는 인간의 개입이 없이 독립적으로 업무를 수행할 수 있는 특성인 자율성을 가지는데, AI가 스스로 인지-판단-제어 과정을 거쳐 만든 출력물이 타인에게 손해를 끼칠 경우, 손해 발생 과정에 인간의 '행위'라고 볼만한 것이 개입되지 않는다. 자율주행시스템이 자동차를 운전하던 중에 교통사고가 발생하는 경우가 대표적인 예이다. 이는 기존에 인간에 의해 이루어질 것이 전제되어 있는 특정 행위(예컨대 운전)가 AI 기술에 의해 완전 자동화되어, 결과적으로 해당 작업에 인간의 행위가 요구되지 않게 되어 발생하는 문제이다. 이러한 특성은 실제 잘못된 사람을 밝혀내야 하는 형사책임 분야에서 더 중대한 문제를 야기하며, 민사책임에서도 일반 불법행위 성립 여부를 판단하는 데 핵심적인 문제이다.

2) 사고 원인 규명의 곤란성

AI 사고는 사고의 원인을 규명하기가 어렵다. AI의 개발 및 이용 과정에서 손해가 발생했다는 현상은 존재하지만, 구체적으로 누구의 잘못으로 어떤 원인에 의해 사고가 발생하여 손해를 야기한 것인지를 밝히기가 어렵기 때문이다. 즉, 책임 요건 중 고의·과실 및 인과관계 입증에 어려워지는 것이다. AI의 복잡성, 불투명성, 예측불가능성, 개방성, 데이터 주도성, 취약성 등 앞서 본 특성 중 대부분이 사고 원인 규명의 곤란성과 관련된다. 전통적 알고리즘이 주어진 문제를 규칙에 따라 단계적으로 처리하여 결과값을 도출하는 순차적 알고리즘이라면, AI 알고리즘은 전통적 알고리즘과 달리 문제를 푸는 방법이 사전에 정해져 있지 않다. 따라서 알고리즘에 의해 얻은 결과가 정답인지 여부를 사후적으로 판단하는 것은 가능하지만, 왜 인공지능 알고리즘이 그러한 결과를 내놓았는지는 알기 어렵고, 그 결과 인공지능의 동작으로 손해가 발생하더라도 그것이 AI 오류 때문인지, 입력된 데이터에 문제가 있기 때문인지, 아니면 다른 문제 때문인지, 데이터에 문제가 있다면 개발 단계에서 사용된 데이터의 문제인지 아니면 출시 후 학습 단계에서 사용된 데이터 때문인지 판단하기가 어렵다. 규제법적 측면에서 AI 사업자에게 설명가능성을 갖출 것을 요구하고 있기는 하나, 규제법상 요구되는 설명가능성이 책임법상 인과관계 및 고의·과실의 근거가 되는 원인 규명으로 치환되는 것은 아니어서 여전히 문제가 남는다. 이러한 원인 규명의 곤란성은 AI 사고 자체의 고유한 특성이라기보다는 정도의 차이는 있지만 첨단 과학 기술이 적용된 분야에서 공통적으로 나타나는 현상이다.

3) 사고 영역의 광범위성

AI는 범용기술로 광범위한 분야에서 활용된다. 따라서 AI 사고 역시 광범위한 영역에서 발생한다. AI가 활용될 수 있는 분야라면 교통, 통신, 의료, 교육, 금융 등 모든 분야에서 AI 사고가 발생 가능한 것이다. 기존의 자동차 사고, 의료사고, 항공기 사고 등이 특정 영역에서 발생하는 점과는 차이가 있다.

기존의 피해 구제 체계는 사고 영역별로 구분되어 운영되고 있다. 자동차 사고의 경우 자배법과 자동차보험이, 의료사고의 경우에는 의료배상책임에 관한 판례 법리와 의료배상책임보험이, 항공기 사고에 대해서는 항공·철도 사고 조사에 관한 법률과 항공기보험이, 금융소비자 권익 침해에 대해서는 금융소비자보호법이 각각 피해 구제 방안을 마련하고

있다. 그런데 AI 사고의 경우 기존의 사고와 달리 특정 분야로 제한되지 않고 광범위한 영역에서 발생할 수 있다. 이러한 AI 사고에 대해 어떻게 피해 구제 체계를 마련할 것인지가 문제 된다.

4) 피해 법익의 다양성

AI 사고 결과 침해되는 법익도 다양하다. 기존의 사고의 경우 위험원의 유형이나 특징에 따라 침해되는 법익의 종류가 제한적이었다. 교통사고 경우 생명·신체·재산상 피해를 야기하고, 의료사고의 경우 환자의 생명·신체상 피해를 야기하며, 금융사고의 경우 재산상 피해를 야기한다. 반면 AI 사고에 의해 침해될 수 있는 법익에는 제한이 없다. 생명, 신체, 재산, 명예, 신용과 같은 전통적·개인적 법익은 물론, 초상권, 성명권 같은 인격권적 법익이 침해될 수도 있고, 선거제도, 민주주의, 지식 및 사상 시장의 신뢰와 같은 공공의 이익이 침해될 수도 있다. 피해 법익의 다양성은 사고 영역의 광범위성과도 관련되지만, 꼭 일치하는 것은 아니다. 예컨대 일반적으로 널리 사용되는 챗봇 형태의 생성형 AI의 경우, 저작권이나 초상권 등의 법익 침해가 일반적이나, 딥페이크 결과물이 사기에 악용되어 피해자의 경제적 손실이 발생하기도 하며, 이용자의 정서와 심리에 영향을 끼쳐 극단적 선택을 유도하는 사례도 있다.

3. AI 사고 책임법제의 쟁점

가. 피해 구제 체계 관련 쟁점

AI 사고 피해 구제 방안을 마련하기 위해서는 먼저 피해 구제 체계에 관한 검토가 필요하다. 사고에 대한 사후적 대응 체계로서의 피해 구제 체계는 앞서 살펴본 바와 같이 포괄적 대응 체계와 개별적 대응 체계로 구분된다. AI 사고의 특성 중 사고 영역의 광범위성 및 피해 법익의 다양성은 피해 구제 체계 측면에 영향을 미친다.

AI 사고에 대한 포괄적 대응 체계는 피해 구제 대상을 'AI 사고'로 보고 AI 활용 영역 및 맥락을 불문하고 모든 종류의 AI 사고에 대응할 수 있는 피해 구제 체계를 마련하는 것이다. 이러한 포괄적 대응 체계는 다시 기존의 책임법제 및 보험제도를 확장하여 적용하는 방안과, AI 사고를 대상으로 하는 별도의 책임법제와 보험제도를 마련하는 방안으로 나누어

볼 수 있다. AI 사고에 대한 개별적 대응 체계는 AI 활용 영역이나 맥락에 따라 피해 구제 체계를 세분화하는 것이다. 앞서 살펴본 자율주행차 사고 책임법제 및 보험제도, 실외이동로봇 보험제도 등이 개별적 대응에 해당한다.

책임법의 영역에서 포괄적 대응 방안에 대한 논의는 매우 더디게 진행되고 있다. 이는 기존의 포괄적 사고 대응 체계라고 할 수 있는 불법행위법을 탄력적으로 활용함으로써 AI 사고에 대응하는 것이 가능하다는 점, 성급하게 책임법을 도입할 경우 AI 산업 발전을 가로막고 실제 피해자 구제 및 공평한 책임 배분에 기여하지 못할 수 있다는 우려 등이 원인인 것으로 보인다. 반면 개별적 대응은 상대적으로 신속하게 이루어지고 있는 것으로 보인다. 자율주행차, 실외이동로봇 등 불특정 다수에게 피해를 야기할 수 있는 모빌리티 분야에서 선제적인 입법이 이루어지고 있다.

AI 사고에 대한 포괄적 대응과 개별적 대응은 반드시 양자택일의 관계에 있는 것은 아니다. 오히려 AI 발전 단계를 고려하여 포괄적 대응과 개별적 대응을 적절히 활용하고 양자가 조화를 이룰 수 있도록 하는 것이 중요하다. 규제법의 영역에서는 이미 포괄적 대응과 개별적 대응이 함께 마련되고 있다. 인공지능 기본법에서 AI 개발 및 활용에 관한 기본 원칙과 핵심 규제 사항을 정하면서 동시에 개별 영역의 특수성을 고려한 AI 활용 가이드라인 등이 별도로 마련되고 있다. 책임법 영역에서도 'AI 사고' 자체에 대한 대응 체계 마련과 개별 영역별 AI 사고에 대한 대응 체계가 함께 논의되어야 할 것이다.

나. 피해 구제 요건 관련 쟁점

AI 사고 특성 중 책임주체의 확정 곤란성 및 사고 원인 규명의 곤란성은 피해 구제 요건 측면 특히 책임주체, 책임 발생 요건으로서의 고의·과실 및 인과관계의 필요 여부 및 입증책임에 영향을 미친다.

1) 책임주체

기존의 책임법은 책임 귀속 주체가 자연인 또는 법인일 것을 전제로 하고 있다. 그런데 AI가 일정 수준의 자율성을 가지고 작업을 수행하던 중 피해가 발생한 경우, 누구에게 그 책임을 귀속시켜야 할 것인지가 문제 된다. 이에 관해 논의 초기에는 AI 자체에 법인격을 부여하여 그 AI에게 책임을 귀속시켜야 한다는 견해가 있었다. 전자인(電子人, Electronic

Person, E-Person)에 관한 논의가 그것이다. 법인이 책임주체인 경우 실제 그 법인의 구성원이나 주주가 최종적인 책임을 부담하는 것처럼, AI에 귀속된 책임도 실제로는 그 AI의 소유자나 등록명의자가 간접적으로 부담하게 된다. EU에서도 논의 초기에는 전자인 제도 도입 여부가 검토되었고, 우리나라에서도 21대 국회에서 전자인 제도를 도입하기 위한 법안이 제출되었다.⁴⁸⁾ 동 법안은 주식회사 등기제도와 유사한 전자인 등기제도를 마련하고, 등기되어 법인격을 취득한 전자인이 사고를 내면 그 책임은 전자인에게 귀속되도록 하되, 실제 책임은 전자인 운영자가 지게 하는 것을 내용으로 한다.

다만 현 시점에서 전자인 제도를 도입해야 한다고 보는 견해는 극소수인 것으로 보인다. AI 관련 법제를 가장 선제적으로 마련하고 있는 EU의 경우에도 법인격 부여에 대해서는 신중한 입장을 취하고 있다. 논의 초기에는 책임 공백 해소를 위해 AI에 법인격을 부여하는 방안이 검토되기도 하였으나,⁴⁹⁾ 이후 책임 귀속을 위해 AI에 법인격을 부여할 필요는 없다고 보는 것이 중론이 되었다(EG-NTF 2019). 우리나라에서도 단순히 책임 귀속을 위해 AI의 법인격을 인정할 필요는 없다고 보는 견해가 유력하다.⁵⁰⁾ 필요성이 없을 뿐 아니라 정당성도 인정되기 어렵다는 견해도 있다.⁵¹⁾ 장기적인 관점에서 법인격 부여가 필요하다고 보는 견해 역시 아직은 먼 미래의 일이라고 보고 있다.⁵²⁾

2) 엄격책임 부과 및 입증책임 완화

피해 구제 요건에 관한 가장 중요한 쟁점은 AI 사고에 대해 별도의 엄격책임을 인정할 것인지 여부이다. 엄격책임이란 과실책임과 대비되는 개념이다. 과실책임은 고의·과실의 존재를 책임의 성립 요건으로 하지만 엄격책임은 고의·과실이 입증되지 않더라도 다른 요건을 통해 책임을 인정한다. 고의·과실 요건을 아예 책임 성립 요건에서 제외하기도 하고, 고의·과실의 입증책임을 전환하여 가해자가 자신의 무과실을 입증하지 못하는 경우 책임을 인정하기도 한다. 앞서 본 일반법 중 행위자 책임은 과실책임에 해당한다. 감독자 책임, 소유자 책임, 특별법 중 제조물책임과 운전자책임은 엄격책임에 해당한다.

48) 전자인법안, 김용 의원 대표 발의, 의안번호 2126687; 임기만으로 폐기됨

49) European Parliament(EP)(2017)

50) 정진명(2024), pp. 65~87

51) 김진아(2022), pp. 273~308

52) 김진우(2021a), pp. 7~51

AI 사고에 대해 기존의 엄격책임을 통해 피해 구제가 가능한지, 아니면 AI 사고에 대한 별도의 엄격책임 도입이 필요한지가 문제 된다. 제조물책임 및 운전자책임과 같은 기존의 엄격책임은 당연히 AI 사고에 대해서도 적용된다. AI가 탑재된 제조물의 결함으로 사고가 발생했거나, AI가 탑재된 자동차의 운행 도중 사고가 발생한 경우에는 기존 엄격책임인 제조물책임과 운전자책임에 의해 피해 구제가 이루어진다. 이 경우 피해자는 AI의 출력물이나 출력물 생성 실패로 손해를 입었다는 사실만 입증하면 되고, 그 출력물 생성이나 출력물 생성 실패에 대해 제조사나 운영자의 고의·과실이 있었는지를 입증할 필요는 없다.

이러한 책임은 모든 AI 사고에 대해 적용되는 것은 아니고, 기존 제조물책임이나 운전자 책임의 범위 내에서만 적용된다. 따라서 모든 AI 사고에 대해 적용되는 별도의 엄격책임 법제가 필요한지 여부가 문제 된다. AI 사고에 대해 엄격책임을 도입하지 않더라도 최소한 AI 사고 피해자의 입증책임은 완화시켜 줄 필요가 있다고 보는 견해도 있다.

〈표 III-2〉 AI 사고의 특성 및 책임법제 이슈

AI의 개념 요소	AI 사고의 특성	내용
① 인간과 유사한 ‘기계지능’ ② 인간의 개입이 필요 없는 ‘자율성’ ③ 스스로 학습 및 개선하는 ‘적응성’ ④ 가상 및 현실 세계에 ‘출력물’ 제공 ⑤ 활용 분야에 제한이 없는 ‘범용성’	책임주체 특정 곤란성	요건 관련 쟁점 • 전자인 제도 도입 여부 • AI 운전자 엄격책임 부과 • 입증책임 완화
	사고 원인 규명 곤란성	
	사고 영역의 광범위성	체계 관련 쟁점 • 포괄적 대응과 개별적 대응
	피해 법익의 다양성	

4. AI 사고 책임보험의 쟁점

가. 책임법제상 쟁점과의 관계

AI 사고에 대한 책임보험의 쟁점은 책임법제 관련 쟁점과 연계된다. 피해 구제 체계와 관련하여 포괄적 체계를 도입하는 경우 그러한 포괄적 체계에 의한 책임을 담보할 수 있는 책임보험이 필요하고, 개별적 체계로 대응하는 경우 개별 영역별 책임보험이 마련되어야 한다. 기존의 책임법제로 AI 사고에 대응하는 경우 기존의 책임보험의 담보 범위를 확장하거나 특약을 부가하여 AI 사고가 담보되도록 해야 하고, 새로운 책임법제가 마련되는

경우 그 책임법제의 내용에 따라 책임보험을 새롭게 마련해야 한다. 만약 전자인 제도가 도입된다면 AI가 보험가입 주체가 되는 책임보험이 필요해진다. 이러한 책임법제와 책임보험의 관련성으로 인해 AI 사고 관련 책임보험에 대한 논의는 독자적으로 이루어지기보다는 책임법제 논의와 연계하여 이루어지고 있다.

나. 기존 보험 활용 방안과 신규 보험 도입 방안

AI 사고에 대해 별도의 책임법제를 도입하는 경우 그 책임법제의 내용을 반영한 새로운 보험이 필요해질 수 있다. 현재까지는 AI 사고에 대한 별도의 포괄적 책임법제는 입법되지 않은 상태이고 개별 영역별 책임법제가 먼저 마련되고 있는바, 보험도 개별 영역별로 먼저 정비되고 있다. 특히 모빌리티 영역을 중심으로 AI 사고 보험제도가 선제적으로 마련되고 있다.

대표적인 사례인 자율주행차의 경우, 기존의 책임법제인 자배법을 자율주행사고에도 적용한다는 전제에서 기존의 보험인 자동차보험을 통해 자율주행차 사고 피해 보상을 실시하되, 다만 자율주행차의 특성을 반영한 특약을 마련하여 자율주행시스템 오작동이나 해킹 등에 의한 사고 시 보상 방안을 명확히 하고 있다. 한편 실외이동로봇의 경우 별도의 책임법제는 마련되어 있지 않으나, 실외이동로봇이 불특정 다수에게 위험을 야기할 수 있다는 점을 고려하여 실외이동로봇보험을 별도로 마련하여 운영자에게 의무적으로 보험에 가입하도록 하고 있다.

이처럼 AI 사고 유형에 따라 기존 책임법제와 보험제도가 있는 경우에는 기존 보험을 확장하여 AI 사고 피해를 구제될 수 있도록 하고, 기존에 보험제도가 마련되어 있지 않은 경우에는 신규 보험을 마련하는 것이 필요할 것이다. AI 사고 영역별로 기존에 마련되어 있는 보험의 확장·보완을 통해 AI 사고에 대응하는 것은 기존 제도를 활용한다는 점에서 효율성과 안정성을 도모할 수 있다. 특히 고위험 영역으로 분류되고 있는 영역에 대해서는 대부분 기존에 별도의 책임법제와 보험제도가 마련되어 있고 AI 사고 영역별 특수성을 고려할 수 있다는 점에서 구체적 타당성도 높다고 볼 수 있다. 구체적인 제도 개선을 위해서는 각 분야별로 해당 분야의 AI 적용 가능성 및 AI 적용 시 사고 위험의 성격 및 크기의 변화 여부, 관련 책임법제, 보험제도 및 보험상품 확장·보완을 위한 구체적 방안을 개별적으로 검토해 보아야 할 것이다.

한편 피해자 구제 필요성이 높으나 기존의 보험제도가 마련되어 있지 않은 경우 별도 책임보험을 도입하는 방안이 가능하다. 특히 불특정 다수의 생명·신체에 위험을 야기할 수 있는 영역에서 보상 공백이 발생하지 않도록 책임보험 도입이 필요한 분야들을 일별하여 적시에 보험이 제공되도록 할 필요가 있다.

모든 AI 사고를 담보하는 포괄적 책임보험을 마련하는 방안도 고려해 볼 수 있다. 이는 EU에서 시도되었던 것으로 AI 사고에 대한 포괄적 책임법제 도입을 전제로 한다. AI 사고 전체 또는 고위험 AI 사고에 대해 AI 운영자에게 엄격책임을 부과하고, 해당 엄격책임의 이행을 담보하는 책임보험제도를 마련하는 것이다. AI 운영자의 책임의 요건은 자동차 운행자책임의 요건과 유사하게 정해지며, 책임보험 역시 자동차 사고 책임보험과 유사하게 구성된다.

장기적인 관점에서 포괄적 책임보험 도입 방안에 대한 검토도 필요하다. 다만 포괄적 책임보험제도 도입을 위해서는 충분한 사전 논의가 필요하고 기술 및 산업의 발전, 국제적 동향과의 조응도 고려해야 한다. 따라서 충분한 사전 검토와 논의를 진행하되, 제도 도입은 신중하게 결정해야 할 것이다.

다. AI 의무보험 도입 필요 여부

1) 의무보험 도입 논의의 전제

보험계약도 계약의 일종이고 계약 자유의 원칙상 계약 체결 여부 및 내용 형성은 당사자의 자유로운 의사에 따라야 한다. 의무보험은 계약 자유의 원칙에 대한 예외인바, 중대한 공익적 필요가 있는 경우 예외적으로 도입되어야 한다. 국민건강보험, 고용보험, 산재보험과 같은 사회보험의 경우 공익적 필요가 크므로 의무보험제도로 운영된다. 민영보험의 영역에서도 예외적으로 보험가입이 의무화되기도 하는데, 대부분의 의무보험은 피해자 구제를 위한 배상책임보험이다. 자동차 사고, 원자력 사고 배상책임보험, 의료기기 사고, 특수건물 화재 사고, 가스 사고, 청소년 수련시설 사고 등 관련 배상책임을 담보하기 위한 보험이 의무화되어 있다. 책임보험은 책임의 이행을 담보하는 것이고 이는 가해자가 배상 자력이 없는 경우 더욱 중요한 역할을 한다. 따라서 의무보험은 가해자의 자력에 비해 손해의 규모가 커서 가해자가 손해배상책임을 이행하기에 충분한 자력을 갖지 못하는 경우에 더욱 필요성이 높다.

요컨대 보험가입을 의무화하기 위해서는 의무화를 통해 연고자 하는 공익이 개인의 계약 체결의 자유를 제약할 만큼 충분히 크다는 점이 인정되어야 하고, 가해자의 배상자력에만 맡겨서는 피해 구제가 충분하지 못하게 될 위험이 있어야 한다.

2) AI 의무보험 도입 필요 여부

AI 의무보험 도입 논의의 쟁점은 AI 활용 전체에 대해 의무보험을 적용할지, 아니면 일부 형태에만 의무보험을 적용할지, 일부에 대해서만 적용한다면 그 기준은 어떻게 정해야 할 지이다.

범용기술인 AI는 인터넷, 전기와 같이 거의 모든 분야에서 활용될 수 있고, 같은 분야에서도 활용되는 형태는 다양하다. AI 활용 영역 및 맥락이 이처럼 다양한 상황에서, AI 의무보험 도입 필요 여부를 일률적으로 판단할 수는 없다. 활용 영역 및 맥락별로 의무보험 도입의 필요성을 개별적으로 검토해야 한다. 의무보험 도입 여부를 판단하기 위해서는 AI 사고 피해의 규모와 잠재적 가해자의 자력을 고려해야 한다.

EP(2020) 및 EG-NTF(2019)는 고위험 AI에 대해 엄격책임과 함께 의무보험을 도입하는 방안을 제안한 바 있다. 그러나 고위험 AI에 포함되는 다양한 영역에 대해 모두 의무보험 도입이 필요한지에 대해서는 의문이 있다.

예컨대 EU AI Act에서 고위험으로 분류되는 영역 중 기준에 위험성이 높다고 여겨지는 분야에서 AI 사고가 발생하는 경우, 사고 피해의 빈도와 심도가 크다. 기계, 장난감, 선박, 승강기, 폭발 관련 장비, 전파 장비, 압력 장비, 공중케이블, 개인보호장구, 가스 기구, 의료 기기, 항공기, 자동차, 농기구, 철도 등이 이에 해당한다.⁵³⁾ 이러한 분야에서 사고가 발생할 경우 주로 생명·신체에 직접적인 피해가 발생할 수 있다. 그중 AI 활용 주체가 개인이나 영세한 사업자일 것으로 예상되는 경우에는 잠재적 책임주체의 배상자력 부족을 보험을 통해 보완할 필요성이 크다. 이러한 영역에 대해서는 AI 보험을 의무화할 필요가 있다. 그런데 이처럼 위험도가 높고 배상자력 부족이 예상되는 분야의 경우 AI 활용 여부와 관계없이 이미 의무보험 도입 필요성이 인정되어 왔고, 실제로 대부분 의무보험제도가 도입되어 있다. 따라서 이러한 영역에서는 기존 의무보험이 해당 분야에서 발생하는 AI 사고까지 보장할 수 있도록 관련 제도를 보완하면 된다.

53) AI Act Annex I

한편 EU AI Act는 교육 및 노동 분야, 금융분야에서의 신용도 평가나 생명보험이나 건강 보험에서의 위험도 평가, 선거·행정·사법 등의 영역에서 AI를 활용하는 경우를 고위험으로 분류하고 있다.⁵⁴⁾ 이러한 영역에서 AI 사고가 발생하는 경우는 생명·신체에 직접적인 피해를 야기하지는 않고 재산상 손해, 평등권, 인격권 등이 침해될 수 있다. 잠재적 가해자는 기업, 금융기관, 정부 등으로 배상자력 부족이 우려되지 않는다. 따라서 이러한 분야에 대해 피해 구제를 위해 의무보험제도를 마련할 필요성은 상대적으로 낮다.

Faure and Li(2022)는 인공지능 사고에 대한 의무보험 도입 여부는 인공지능 자체의 위험성을 기준으로 정할 것이 아니라 사고에 대한 배상 의무자의 경제적 파산 가능성이 있는지를 기준으로 정해야 한다는 의견을 제시한다. 보험가입 의무화에는 상당한 비용이 소요되므로, 그 증거에 기반하여(Evidence based manner) 도입 필요 여부를 판단해야 한다고 본다. 또한 피해 구제 수단으로 보험 외에 자가보험이나 공제 등 보험 이외의 형태의 위험 분산 방안도 활용될 수 있어야 한다고 보고 있다.⁵⁵⁾

3) 소결

AI의 범용기술로서의 특성을 고려할 때 모든 AI 사고에 대해 보험가입을 의무화하는 것은 합리적이라고 보기 어렵다. 또한 고위험으로 분류되는 영역에서도 사고의 빈도, 심도, 피해 범위의 종류, 잠재적 책임자의 자력 등이 천차만별이다. 따라서 모든 AI 내지 모든 고위험 AI에 대해 의무보험 도입이 필요하다고 보기는 어렵다. 따라서 현재 마련되어 있는 의무보험 체계 내에서 AI 사고 위험을 담보할 수 있도록 하고, 불특정 다수의 생명·신체에 위험을 야기할 수 있음에도 기존 의무보험에 의해 담보되지 않는 영역에 대해서는 해당 분야에 적합한 의무보험을 도입하는 방향으로 대응하는 것이 합리적일 것이다.⁵⁶⁾

54) AI Act Annex III

55) Faure and Li(2022), pp. 16~19

56) 황현아(2024a), pp. 149~150

1. 주요국 동향

가. 개관

AI 사고 피해 구제 체계 및 요건에 관해 가장 선도적으로 법제를 정비하고 있는 곳은 EU이다. EU는 세계 최초로 AI에 대한 포괄적 규제법인 AI Act를 제정하였고, 포괄적 AI 책임 법인 AILD 초안을 발표하였으며, AI 사고의 특성을 반영하여 제조물책임지침을 개정하였다. EU 입법의 영향으로 우리나라도 인공지능 기본법을 제정하였고, 캐나다에서도 AI & Data Act 법안이 발의되었다.

반면 미국은 상대적으로 법제 정비가 본격적으로 추진되지는 않고 있다. 미국은 AI 이니셔티브법을 통해 AI 진흥에 관한 법적 체계를 마련하였으나, 규제나 책임에 관한 연방 차원의 입법 조치가 이루어지지는 않고 있다. 대통령의 행정명령을 통해 각 부처에 과제를 부과하는 방식으로 대응이 이루어지고 있으나, 바이든 정부에서 발효된 AI의 안전성과 신뢰성 확보를 위한 행정명령⁵⁷⁾이 트럼프 2기 정부 출범 직후 폐기되고⁵⁸⁾ 진흥 정책 중심의 새로운 행정명령⁵⁹⁾이 발표되는 등의 상황을 볼 때 아직까지 법적 안정성과 신뢰성을 갖춘 규제 및 책임 체계를 갖추었다고 보기는 어렵다. 연방 차원의 포괄적 규제 프레임워크의 부재로 개별 주 차원에서 AI 관련 법안이 급증하고 법적·제도적 혼란이 가중되고 있는 상

57) Executive Order 14110, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"; 바이든 정부의 행정명령은 미국 연방정부 차원에서 AI의 안전성, 보안성, 신뢰성을 확보하고, AI 기술 발전에 따른 잠재적 위험을 예방하기 위한 최초의 포괄적 규제 조치로, 2023. 11. 1. 발효되어 2025. 1. 20. 트럼프 대통령 취임과 동시에 폐기되었음

58) Executive Order 14148, "Initial Rescissions of Harmful Executive Orders and Actions"; 2기 트럼프 정부는 임기 첫날 바이든 정부의 행정명령 중 다양성·형평성·포용성 정책, 기후·환경 규제, 이민·국경 정책 등에 관한 68개 행정명령을 폐지한바, 바이든 정부의 AI 관련 행정명령인 EO 14110도 폐지 대상에 포함되었으며, 다만 AI 데이터센터에 대한 연방 지원 정책만 유지하도록 하였음

59) Executive Order 14179, "Removing Barriers to American Leadership in Artificial Intelligence"; AI 기업에 부과되는 안전성 평가·보고의무 등 규제를 중단하거나 재검토하고, 기업의 자율성과 혁신 환경을 강조하는 것을 주요 내용으로 하는 EO 14179를 발표하였음

황이라는 평가가 이루어지고 있다.⁶⁰⁾

일본의 경우 AI 개발·활용을 위한 제도 마련을 위해 AI 전략회의 및 AI 제도 연구회를 운영하고 있으나, 현 단계에서 구체적인 입법을 추진하기 보다 AI 도입에 따른 법제 정비 수요 등을 점검하고 중요한 사안에 대해서는 기존 법제에 따라 대응하는 것을 원칙으로 하되 필요한 지원을 제공하는 방안을 채택하는 등 산업 발전에 대한 효율적 지원을 중심으로 법제도 정비 전략을 수립하고 있다.⁶¹⁾

AI에 관해 주요국에서 다양한 입법적 노력을 기울이고 있으나, AI 사고 책임에 대해 체계적이고 구체적인 입법적 조치를 취한 것은 현재 EU가 유일한 것으로 보인다. 이하에서는 주요국 논의를 전반적으로 살펴본 후 EU AI 관련 입법의 체계와 내용에 대해 구체적으로 살펴보기로 한다.

나. 미국

1) 논의 경과

미국의 AI 관련 규제 논의는 2016년 10월 오바마 행정부가 'AI의 미래를 위한 준비 (Preparing for the Future of Artificial Intelligence)⁶²⁾' 보고서를 발표하면서 본격화되었다. 오바마 행정부는 해당 보고서를 통해 AI의 경제적·사회적 파급효과를 분석하고, 동시에 개인정보 보호, 알고리즘 편향, 고용 구조 변화, 국가안보 위협과 같은 윤리적·법적 쟁점을 주요 의제로 다루었다. 오바마 행정부는 규제 도입보다는 정책 연구 기반을 마련하고, 학계·산업계·시민사회와의 지속적 대화를 통해 향후 대응 방향을 모색하는 데 중점을 두었다.

이후 트럼프 행정부는 2019년 2월 '미국의 AI 리더십의 유지' 행정명령 제13859호(이하 'E.O. 13859'⁶³⁾)라 함)를 발령하여 본격적인 국가 전략을 수립하였다. 이 전략은 AI 연구개발에 대한 연방정부 지원 확대, 공공데이터 개방, 국제 협력 강화, 규제의 최소화를 핵심

60) 한국지능정보사회진흥원(2025a), p. 25

61) 한국지능정보사회진흥원(2025b), p. 14

62) White House(2016)

63) Executive Order 13859: Maintaining American Leadership in Artificial Intelligence

으로 하여 미국을 글로벌 AI 경쟁에서 선도적 위치에 두는 것을 목표로 하였다. 이를 제도적으로 뒷받침하기 위해 2020년 3월 'AI 이니셔티브법'이 발의되었으며, 이 법은 AI를 단순한 기술 혁신이 아니라 경제와 안보를 지탱하는 국가 차원의 전략 자산으로 규정하였다. 해당 법률을 통해 AI 연구·개발 및 정책 조정을 위한 국가 차원의 체계를 확립함으로써 미국의 AI 경쟁력 확보를 위한 장기적인 틀이 마련되었다.

바이든 행정부는 출범 이후 AI 위험관리와 사회적 신뢰성 확보에 중점을 두고, 2023년 10월 '안전·안보·신뢰성 있는 AI 개발과 이용에 관한 행정명령(이하 'E.O. 14110'⁶⁴)이라 함)'을 통해 안전·안보·신뢰성 있는 AI 개발 및 이용을 범정부적 과제로 설정하였다. 이 명령은 안전과 안보, 혁신과 경쟁, 노동과 시민권 보호, 소비자 권익, 프라이버시 보장, 공공 부문 활용, 국제 협력 등 다양한 분야를 포괄하는 정책 목표를 설정하였으며, 특히 생성형 AI와 이중용도 기반 모델의 위험 평가, 사이버보안 및 화생방핵(Chemical, Biological, Radiological, and Nuclear: CBRN) 위협 대응, 주요 기반시설 보호를 위한 리스크관리 체계를 포함하였다. 이와 함께 공공 조달 확대 및 인재 유치를 통해 미국 내 AI 생태계를 강화하려는 정책적 노력이 병행되었다. 바이든 정부의 접근은 포괄적이고 일률적인 규제보다는 각 부문과 맥락에 맞춘 맞춤형 규율 및 유연한 거버넌스를 지향한 것이 특징이었다.

반면 2025년 출범한 트럼프 행정부 2기는 바이든 정부의 E.O. 14110을 '급진 좌익적 발상'으로 평가하며 규제 중심의 접근을 철회하고,⁶⁵ 2025년 7월 'AI 액션 플랜(America's AI Action Plan)⁶⁶'을 발표하는 동시에 ① 연방정부 조달 시 편향 배제,⁶⁷ ② 데이터센터·반도체·에너지 설비 등 AI 인프라 구축 허가 절차 간소화,⁶⁸ ③ 미국 AI 기술 동맹국 수출

64) Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

65) Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence)

66) White House(2025)

67) Executive Order 14319: Preventing Woke AI in the Federal Government; 기술적 편향이 아닌 정치적·이념적 중립성을 연방정부가 조달하는 AI 시스템에 의무화하겠다는 뜻으로 대규모 언어모델(Large Language Model; LLM)이 특정 정치적·사회적 가치관에 편향되지 않도록 이념적 편향(Woke bias)을 차단하고 연방정부가 사용하는 AI는 객관적이고 사실에 기반한 응답을 제공해야 한다는 원칙을 강조함. 이에 따라 정부 기관이 LLM 등을 구매 또는 활용할 때 공급업체는 자사의 AI가 객관적 진실(Objective truth)에 기반한 응답을 제공한다는 것을 보증해야 함

68) Executive Order 14318: Accelerating Federal Permitting of Data Center Infrastructure; 5억 달러 이상 투자 또는 100MW 이상 전력 부하 등의 요건을 충족하는 AI 데이터센터·반도체·에너지 프로젝트에 대해 국가환경정책법(National Environmental Policy Act of 1969; NEPA) 등 적용되는 환경 규제를 완화하고, 연방 토지와 오염 부지를 활용하며, 대출·보조금·세제 인센티브를 제공함

및 경쟁국 제한⁶⁹⁾을 위한 세 가지 행정명령을 발령하며 경제 성장 및 국제 경쟁력 확보를 우선시하는 새로운 방향을 제시하였다. 또한 이를 실행에 옮기기 위해 2025년 11월 AI 부문의 ‘맨해튼 프로젝트(Manhattan Project)’로 불리는 범정부적 이니셔티브 ‘제네시스 미션(Genesis Mission)’을 공식 출범⁷⁰⁾하는 등 AI 액션 플랜에 따른 행보를 이어가고 있다.

2) America’s AI Action Plan의 주요 내용

AI 액션 플랜은 크게 ① AI 혁신 가속화(Accelerate AI Innovation), ② 미국 내 AI 인프라 구축(Build American AI Infrastructure), ③ 글로벌 AI 외교 및 안보 관련 리더십 확보(Lead in International AI Diplomacy and Security)의 세 축으로 구성되어 있다.

첫 번째로 ‘AI 혁신 가속화’ 부분에서는 연방 및 주 차원에서 규제를 최소화하고, 오픈소스 모델과 데이터셋 공유를 적극 장려하며, 연방정부 조달 시 ‘이념적 편향 없는 객관적 진실’을 기준으로 삼아 대규모 언어모델(LLM) 공급 요건을 재설정하였다. 이를 통해 규제 부담을 줄이는 동시에 민간과 학계가 공공 영역 수요에 적극 참여하도록 유도하여 경쟁력을 높이려는 의도를 담고 있다.

두 번째로 ‘미국 내 AI 인프라 구축’ 부분에서는 데이터센터, 반도체 공장, 에너지 설비 등 대규모 인프라 확충을 위해 허가 절차를 간소화하고, 연방 토지를 인프라 용도로 개방하는 방안이 포함되었다. 또한 전력 수요 증가에 대응하기 위해 원자력과 재생에너지를 동시에 확대하고, 직업 훈련과 산학 연계 교육, 해외 우수 인재 유치 정책을 통해 AI 전문 인력을 확보하려는 노력이 병행된다.

마지막으로 ‘글로벌 AI 외교 및 안보 관련 리더십 확보’ 부분에서는 동맹국과 AI 기술과 표준을 공유하여 민주주의 진영 내 기술 협력을 강화하는 동시에, 중국 등 경쟁국에는 핵심 기술 접근을 제한하는 이중 전략을 채택하였다. 이를 통해 미국은 AI를 단순한 기술 경쟁 수단이 아닌 국가안보와 직결되는 핵심 기반으로 명확히 규정하고 있다.

AI 액션 플랜에 대해서는 안전성과 신뢰성을 희생하고 단기 성장 및 경쟁력 확보에만 집

69) Executive Order 14320: Promoting the Export of the American AI Technology Stack; 동맹국에 미국의 AI 기술(모델·반도체·데이터센터 인프라 등)을 적극 수출하고, 중국 등 경쟁국에 대해서는 첨단 AI 기술 접근을 차단하는 이중 전략을 명문화한 행정명령임

70) White House, LAUNCHING THE GENESIS MISSION(<https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/>) (2026. 2. 4. 최종 접속)

중하다는 비판적 시각이 있으며, 대규모 데이터센터 건설에 따른 환경 부담, 노동·소비자 보호 부족, 콜로라도주 AI법과 같은 주(州) 차원의 AI 규제와의 충돌 가능성⁷¹⁾에 대한 우려도 제기되고 있다. 또한 AI 액션 플랜은 AI 기술 및 산업 진흥에만 중점을 두어 AI 사고에 대한 책임 및 피해 구제 내용은 반영되어 있지 않으며, AI 개발자 및 사업자에 대한 과태료 등 제재 방안에 대해서도 정하고 있지 않다.

다. 영국

1) 논의 경과

영국은 2021년 9월 ‘국가 AI 전략(National AI Strategy)⁷²⁾’을 발표하여 향후 10년간 ① AI 생태계에 대한 장기적 투자 및 계획 수립을 통해 과학·AI 강국의 지위를 유지하고, ② AI 혁신의 혜택을 전 산업·지역으로 확장하며, ③ 적절한 국내외 거버넌스 확립을 통해 혁신과 투자를 장려하는 것을 목표로 하였다. 이 전략은 장기적인 비전을 중심으로 인프라, 인재, 거버넌스 전반을 아우르는 기초 설계에 초점을 두었다.

영국 정부는 2022년부터 2023년까지 AI 규제 체계 설계와 관련하여 심화된 논의를 진행하였고, 2023년 3월 ‘AI 규제에 대한 혁신적 접근 방식(A pro-innovation approach to AI regulation)⁷³⁾’이라는 제목의 AI 백서를 발표하였다. 이 백서는 EU AI Act와 같은 포괄적 단일 법률이 아닌 원칙 중심의 부문별 접근 방식을 제시하였는데, AI의 안전하고 혁신적인 사용을 촉진하기 위해 고려해야 할 5가지 원칙으로 ① 안전, 보안 및 견고성(Safety, security and robustness), ② 투명성 및 설명 가능성(Appropriate transparency and explainability), ③ 공정성(Fairness), ④ 책임 및 거버넌스(Accountability and governance), ⑤ 경쟁 가능성 및 보상(Contestability and redress)을 명시하였다. 또한 AI 혁신을 저해하는 강압적인 법안을 피하고, 새로운 단일 규제기관을 만드는 대신 기존에 있던 관련 기관들에 권한을 부여하여 상황에 맞는 맞춤형 접근 방식을 제시하였다.

71) 콜로라도주 AI법(SB 24-205, Colorado AI Act)은 AI에 대해 강력한 책임 및 투명성 요건을 부과하고 포괄적 규제를 도입하고 있는 반면, AI 액션 플랜은 연방 차원의 규제 완화 정책을 포함하고 있어 주 차원의 규제 조치를 과도한(Burdensome) 규제로 간주함

72) HM Government(2021); 해당 전략은 AI가 COVID-19 대응에 기여하였음을 언급하며, NHS와 같은 보건 분야에도 혁신적인 AI가 상용화될 수 있도록 하는 것을 목표로 함. 이에 따라 NHS AI 연구소(NHS AI Lab)는 ‘보건 및 사회복지 분야 AI 국가 전략(National Strategy for AI in Health and Social Care)’을 수립함

73) Department for Science, Innovation & Technology(DSIT)(2023a)

2023년 하반기 영국 정부는 고위험·범용 AI(Frontier AI)의 확산에 대응하기 위해 ‘프런티어 AI 태스크포스(Frontier AI Taskforce)’를 출범하였다.⁷⁴⁾ 이 조직은 프런티어 AI의 잠재적 위험을 연구하고 사전평가하는 기능을 전담하며, 국가 차원의 안전 관리 인프라를 구축하는 출발점이 되었다. 같은 해 11월, 영국은 제1회 AI 안전 정상회의(AI Safety Summit)를 주최하고 ‘블레츨리 선언(Bletchley Declaration)⁷⁵⁾’을 채택해 다수 국가 간 프런티어 AI의 위험 평가 및 안전성 연구를 위한 국제 협력 체계를 마련하였다. 이를 계기로 프런티어 AI 태스크포스가 ‘AI 안전연구소(AI Safety Institute; AISI)’로 격상되었으며, AI 관련 평가 및 검증 등에 관한 실질적인 기능을 수행하는 핵심 기관이 되었다.⁷⁶⁾

이외에도 경쟁시장청(Competition and Markets Authority; CMA)은 2023년 ‘기초모델 경쟁·소비자 영향 분석⁷⁷⁾’을 실시하고 2024년 2차 보고서를 통해 경쟁·소비자보호 원칙을 제시하였으며⁷⁸⁾ 정보위원회(Information Commissioner’s Office; ICO)는 2023년 ‘AI와 데이터 보호 가이드라인(Guidance on AI and data protection)⁷⁹⁾’을 개정하여 AI 개발·운영 단계별 개인정보보호 원칙과 법적 의무를 구체적으로 명시하였다. 2023년 제정된 ‘온라인 안전법(Online Safety Act)’은 AI 생성 콘텐츠를 포함한 불법·유해 정보 확산 방지를 위한 플랫폼의 의무를 규정하였다.

영국은 이상의 정책 및 제도적 대응을 축적해 나가며 AI 관련 영국의 독자적 규제모델을 형성하였고, 이를 반영하여 2025년 1월 AI 산업 및 인프라 육성을 위한 실행 계획인 ‘AI 기회 행동 계획(AI Opportunities Action Plan)⁸⁰⁾’을 발표하였다.

74) Department for Science, Innovation & Technology(DSIT)(2023b); Department for Science, Innovation & Technology(DSIT)(2023c)

75) Department for Science, Innovation & Technology(DSIT)(2023d)

76) UK Government(2023)

77) CMA(2023)

78) CMA(2024); Initial Report에서 7대 원칙으로 책임(Accountability), 접근성(Access), 다양성(Diversity), 선택권(Choice), 유연성(Flexibility), 공정 거래(Fair dealing), 투명성(Transparency)이 제안되었으나, ‘유연성’이 ‘선택권’으로 통합되어 최종적으로 6대 원칙이 제시됨

79) ICO(2023); 2025년 6월 발효된 ‘Data(Use and Access) Act’에 따라 변경 검토 중임

80) Department for Science, Innovation & Technology(DSIT)(2025)

2) AI opportunity action plan의 주요 내용

AI 기회 행동 계획은 향후 10년간 국가 전반에 걸쳐 AI 역량을 확장 및 심화하기 위해 마련된 종합 실행 전략으로, 연구개발 지원, 인프라 확대, 인재 확보 및 규제 혁신 등 AI를 최적으로 설계하기 위한 50가지 실행 계획이 포함되어 있다. 영국 정부는 동 계획에서 ① AI 활성화 생태계 조성, ② 공공·민간 분야의 AI 활용 가속화, ③ 자주적인 AI 추진이라는 세 가지 방향성을 제시하고 있다.

영국 정부는 먼저 첨단 AI 연구개발(R&D)에 대한 투자를 확대하고, 국가 AI 연구소 및 대학·산업 간 협력 네트워크를 강화하여 기초·응용 연구 분야에서 세계적 경쟁력을 확보하고자 한다. 이를 위해 AI 연구자원(AI Research Resource: AIRR)의 수용력을 확대하고, AI 데이터센터 구축 가속화를 위해 AI 성장구역(AI Growth Zones: AIGZs) 조성 계획을 밝혔다. 또한 고품질 데이터의 확보 및 제공을 위해 공공 및 민간 부문을 장려하고, AI 생태계에 필요한 전문 인재를 장기적으로 확보하기 위해 교육 및 훈련 프로그램 확대 및 인재 유치를 위한 방안을 마련할 계획이다.

다음으로 전 산업 및 공공 부문의 AI 도입 가속화를 위해 정부 조달, 시범 사업, 국책 과제를 통해 공공 부문이 초기 수요를 창출하며, 이를 민간 확산의 발판으로 삼고자 하였다. 특히 정부가 탐색(Scan) → 시험 적용(Pilot) → 대규모 확산(Scale) 접근 방식을 채택하여 영국 내 스타트업 및 혁신 생태계를 지원하고, 공공 부문과 민간 부문이 상호 보완적인 역할을 수행하여야 함을 강조하며, AI 도입을 통해 지역별·기업별 격차를 해소할 수 있도록 지원하기로 하였다.

끝으로, 수동적으로 AI를 받아들이는 것이 아닌, 자주적인 AI를 개발하여 최첨단 AI 분야에서 영국의 점유율을 극대화하는 것을 목표로 하였다. 이를 위해 UK Sovereign AI를 창설하고, 민간 부문과 협력할 수 있는 권한을 부여하였다. 이 부서는 민간 및 학계가 국제 협력과 합작 투자, 육성 등을 통해 각각의 강점을 최대한 발휘할 수 있도록 지원하며, Innovate UK의 지원을 받아 새로운 회사의 설립, 스타트업 규모 확장, AI 기술 발전 등에 투자할 수 있는 권한을 갖고 있다.

본 계획은 영국 정부가 AI에 대해 장기적인 계획을 가지고 즉각적인 조치를 취할 것임을 밝히고 있으나, AI 사고에 따른 책임 규명이나 피해 보상 절차, 개발자 또는 사업자에 대한 법적 제재와 관련된 내용은 포함하고 있지 않다. AI 관련 포괄적 규제에 해당하는 법안

인 Artificial Intelligence (Regulation) Bill⁸¹⁾가 상원에 발의되어 있으나, 정부 주도 입법이 아니라 의원 개인 발의 입법이라는 점 및 AI 혁신 친화적이고 기업 우호적인 영국 정부 정책 방향과 조화되지 않는다는 점에서 통과 가능성은 높지 않은 것으로 여겨지고 있다.⁸²⁾ 영국 정부는 2026년 포괄적 AI 패키지 법제 도입을 검토하겠다고 하였으나 아직 초안이 발표되지 않은 상태이며, 패키지에 AI 사고 책임에 관한 내용이 포함될지는 불분명한 것으로 보인다. 한편 영국 법률위원회(Law Commission)는 AI 도입 및 활용을 고려한 제조물책임법 개정을 검토 중에 있으며, 주된 쟁점은 EU의 개정 PLD와 대체로 유사한 것으로 보인다.⁸³⁾

라. 일본

1) 논의 경과

일본의 AI 관련 제도 논의는 2015년 12월 ‘Society 5.0’ 개념의 구상에서 출발하였다. 2016년 ‘제5기 과학기술기본계획(第5期科学技術基本計画)⁸⁴⁾’에서 처음 제시된 이 비전은 물리적 공간과 사이버 공간이 통합된 초연결·초스마트 사회 실현을 위해 AI의 중요성을 강조하였으며, 국가 차원에서 AI의 연구개발 및 사회적 활용을 추진할 필요성이 있음을 언급하였다. 이후 2021년 9월, 일본 정부는 국가 디지털 전환의 컨트롤타워로 디지털 청(デジタル庁)을 출범⁸⁵⁾시키고 공공서비스와 사회 인프라 전반에서 AI 활용 기반을 확충하기 시작하였다.

국제적으로 AI 기술이 급속히 발전하고 ChatGPT 등 생성형 AI가 사회적 논의를 촉발한 2023~2024년 무렵, 일본 내에서도 AI의 잠재적 위험과 이에 대응하기 위한 AI 규제 필요성 논의가 대두되었다. 2023년 5월 일본 정부는 G7 정상회의를 계기로 ‘히로시마 AI 프로

81) UK Parliament, Artificial Intelligence (Regulation) Bill [HL](<https://bills.parliament.uk/bills/3942/>) (2026. 2. 4. 최종 접속)

82) Clyde&Co(2025), “The relaunched UK AI Regulation Bill – a step towards statutory regulation of AI in the UK? Insight Article”(<https://www.clydeco.com/en/insights/2025/03/the-relaunched-uk-ai-regulation-bill-a-step-toward>) (2026. 2. 20. 최종 접속)

83) Clyde&Co(2024), “Product Liability legislative reforms: A snapshot of the UK’s position”(<https://www.clydeco.com/en/insights/2024/12/product-liability-legislative-reforms>) (2026. 2. 20. 최종 접속)

84) 内閣府(2016)

85) ‘디지털청설치법(デジタル庁設置法)’에 따라 설립된 정부 기관으로, 법 자체에 ‘AI(人工知能)’라는 용어가 직접적으로 등장하지 않으나 디지털 사회의 형성, 정보 시스템의 정비, 행정의 디지털화 등을 목적으로 하는 기관으로서 시를 포함한 최신 ICT 기술 활용에 관한 부분이 업무 범위 중 하나임

세스(広島AIプロセス)를 발족시켰고, 이후 이 프로세스는 49개국과 지역이 참여하는 자발적 국제 협력체로 발전하여 AI 거버넌스의 글로벌 프레임워크 구축을 목표로 하였다. 2024년 4월에는 일본전신전화 주식회사(Nippon Telegraph and Telephone Corporation; NTT)와 요미우리신문이 생성형 AI의 규제 및 조치 필요성에 관한 공동성명을 발표⁸⁶⁾하며 정부에 제도 및 대응 방안 마련을 촉구하기도 하였다.

이 시기 일본은 법률보다는 가이드라인 중심의 '소프트 로(Soft law)' 접근을 유지하였다. 총무성(総務省)은 '국제적인 논의를 위한 AI 개발 가이드라인(国際的な議論のための AI 開発ガイドライン案)'⁸⁷⁾, 'AI 활용 가이드라인(AI 利活用ガイドライン)'⁸⁸⁾, 'AI 활용 가이드라인: AI 활용을 위한 실질적인 레퍼런스(AI 利活用ガイドライン~AI 利活用のためのプラクティカルリファレンス~)'⁸⁹⁾를, 경제산업성(経済産業省)은 'AI 원칙 실천을 위한 거버넌스·가이드라인 Ver. 1.1(AI 原則実践のためのガバナンス・ガイドライン Ver. 1.1)'⁹⁰⁾ 등을 공표하였으며, 2024년 4월에는 양 부처가 공동으로 'AI 사업자 가이드라인(AI 事業者ガイドライン案)'⁹¹⁾을 발표하였다. 그러나 이러한 노력에도 불구하고 데이터 수집 및 개인정보 보호, 저작권 및 의료 분야 등에서 법적 공백 및 논란이 계속되었다.

이러한 사회적 요구를 바탕으로 2025년 6월 일본은 'AI 관련 기술의 연구개발 및 활용 추진에 관한 법률(人工知能関連技術の研究開発及び活用の推進に関する法律)'을 제정·공표하였다.

2) AI 관련 기술의 연구개발 및 활용 추진에 관한 법률의 주요 내용

일본의 'AI 관련 기술의 연구개발 및 활용 추진에 관한 법률(이하 '일본 AI 법'이라 함)'은 AI 관련 국가의 기본 방침을 정하는 기본법적 성격을 가진 일본 최초의 AI 종합 기본법으로, AI 기술의 연구개발부터 사회적 활용까지 국가 차원에서 종합적·체계적 AI 활용을 추진하기 위한 기본 이념과 시책, 추진 체제를 규정한다.

86) 株式会社読売新聞グループ本社, 日本電信電話株式会社(2024)

87) AI 네트워크 사회推進會議(2017)

88) AI 네트워크 사회推進會議(2019a)

89) AI 네트워크 사회推進會議(2019b)

90) AI 原則の実践の在り方に関する検討会, AI ガバナンス・ガイドライン WG(2019)

91) 総務省, 経済産業省(2024); 현재는 2025. 3. 28. 개정된 1.1판이 최신 버전임

법률은 총 4장으로 구성되어 있는데 제1장과 제2장은 2025년 6월에 시행되었고, 제3장과 제4장은 9월 1일부터 시행되었다. 제1장 ‘총칙’에서는 법률의 목적과 AI 관련 기술의 정의, AI 관련 기술의 연구개발 및 활용의 추진을 위한 기본 이념 및 국가, 지자체, 연구개발 기관, AI 활용 사업자 및 국민의 책무를 규정하고, 제2장 ‘기본적 시책’에서는 연구개발 환경 조성 및 데이터 및 인프라 확충, 인재 양성, 윤리 및 적합성 확보, 국제 협력 등에 관해 규정하며, 제3장(제18조)에서는 정부의 AI 기본계획 수립 의무를 규정한다. 제4장은 총리를 본부장으로 하는 AI 전략본부를 설치하도록 하여 부처 간 조정, 기본계획 실행, 협력 등을 총괄하도록 하였다.

일본 AI 법은 EU의 ‘AI Act’처럼 금지 규정이나 제재 규정을 중심으로 한 규제형 법률이 아니라 정부의 정책 방향과 행정 조직의 역할을 명확히 하는 기본법 성격이 강하며, AI 활용에 대한 사회적 기대와 우려를 균형 있게 반영하면서도 혁신 촉진을 저해하지 않는 소프트 로(Soft law) 접근 방식을 유지한 것이 특징이다. 즉, AI 관련 사고에 대한 법적 책임이나 피해 구제에 관한 내용 또는 AI 사업자에 대한 제재 등에 대한 조항은 포함되어 있지 않다. 제7조에서 ‘활용 사업자의 책무’를 규정하고 있으나, 이는 국가 정책에 대한 협력 의무에 관한 것으로 위반 시 벌칙이나 제재 조항은 없다.

마. 소결

주요국이 다양한 입법적 노력을 기울이고 있으나, AI 사고 책임에 대해 체계적이고 구체적인 입법적 조치를 취한 것은 현재 EU가 유일한 것으로 보인다. 이하에서는 EU AI 관련 입법의 체계와 내용에 대해 구체적으로 살펴보기로 한다.

〈표 IV-1〉 주요국 AI 법제 동향(요약)

국가명	주요 동향
미국	<ul style="list-style-type: none"> • 2016년 오바마 행정부의 ‘인공지능의 미래를 위한 준비’ 보고서를 통해 규제 논의 본격화 • 2019년 트럼프 행정부의 미국의 AI 리더십 유지 행정명령을 통해 국가 전략 수립(정부 지원 확대, 규제 최소화) • 2023년 바이든 행정부 출범 후 안전, 안보, 신뢰성 있는 인공지능 개발과 이용에 관한 행정명령 발표 • 2025년 트럼프 행정부 출범 후 바이든 정부의 행정명령 철회 및 AI 산업 발전 및 국제 경쟁력 확보를 최우선으로 하는 정책 추진

〈표 IV-1〉 계속

국가명	주요 동향
영국	<ul style="list-style-type: none"> • 2021년 국가 AI 전략을 발표하여 장기 투자 계획 수립 • 2023년 'AI의 규제에 대한 혁신적 접근 방식'을 발표하여 원칙 중심 부문별 규제 방식 제시 • 2023년 하반기 고위험, 범용 인공지능 확산 대응을 위한 프런티어 AI 태스크포스 출범 • 2023년 11월, 제1회 AI 안전 정상회의 주최 및 블레츨리 선언 채택으로 AI 위험 평가 및 안전성 연구를 위한 국제 협력 체계 마련 • 2025년 산업 및 인프라 육성을 위한 AI 기회 행동 계획 발표
일본	<ul style="list-style-type: none"> • 2016년 제5기 과학기술기본계획에서 AI 관련 제도 논의 시작 • 2023년 G7 정상회의를 계기로 히로시마 AI 프로젝트 발족(49개국 참여 AI 거버넌스 글로벌 프레임워크 구축 목표) • 연성규제 중심 규제 마련(AI 활용 가이드라인, AI 원칙 실천을 위한 거버넌스 가이드라인, AI 사업자 가이드라인 등) • 2025년 6월, 인공지능 관련 기술의 연구개발 및 활용 추진에 관한 법률(진흥법) 제정

2. EU AI 관련 입법 현황

EU의 AI 법제는 사고에 대한 사전 예방 수단인 규제법과 사후 구제 수단인 책임법으로 구분된다. 입법 순서상으로는 규제법이 먼저 정비되었고 책임법은 아직 논의가 진행 중인 상황이다. 법제 내용상으로는 규제법에서 정해진 AI 사고 위험, 이해관계자의 역할과 책무를 전제로 책임법 논의가 이어지고 있기도 하고, 역으로 책임법에 대한 고려가 규제법 내용 형성에 영향을 미치기도 한다.

EU AI 법제의 핵심은 EU AI Act이다. AI Act는 AI가 활용되는 모든 영역에 적용되는 포괄적 규제법이다. AI Act는 위험도에 따라 AI를 4등급으로 분류하여 차등 규제를 적용하는 '위험 기반 접근 방식'을 채택하고 있다. 특히 고위험 AI 개발 및 활용에 대해서는 엄격한 규제를 적용한다. AI Act는 회원국 국내법과 동일한 효력을 갖는 '규정(Regulation)'에 해당하므로, 회원국의 별도 조치 없이도 자동 적용된다. AI Act는 2024년 8월 발효되었으나, 항목별로 시행 시기가 달라 발효 후 2년간 단계적으로 시행된다. 핵심이라 할 수 있는 고위험 AI에 대한 규제는 2026년 8월부터 시행될 예정이었으나, 최근 시행 시기를 뒤로 미루는 방안에 대한 논의가 진행 중이다.⁹²⁾

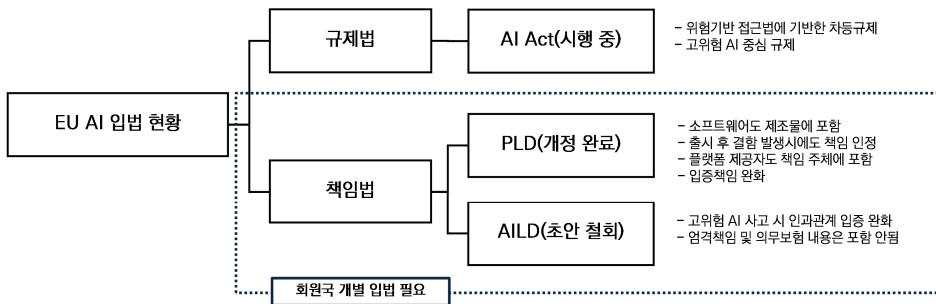
92) <https://www.europarl.europa.eu/news/en/press-room/20260323IPR38829/artificial-intelligence-act-delayed-application-ban-on-nudifier-apps> (2026. 4. 8. 최종 접속)

EU AI 책임법에는 개정 제조물책임지침(Product Liability Directive; PLD)과 AI 책임지침(AI Liability Directive; AILD)(초안)이 있다. 2024년 12월 발효된 개정 PLD는 AI 사고 발생 시 제조물책임 법리를 통해 대응할 수 있도록 기존 법제를 정비한 것이다. 개정 PLD의 핵심은 디지털 신기술이 적용된 제조물의 결함으로 사고가 발생한 경우 책임주체를 확대하고 소비자의 입증책임을 완화한 것이다. PLD는 회원국이 그 취지에 따라 국내법 개정 의무를 부담하는 '지침(Directive)'에 해당하며, 회원국들은 발효 후 2년이 되는 시점인 2026년 12월 9일까지 개정 PLD의 취지에 따라 자국의 제조물책임법을 개정할 의무를 부담한다.

2022년 9월 발표된 AILD 초안은 AI 사고에 대한 포괄적 책임법제로, AI에 의한 불법행위 상황 발생 시 책임 인정의 요건을 정하고 있다. 특정 위험원에 대한 별도의 책임법이라는 점에서 자배법, 원자력손해배상법 등과 유사하다. 법적 효력은 PLD와 마찬가지로 회원국 국내법 개정을 요하는 지침에 해당한다. 최근 EU 집행위원회는 AILD 초안을 공식적으로 철회하였으나, 관련 논의는 계속되고 있다.

이하에서는 EU AI Act, 개정 PLD 및 AILD 초안의 내용을 살펴보고, EU의 AI 관련 법제의 체계와 내용이 주는 시사점을 살펴보기로 한다.

(그림 IV-1) EU AI 관련 입법 현황



3. EU AI 규제법: AI Act

가. 개관

EU는 2018년경부터 AI 관련 법제 정비 논의를 시작하여, 먼저 2019년 연성규범인 윤리 가이드라인을 발표하였다. 이후 유럽 집행위원회는 AI 규제 방안을 보다 구체화한 AI 백서를 발표하고 의견 수렴 과정을 거쳐 2021년 AI Act 초안을 발표하였다. 초안에 대해 의회 및 회원국들의 논의를 거쳐 AI Act 최종안이 승인되었고 2024년 8월 1일 발효되었다. AI Act는 전문 180개 호와 113개 조항 및 13개 부속서로 구성되어 있다. 여기에서는 책임법제 논의에 필요한 범위 내에서만 개괄적으로 살펴보기로 한다.⁹³⁾

〈표 IV-2〉 AI Act의 구성

장	제목		조항	시행일
제1장	일반조항		제1조~제4조	2025년 2월
제2장	금지되는 AI		제5조	2025년 2월
제3장	고위험 AI	(1) 고위험 AI 시스템 분류	제6조~제7조	2026년 8월
		(2) 고위험 AI 시스템의 요건	제8조~제15조	2026년 8월
		(3) 고위험 AI 시스템 제공자, 배포자 기타 관련자들의 의무	제16조~제27조	2026년 8월
		(4) 관할 당국 등에 대한 통지	제28조~제39조	2025년 8월
		(5) 표준, 적합성 평가, 인증서, 등록	제40조~제49조	2026년 8월
제4장	특정 AI 시스템 제공자 및 배포자의 투명성 의무		제50조	2026년 8월
제5장	범용 AI 모델	(1) 분류 기준	제51조~제52조	2025년 8월
		(2) 범용 AI 모델 제공자의 의무	제53조~제54조	2025년 8월
		(3) 시스템 위험이 있는 범용 AI 모델 제공자의 의무	제55조~제56조	2025년 8월
제6장	혁신지원 조치		제57조~제63조	2026년 8월

93) EU AI Act의 내용 및 시사점에 대한 분석은 심소연(2024) 및 채은선(2024)을 참고 바람; EU AI Act 전문 번역은 국회도서관 버전과 한국법제연구원 버전이 있음

〈표 IV-2〉 계속

장	제목		조항	시행일
제7장	거버넌스	(1) EU 거버넌스	제64조~제69조	2025년 8월
		(2) 국가별 관할 당국	제70조	2025년 8월
제8장	고위험 AI 시스템에 대한 EU 데이터베이스		제71조	2026년 8월
제9장	판매 후 모니터링, 정보 공유 시장 관찰	(1) 판매 후 모니터링	제72조	2026년 8월
		(2) 중대한 사고에 대한 정보 공유	제73조	2026년 8월
		(3) 시행	제74조~제84조	2026년 8월
		(4) 구제 수단	제85조~제87조	2026년 8월
		(4) 범용 AI모델 제공업자에 대한 감독, 조사, 집행 및 모니터링	제88조~제94조	2026년 8월
제10장	자율 협약 및 지침		제95조~제96조	2026년 8월
제11장	권한 위임 및 위원회 절차		제97조~제98조	2026년 8월
제12장	벌칙		제99조~제101조	2025년 8월
제13장	최종 조항		제102조~제113조	2026년 8월
부속서	-		I ~ XIII	-

주: 고위험 AI 관련 규제 등 일부 항목은 시행 시기가 연기될 예정임
 자료: 심소연(2024), p. 2를 참고하여 일부 보완함

AI Act 전문은 AI 시스템의 개발, 시장 출시, 서비스 제공 및 사용에 대한 통일된 법적 프레임워크를 마련하여 유럽 연합 내부 시장의 기능을 개선하는 것이 법 제정의 목적임을 명시하고 있다. 또한 유럽 연합의 가치에 따라 인간 중심의 신뢰할 수 있는 AI의 도입을 촉진하면서도 건강, 안전, 유럽연합 기본권 헌장에 명시된 기본권(민주주의, 법치주의, 환경 보호 포함)에 대한 높은 수준의 보호를 보장하기 위한 것이라고 밝히고 있다.⁹⁴⁾

한편 AI Act는 AI 관련 주요 개념을 정의하고 있는바, 앞서 본 AI 시스템 외에도 범용 AI 시스템(General purpose system), 공급자(Provider), 운영자(Operator), 배치자(Deployer) 등 주요 개념을 정의하는 한편, 각 위험 유형별로 관련 이해관계자들의 의무를 정하고 있다.

94) AI Act Recitals (1)

〈표 IV-3〉 AI Act의 주요 개념

주요 개념	의의
AI 시스템 (AI System)	다양한 수준의 자율성을 가지고 작동하도록 설계된 기계 기반 시스템으로, 배포 후 적응성을 나타낼 수 있으며, 명시적·묵시적 목표를 달성하기 위해 입력된 정보로부터 예측, 콘텐츠, 추천 또는 결정을 생성할 수 있는 방법을 추론하는 것을 의미함
공급자 (Provider)	유상 또는 무상으로 AI 시스템을 개발하거나, 개발된 AI 시스템을 소유하면서 자신의 명의 또는 상표로 이를 시장에 출시하거나 서비스를 제공하는 자연인 또는 법인, 공공기관, 대행기관, 기타 법적 주체를 지칭함
배포자 (Deployer)	자신의 권한에 근거하여 AI 시스템을 사용하는 자연인 또는 법인, 공공기관, 대행기관, 기타 법적 주체를 총칭함. 단, 사적이고 비업무적인 활동 과정에서 사용되는 경우는 제외함(*기존 발표된 초안에서는 'user'라는 용어를 사용하였으나, 최종 채택된 법 규정에서는 'deployer'로 변경됨)
운영자 (Operator)	공급자, 제조업자, 배포자, 수권대리인, 유통업자, 수입업자를 총칭함

자료: AI Act Article 3

나. 위험 기반 접근법

1) 개관

EU AI Act의 핵심 내용은 AI를 위험도에 따라 구분하여 차등 규제를 적용한다는 것이다. EU AI Act는 AI의 위험수준을 금지되는 AI, 고위험 AI, 제한적 위험 AI, 최소 위험 AI로 구별하고, 고위험 AI에 대해서는 시스템 요건을 부과하고 제공자·배치자 등 이해관계자에게 의무를 부과하는 등 강한 규제를 적용하고 있다. 고위험 AI는 자동차, 의료기기와 같이 기존에도 강한 안전규제가 적용되는 분야에 AI가 적용되는 경우와 기본권 침해 가능성이 큰 영역에서 AI가 활용되는 경우로 구분된다. 전자의 경우는 기존 안전규제와 AI Act 규제가 중첩 적용되고, 후자의 경우에는 EU AI Act에 의해 비로소 규제 대상이 된다. 범용 AI 모델에 대해서는 시스템 위험이 있는 경우 모델 평가, 정보 기록, 보고 등의 추가적인 의무가 부과된다. 이상과 같은 EU AI Act의 내용은 다음에 살펴볼 PLD와 AILD의 전제가 된다. PLD와 AILD는 EU AI Act의 위험 기반 규제를 전제로 하고 있으며, 특히 고위험 AI에 대한 책임법제적 대응을 모색하는 것을 주 목적으로 하기 때문이다.

2) 위험도에 따른 AI 구분

가) 금지되는 AI

금지되는 AI는 말 그대로 개발이나 이용이 금지되는 AI를 의미한다. ① 인간의 잠재의식에 영향을 미치거나, 조작·기만을 위한 기술, ② 인간 행동 왜곡을 목적으로 개인의 취약성을 이용하는 기술, ③ 사회적 점수(Social scoring)를 통해 개인이나 단체를 평가하거나 분류하는 기술, ④ 범죄를 저지를 가능성을 평가하거나 예상하기 위한 기술(객관적 사실에 입각하여 평가하는 경우는 허용), ⑤ 얼굴 이미지를 수집하여 안면인식 데이터베이스를 생성·확장하는 기술, ⑥ 직장 및 교육기관에서의 감정추론 AI, ⑦ 민감정보(인종, 정치적 성향, 노조 가입, 종교, 신념, 성적지향) 추론을 위해 생체인식데이터를 기반으로 인간을 분류하는 시스템, ⑧ 법 집행 목적으로 공개된 장소에서 하는 실시간 원격생체식별시스템 이용(인신매매 등을 방지하기 위한 경우는 예외) 등이 이에 해당한다. 유형에 따라 예외적으로 허용되는 경우도 있어 완전히 금지된다고 보기는 어려우나 고위험 AI에 비해 기본권 침해 가능성이 현저히 높기 때문에 금지되는 AI로 분류된다.

나) 고위험 AI

EU AI Act의 핵심 규제 대상인 고위험 AI는 크게 두 그룹으로 구분된다. 첫째, 기존에 EU의 안전 관련 규제의 적용을 받는 위험군에 해당하는 제품에 AI가 장착된 경우이다. 기계, 장난감, 선박, 승강기, 폭발 관련 장비, 전파장비, 압력장비, 공중케이블, 개인보호장구, 가스기기, 의료기기, 항공기, 이륜차, 농기구, 해양장비, 철도, 자동차 등이 이에 해당하며, AI Act 부속서 I에 열거되어 있다. AI 탑재 여부와 관계없이 본래 위험성이 높다고 평가되어 엄격한 안전규제가 적용되는 영역이라고 볼 수 있다.

둘째, AI 이용 시 기본권 침해 가능성이 높은 것으로 예상되어 AI Act 부속서 III에서 고위험 AI로 분류한 경우이다. ① 원격 생체 식별, ② 핵심 기반 시설, ③ 교육 및 직업 훈련, ④ 고용, 근로자 관리, 자영업, ⑤ 필수 공공서비스, 민간서비스(신용도평가-금융사기 적발 제외, 생명 및 건강보험 관련 위험 평가), ⑥ 법 집행, ⑦ 이민, 망명, 국경통제, ⑧ 사법행정, 선거, 투표 등이 이에 해당한다. 다만 위 항목에 해당하더라도 의사결정 과정에 실질적으로 영향을 미치지 않고, 건강, 안전 및 기본권에 대해 중대한 위험을 야기하지 않는

경우에는 고위험 AI에서 제외될 수 있다.

고위험 AI로 분류되는 경우, AI Act에서 정한 시스템 요건을 충족하여야 하고, 공급자, 수입업자, 유통업자, 배포자에게 일정한 의무가 부과된다. AI 시스템 요건으로 위험관리체계 수립·이행, 데이터 거버넌스 구축 및 데이터 관리, 기술문서 작성, 로그 관리, 배포자에 대한 투명성 및 정보 제공 등이 요구된다. 이러한 요건의 준수는 공급자가 보장하여야 한다. 한편, AI 공급자는 품질관리 체계 마련, 문서 관리, 로그 보관, 시정조치 및 정보 제공, 관할 당국과의 협력, 적합성 평가 절차 이행, 적합성 보증에 서명하고 이를 제품에 표시할 의무를 부담한다. 수입업자와 유통업자는 시장 출시 전 적합성 심사 조치들의 이행을 확인할 의무를 부담한다. 배포자는 고위험 AI 시스템 사용에 필요한 기술적, 관리적 조치를 이행하여야 하고, 인간에게 AI를 감독하도록 해야 하며, 사업장 사용 전 근로자에게 고지하여야 한다. 공공기관, 공공서비스를 제공하는 민간단체 및 신용평가나 생명보험 위험평가를 실시하는 배포자의 경우 해당 시스템의 사용으로 인한 기본권에 대한 영향을 평가하여야 한다.

다) 제한적 위험 AI

제한적 위험 AI는 크게 4가지 경우로 구분되며, 투명성 의무가 부과된다. ① 자연인과 직접 상호작용하기 위한 AI 시스템(챗봇)의 경우 AI임을 표시해야 한다. ② 범용 AI 포함 생성형 AI는 그 결과물이 AI에 의해 생성됨을 표기해야 한다. ③ 감정인식 시스템, 생체인식 분류 시스템의 경우, 해당 시스템이 운영되고 있다는 사실을 상대방에게 알려야 한다. ④ 딥페이크 생성 AI 시스템의 경우, AI 배치자는 라벨링 등을 통해 딥페이크 생성물이라는 사실을 알려야 한다.

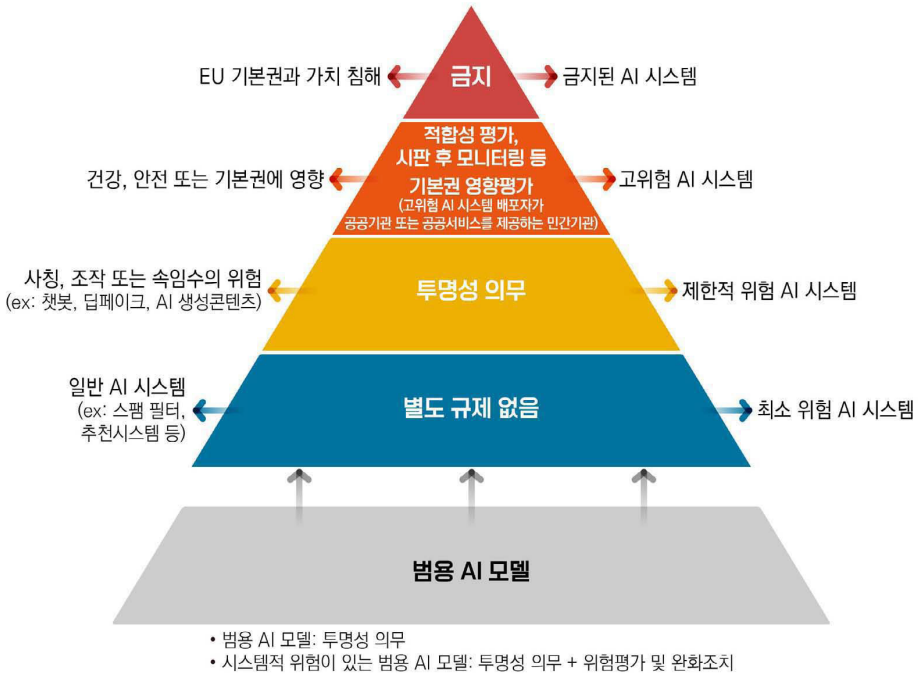
라) 최소 위험 AI

이상의 유형에 해당하지 않는 AI 시스템은 원칙적으로 별도 규제 없이 자유롭게 개발하고 활용할 수 있다. 다만, EU는 이러한 영역에 대해서도 자율규제 마련을 장려하고 있다.

마) 범용 AI 모델

AI Act는 범용 AI 모델의 경우 공급자에게 추가적인 의무를 부과한다. 범용 AI 모델 제공자는 기술문서 작성 의무, 범용 AI를 자신의 AI에 통합하고자 하는 공급자를 위한 정보 제공 및 문서 작성 의무, 저작권법 준수 정책 마련 및 AI 학습용 콘텐츠 관련 정보 공유 의무 등을 부담한다. 한편, 시스템 리스크를 야기할 수 있는 범용 AI 모델의 제공자는 구조적 위험을 파악하고 경감하는 관점에서 모델을 평가하여야 하고, 해당 모델에 의해 발생 가능한 시스템적 위험을 평가하고 감경시켜야 하며, 중대한 사고 정보 및 이에 관한 시정조치를 기록 및 보고하여야 하고, 보다 높은 사이버 보안 의무를 준수해야 한다.

〈그림 IV-2〉 AI 시스템 위험도에 따른 차등 규제



자료: 심소연(2024), p. 3

4. EU AI 책임법: PLD 및 AILD

가. 개관

AI 책임법 관련 논의는 크게 AI 사고에 대해 기존의 책임법제로 대응할 것인지, 아니면 AI 사고에 대한 별도의 책임법제를 마련할 것인지로 나누어진다. 초기에는 AI 사고의 특수성 및 기존 책임법제의 한계에 주목하여 AI 사고에 대한 별도 책임법제 마련 필요성이 강조되었다. AI 자체에 법인격을 부여하는 방안이나 AI 사고에 대해 특정 이해관계자에게 무과실책임을 부과하는 논의가 이에 해당한다. 그러나 논의가 진행될수록 AI 사고에 대한 책임법제를 도입할 경우 발생할 수 있는 각종 부작용이 드러났고, AI 기술 발전 단계를 고려할 때에도 별도 책임법제 도입은 시기상조라는 견해가 대두되었다.

이러한 논의를 반영하여 EU는 2025년 기존 책임법제를 일부 수정·보완함으로써 AI 사고 책임 문제를 해결하는 방향으로 입법을 완료하였다. AI 사고의 책임주체 불명확성 및 입증 곤란성을 고려하여 기존 책임법제인 PLD를 개정하여 책임주체를 확대하고 입증책임을 완화하였다. 반면, 별도 책임법제에 해당하는 AILD는 논의 초기에 제안되었던 것에 비해 상당히 축소된 내용으로 초안이 마련되었고 그마저도 관철되지 못하고 최근 철회되었다. 이하에서는 AI 책임법제에 관한 EU의 논의 경과를 살펴보고, 개정 PLD의 내용 및 AILD 초안의 내용을 정리해 보기로 한다.

나. 논의 경과

1) 2019년 EU 집행위원회 산하 전문가그룹 EG-NTF 보고서 발표

가) 개관

EU의 AI 책임법 논의의 기초가 된 것은 집행위원회 산하 전문가 그룹인 'Expert Group on Liability and New Technologies-New Technologies Formation(EG-NTF)⁹⁵⁾가 2019년 발

95) EG-NTF는 EU 집행위원회가 2018년 설립한 독립 전문가 그룹으로, 인공지능(AI), 사물인터넷(IoT), 분산원장기술 등 신기술로 인한 법적 책임 문제를 연구하고 기존 법체계의 적합성을 검토하여 개선 방향을 제안하는 것을 목적으로 함. 집행위원회는 'Expert Group on liability and new technologies'라는 전문가 그룹을 구성하여 동 그룹 내에 'Product Liability Directive(PLD) Formation'과 'New Technologies Formation'이라는 두 개

간한 'Liability for Artificial Intelligence and Other Emerging Digital Technologies'이다. EG-NTF(2019)는 AI를 비롯한 신형 디지털 기술이 야기하는 새로운 위험과 피해에 대해 기존의 법률 체제로 대응하는 데 한계가 있는지를 점검하고, 원활한 피해 보상, 공정한 책임 배분 및 사고 예방을 위해 총 18의 쟁점에 대해 34개의 핵심 내용(Key-findings)을 제시하였다. AI 사고에 대해 운영자에게 엄격책임을 부과하는 방안, 제조물책임 확대 방안, 입증 책임 전환 및 완화 방안, 보험가입 의무화 방안 등이 핵심 내용이다.

나) 기존 책임법제의 한계 및 AI 책임법제의 기본 전제

EG-NTF(2019)가 AI 사고 관련 기존 책임법제의 한계로 지적한 사항은 크게 3가지이다. 첫째, 기존 책임법제로는 AI 사고로 인한 손실을 공정하고 효율적으로 분배하기가 어렵다는 점이다. 일반적으로 어떠한 위험원에 의해 사고가 발생한 경우 사고 유발 행위를 한 자, 위험원로부터 이익을 얻는 자, 위험원을 관리하는 자, 최소 비용으로 위험을 회피할 수 있는 자, 최소 비용으로 보증을 취득할 수 있는 자 등을 책임주체로 고려해 볼 수 있다. 책임법은 이들에게 손실을 적절히 배분하는 기능을 해야 한다. 그런데 AI 사고는 책임주체가 불명확한바, 기존 책임법제를 그대로 적용할 경우 손실 배분에 불공정이나 비효율이 발생할 수 있다. 둘째, 개인의 권리에 발생하는 위험에 대해 법체계가 일관성 있는 적절한 대응을 하지 못할 수 있다. 예컨대, 피해의 종류나 크기가 동일함에도 그 원인이 인간의 행위나 기존 기술에 의한 것인지 아니면 AI 등 신기술에 의한 것인지에 따라 권리 구제의 절차 및 결과에 차이가 발생할 수 있다. 셋째, 효과적인 사법 접근성을 저해할 수 있다. 특히 AI 사고의 경우 피해자가 소송을 제기하는 것이 현저히 곤란하거나 과도한 비용이 발생하는 경우가 있을 수 있다.⁹⁶⁾ EG-NTF(2019)가 제시하고 있는 3가지 기존 책임법제의 한계점은 앞서 본 AI 사고 관련 책임법제의 책임주체 및 입증책임 문제와 관련된다.

이러한 한계점에 대한 고려는 AI 사고의 특성을 고려한 책임법제 정비 필요성에 대한 검토로 연결된다. AI 책임법제 정비는 기존 책임법제의 수정·보완과 새로운 책임법제 마련을 포괄하는 개념이다. EG-NTF(2019)는 AI 사고 책임법제 정비의 기본 전제(Bases of liability)로 3가지를 제시한다. 첫째, 유사한 위험에 대해서는 동일한 책임 체계가 적용되어

의 소그룹을 두었으며, 그중 New Technologies Formation(NTF)은 신기술 관련 책임법의 원칙 및 지침 개발을 담당하였음. 현재는 활동이 종료되었음

96) EG-NTF(2019), p. 34

야 한다. 이는 책임의 근거뿐 아니라 손해의 유형 및 보상 범위를 정할 때도 마찬가지이다. 둘째, 엄격책임과 과실책임이 함께 적용되어야 한다. 두 가지 책임이 모두 성립되는 경우 피해자는 각 책임자 모두에게 손해배상을 청구할 수 있어야 한다. 이때 복수의 책임자들은 공동불법행위자(Multiple tortfeasors)로 본다. 셋째, 특정 영역의 사고에 대해서는 불법행위책임과는 별도로 계약상 책임(Contractual liability)이나 다른 형태의 보상 체계⁹⁷⁾가 적용되며,⁹⁸⁾ 기존 책임법제 수정·보완 시 이러한 여타 보상 체계를 고려하여야 한다.

요컨대 AI 책임법제는 책임 배분의 공정성과 효율성, 피해 구제의 일관성과 적절성, 사법 시스템의 접근성과 실효성 관점에서 기존 책임법제가 AI 사고 대응 수단으로 적합한지에 대한 평가를 전제로, 동일한 위험에 대해 동일한 책임을 적용하고, 엄격책임과 과실 책임을 병용하며, 계약책임 및 기타 보상 체계와 불법행위책임의 역할 배분을 고려하여 마련되어야 한다고 볼 수 있다.

다) 엄격책임(운영자 엄격책임 신규 도입, 제조업자 엄격책임 강화)

EG-NTF(2019)가 가장 우선적으로 검토한 책임법 이슈는 엄격책임(Strict liability)이다. AI 사고의 책임주체의 불명확성과 사고 원인 규명 곤란성을 고려할 때, 위험원과 특별한 관계를 맺고 있는 특정인에게 고의·과실 여부를 불문하고 책임을 귀속시키는 것이 적절할 수 있기 때문이다. AI 책임 논의 초기에는 AI 자체에 법인격을 부여하는 방안이 검토되기도 하였으나, 적어도 책임법의 한계 때문에 AI에게 법인격을 부여할 필요는 없다는 것이 현재의 중론이다.

EG-NTF(2019)는 운영자와 제조업자에게 엄격 책임을 부과할 필요가 있다고 제안한다. 먼저 AI 운영자에 대한 엄격책임 부과 방안을 살펴보자. AI 운영자(Operator)는 AI의 운영과 관련된 위험을 통제하고 그 운영으로부터 이익을 얻는 자를 의미한다.⁹⁹⁾ 우리나라 자배법

97) 불법행위책임 외에 다른 형태의 책임을 의미하며, 계약상 책임, 민법상 하자담보책임 등이 여기에 해당할 수 있음

98) 예컨대, 오스트리아의 경우 제조물 결함에 의한 사고의 피해자가 제조업자에게 직접 책임을 묻는 방법으로 엄격책임인 제조물책임과 함께 제3자에 대한 관계에서 보호 의무를 부과하는 계약(the concept of a contract with protective duties in relation to third parties)(주: 제조업자에게 해당 제조물을 구매한 소비자를 제3자의 손해배상청구로부터 보호해야 할 의무를 부과하는 계약으로 이해됨)도 인정되고 있음(EG-NTF(2019), p. 37, 각주 96)

99) EG-NTF(2019), p. 39; EG-NTF(2019)는 Operator를 'the person who is in control of the risk connected with the operation of emerging digital technologies and who benefits from their operation'이라고 정의하고 있음

상 운전자 개념과 유사하다. 자배법상 운행자는 자동차 운행에 대한 위험을 지배하고 그 운행으로부터 이익을 얻는 자를 의미한다. 위험원에 대한 지배와 이익을 근거로 위험원이 야기한 사고에 대한 책임을 귀속시키는 것이다. 이는 불법행위법 영역에서는 위험책임, 보상책임 등으로 불리며, 과실책임 원칙의 예외로 인정되어 온 것이다. 이를 AI 사고에 대해서도 적용하여, AI 운영 관련 위험을 통제하고 그 운행으로부터 이익을 얻는 자를 AI의 운영자로 지정하여 사고 발생 시 그 운영자가 과실이 있는지 여부를 불문하고 책임을 귀속시키는 것이다. EG-NTF(2019)는 공공장소에서 활용되고 심각한 피해를 유발할 수 있는 경우 운영자에게 엄격책임을 부과할 것을 제안한다. 만약 복수의 운영자가 있는 경우에는 관련 기술의 활용을 주로 결정하고 이득을 얻는 프런트엔드 운영자(Front-end operator), 관련 기술의 특성을 지속적으로 정의하면서 필수적이고 지속적인 백엔드 지원을 하는 자(Back-end operator) 중 운영 위험에 대해 더 많은 통제권을 갖는 자에게 엄격책임을 부과해야 한다고 보고 있다. 통상적인 경우 프런트엔드 운영자가 더 많은 통제권을 갖지만, 그렇지 않은 경우도 있을 수 있으며, 입법자가 엄격책임을 지는 운영자가 누구인지를 구체적으로 정하는 것이 바람직하다고 본다.

한편 제조업자의 경우 기존 제조물책임법에 따라 엄격책임을 부담하고 있는바, 이러한 제조업자의 엄격책임은 AI 사고 피해 구제에도 핵심적인 역할을 할 것으로 기대된다. AI는 적응성을 가지므로 AI 제조업자는 제품 유통 이후 발생한 결함에 대해서도 제조물책임을 부담하며, AI 제조업자가 업데이트 및 업그레이드 의무를 부담하는 경우에는 제품 공급 시점을 기준으로 한 ‘개발 위험의 항변’은 적용되지 않는다. AI 기술의 특성으로 인해 제조물책임 성립 요건 입증에 어려운 경우에는 그 입증책임은 전환하거나 완화되어야 한다.

라) 과실책임(운영자 및 제조업자의 주의의무)

EG-NTF(2019)는 엄격책임 도입 여부와 별개로 일반적·포괄적 책임법제인 과실책임을 AI 사고에 어떻게 적용할 것인지에 대한 정비도 필요하다고 보았다. 과실책임 적용을 위해서는 먼저 과실의 근거가 되는 주의의무를 확정할 필요가 있는바, EG-NTF(2019)는 AI 운영자 및 제조업자의 주의의무의 내용을 제안한다.

운영자는 적합한 시스템 선택, 시스템 모니터링 및 유지관리 관련 주의의무를 부담한다. 이때 주의의무의 구체적인 내용은 규제법에서 정해질 수도 있고, 법원 판례에 의해 구체화될 수도 있다. 예컨대 EU AI Act에서 고위험 AI 개발 및 활용 시 준수해야 할 사항을 정

한 경우, 이를 위반하였다면 주의의무 위반에 해당하는 것으로 볼 수 있다. 한편 제조업자는 운영자가 시스템의 선택, 모니터링, 유지관리를 할 수 있도록 제품을 설계, 설명, 판매해야 하고, 제품을 유통시킨 후에도 적절히 모니터링해야 할 주의의무를 부담한다. 제조업자가 직접 운영하는 경우에는 두 가지 의무를 중첩적으로 부담하게 된다.

마) AI 감독자 책임(사용자책임과 동등한 수준의 자동화기술 운영자책임 부과)

타인을 사용하여 사무에 종사하게 한 자는 그 피용자가 사무집행에 관하여 제3자에게 가한 손해를 배상할 책임이 있다(민법 제756조). 이를 사용자책임이라 하며, 감독자 책임의 일종이다. AI 기술 발전으로 사람을 이용하여 사무에 종사하게 하는 대신 AI를 동일한 사무에 종사하도록 할 수 있다. AI 에이전트가 비서 업무를 대체하고 자율주행시스템이 운전자를 대신하게 되는 경우가 이에 해당한다. EG-NTF(2019)는 인간 보조자를 고용하는 것과 기능적으로 동등한 방식으로 사용된 자동화 기술로 인해 손해가 발생한 경우 운영자에게 부과되는 책임은 인간 보조자 활용에 대해 적용되는 감독자 책임 체계와 일치해야 한다고 제안하고 있다. 사람 대신 AI를 활용하다가 사고가 발생한 경우에는 사람을 사용했을 때 인정되는 것에 상응하는 책임이 인정되어야 한다는 것이다. 또한 감독자의 책임과 관련하여 자동화 기술 작동의 적정성을 평가할 때에는 인간 보조자의 행위를 평가할 때와 같은 기준을 적용해야 한다고 제안하고 있다. 즉, 기존에 인간이 수행하던 업무를 AI가 대체한 경우 AI 업무 수행 적절성 여부는 같은 조건에서 인간이 업무를 수행했을 때와 같은 기준으로 평가해야 한다는 것이다.

이러한 제안은 AI에 의한 자동화 상황에서 발생할 수 있는 책임 문제를 해결하는데 유용한 기준이 된다. 자율주행차의 경우를 예로 들어 보자. 자동차 소유자나 운수업체 운영자가 인간 운전기사를 고용하여 운전하게 하는 경우 자동차 소유자나 운수업체 운영자는 자배법상 운행자책임 및 민법상 사용자책임을 부담한다. EG-NTF(2019)의 제안에 따르면 자동차 소유자나 운수업체 운영자는 인간 운전기사 대신 자율주행시스템을 이용하여 운전한 경우에도 동일한 책임을 부담해야 한다. 한편, 자율주행시스템에 의한 운전의 적정성을 평가할 때에는 인간 운전자의 운전 행위를 평가할 때와 같은 기준을 적용해야 한다. 인간 운전자가 실수할 수 있는 상황이라면 자율주행시스템도 실수할 수 있고, 인간 운전자가 피할 수 없는 상황이라면 AI도 피할 수 없다고 보아야 하는 것이다.

이러한 제안에 따라 AI에 의해 자동화된 영역에 대해 기존의 사용자책임 체계를 적용하고,

AI 작동의 적정성 여부를 같은 상황에서 인간의 행위에 대한 적정성과 동일한 기준으로 평가한다면 기존 책임법제 적용으로도 책임 공백을 상당 부분 해소할 수 있을 것으로 보인다.

바) 입증책임

입증책임 문제는 AI 사고 피해 구제의 핵심 쟁점이 될 가능성이 높다. AI 운영자나 제조업자에게 엄격책임이 부과되더라도 이는 고의·과실에 대한 입증을 요구하지 않는 것일 뿐 그 밖의 요건에 대해서는 여전히 사고 피해자가 입증책임을 부담하기 때문이다. AI 운영자나 제조자 역시 종국적으로는 사고 원인을 규명하여 책임 비율에 따라 책임을 최종 배분해야 하므로, 사고 원인을 규명할 필요가 있다. EG-NTF(2019)는 입증책임 문제를 해결하기 위해 ‘로깅 바이 디자인’, 안전기준 위반 시 고의·과실 및 인과관계 인정, 고의·과실 및 인과관계 입증책임 완화 등을 제안하고 있다.

‘로깅 바이 디자인(Logging by design)’은 설계 시 AI 운영 관련 정보를 생성·기록하는 수단을 장착할 의무를 부과하는 것이다. 위험 현실화 여부를 판단하는 데 필수적인 정보를 생성·기록하도록 하되, 로깅의 기술적 실행 가능성, 비용, 대체 수단 활용 가능성, AI가 야기하는 위험의 유형 및 규모, 타인의 권리에 미칠 수 있는 부정적인 영향 등을 종합적으로 고려하여 의무 부과 범위를 정해야 하며, 특히 개인정보나 영업비밀 보호를 침해하지 않도록 해야 한다. 만약 제조업자가 이러한 로깅 바이 디자인 의무를 위반하여 정보가 생성·기록되지 않았다면 해당 정보를 통해 입증하고자 하는 요건사실은 입증된 것으로 보게 되고, 이로 인해 AI 운영자가 손해배상책임을 부담하게 되는 경우 제조업자에게 구상을 청구할 수 있다.

AI 관련 ‘로깅 바이 디자인’의 가장 대표적인 예는 자율주행차에 장착되는 정보기록장치이다. 자율주행차 제작사는 자율주행차에 자율주행정보기록장치를 장착할 의무를 부담한다(자배법 제39조의17 제1항). 자율주행차가 사고가 발생하였는데 그것이 자율주행시스템 오작동 때문인지 여부가 문제 될 경우 자율주행정보기록장치에 기록된 정보를 토대로 이를 밝힐 수 있을 것이다. 만약 제작사가 자율주행정보기록장치를 장착하지 않거나 장치 내에 필요한 정보가 기록되지 않았다면 자율주행시스템 오작동이 존재하는 것으로 인정되고, 이로 인해 자율주행차 운행자가 책임을 부담하는 경우 운행자는 제작사에게 구상을 청구할 수 있다. 이처럼 로깅 바이 디자인 개념은 AI 관련 사고 원인 규명 수단으로 활용될 수 있을 것이다.

입증책임 문제를 해결하기 위한 또 다른 방안은 바로 안전기준 위반 사실이 인정되면 곧바로 고의·과실, 결함, 인과관계의 존재를 추정하는 것이다. AI에 대해서는 활용 영역별로 안전기준이 마련되어 있다. 안전기준이란 AI 사고를 방지하고 안전성을 확보하기 위해 마련된 각종 규범력 있는 기준들을 의미한다. 자율주행차에 대해서는 자율주행차 안전기준이, 금융 AI에 대해서는 금융 AI 가이드라인이 마련되어 있으며, EU AI Act는 고위험 AI에 대한 포괄적 규제를 마련하고 있다. 이러한 개별적, 포괄적 안전규제 위반 여부는 고의·과실, 인과관계, 결함 등에 비해 입증하기가 용이하다는 점에서 피해자의 입증 부담을 크게 덜어주게 된다.

또한 기술의 특성을 고려하여 인과관계 및 고의·과실에 대한 입증책임을 완화할 것을 제안한다. 입증책임을 완화는 입법을 통해서도 가능하지만 법원의 판결을 통해서도 가능하다. AI 기술과 손해 발생 사이의 관련성, 정보 접근 가능성, 피해의 종류 등을 종합적으로 고려하여 입증책임을 균형을 맞출 필요가 있는 경우 인과관계에 대한 입증책임을 완화할 수 있다고 보고 있다. 또한 AI 관련 사고 책임 중 과실책임이 적용되는 영역에 대해서는 주의의무의 수준을 확립해야 하고 그 입증이 지나치게 어렵거나 고비용인 경우에는 입증책임을 전환해야 한다고 본다.

끝으로 복수의 공급자가 계약이나 기타 근거에 의해 상업적·기술적 단위상 서로 구분되는 요소를 제공하였을 때, 여러 개의 상업적·기술적 단위 중 어느 요소가 피해를 유발했는지를 입증하기 어려운 경우에는 잠재적 불법행위자들이 공동불법행위책임을 부담하는 방안을 제안하고 있다. AI 기반 제품이나 서비스는 여러 기술적 수단이 복합적으로 작용할 수 있는바, 해당 제품 및 서비스 공급에 관여한 주체들이 모두 연대하여 책임을 부담하도록 함으로써 피해자의 입증 부담을 줄여주는 효과를 얻을 수 있다.

사) 보험 및 보상기금

EG-NTF(2019)는 보험 및 보상기금 관련 방안도 제안하고 있다. 먼저 보험에 관해서는 AI 기술로 인해 발생하는 피해의 빈도나 심도가 높고, 운영자가 개별 피해자에게 배상하는 것이 어려울수록 의무보험 도입의 필요성이 크다고 보고 있다. 통상 의무보험은 엄격책임과 연계되는 경우가 많다. 자동차보험이 가장 대표적인 예이다. 이러한 방식을 AI 사고에 적용할 경우 AI 운영자에게 엄격책임과 그 책임의 이행을 담보하기 위한 보험에 가입할 의무를 함께 부과하게 된다.

다만 EG-NTF(2019)는 의무보험제도 도입을 위해서는 AI 운영자의 배상자력을 담보하기 위해 의무보험이 필요한지에 대한 신중한 분석이 필요하다고 하였으며, 분야별로 의무보험제도 필요 여부가 달리 판단될 수 있다고 보고 있다.

한편 보상기금은 책임 체계에 의한 피해자 구제가 이루어지기 어려운 경우 보충적인 역할을 해야 한다고 보고 있다. 자동차 사고의 경우 원칙적으로 자동차보험에서 피해자 구제가 이루어지고, 뺑소니, 무보험차와 같이 자동차보험으로 보상이 어려운 경우 정부보장사업에 의해 피해자 구제가 이루어지는 것과 마찬가지로이다. 한편, 해킹 등 사이버범죄에 의한 피해는 가해자를 특정하여 그에게 책임을 묻는 것이 어렵기 때문에 노폴트 보상 제도를 적용하는 것이 적합하다고 보고 있다.

아) 요약

이상과 같이 EG-NTF(2019) 보고서는 AI 사고 책임과 관련한 기존 책임법제의 한계, AI 책임법제의 기본 전제, AI 책임법제 마련 시 고려해야 할 주요 쟁점 및 각 쟁점에 대한 합리적인 해결 방안 등을 제안하고 있다. EU AI 책임법에 관한 이후의 논의는 이 보고서에서 다루어진 쟁점들과 제안들을 토대로 이루어지고 있다고 볼 수 있다.

〈표 IV-4〉 EG-NTF(2019)의 책임법제 관련 쟁점 사항

항목	쟁점	핵심 사항
1	책임법에 대한 신기술의 도전	<ul style="list-style-type: none"> AI·디지털 기술의 복잡성, 불투명성, 개방성, 자율성, 낮은 예측가능성, 데이터 주도성, 취약성은 책임법제에 영향을 미침 이러한 영향은 혼란을 초래할 우려가 있음
2	도전에 대한 평가 및 조정 필요성	<ul style="list-style-type: none"> 기존 책임법은 기본적 해결책을 제공하나, 손해 부담의 불공정, 피해자 보호 미흡, 사법 접근성 저하 문제가 발생 가능함 기존 책임법제에 대한 조정 및 법 개정 필요성에 대한 검토가 필요하며, 다양한 위험에 대해 단일한 해결책을 제시하는 것은 불가능함
3	책임의 기본 전제	<ul style="list-style-type: none"> 유사한 위험에 대해서는 유사한 책임 체계가 적용되어야 함 엄격책임과 과실책임은 병존되어야 하고, 피해자는 복수의 책임자에게 배상을 청구할 수 있어야 함 일부 디지털 생태계에서는 계약책임 및 다른 보상시스템이 불법행위책임과 함께 또는 이를 대체하여 적용될 수 있음
4	AI 법인격	<ul style="list-style-type: none"> 책임 귀속을 위해 시에 법인격을 부여할 필요는 없음

〈표 IV-4〉 계속

항목	쟁점	핵심 사항
5	운영자의 엄격책임	<ul style="list-style-type: none"> • 디지털 신기술이 공공장소에서 운영되고 심각한 피해를 야기할 수 있는 경우 엄격책임을 적용하는 것이 적절함 • 엄격책임은 위험을 통제하고 운영으로부터 이익을 얻는 운영자(Operator)에게 귀속되어야 함 • 프런트엔드 운영자(Front-end operator)와 백엔드 운영자(Back-end operator)가 있는 경우 위험에 대해 더 많은 통제 권한을 가지는 자에게 엄격책임을 부과해야 함 • 인간에 의한 통제라는 전통적 개념에 기반한 항변사유나 면책사유는 재고될 필요가 있음
6	제조물책임	<ul style="list-style-type: none"> • 물리적 실체가 있는지 여부를 불문하고 제조물책임이 인정됨 • 유통 후 결함이 확인된 경우에도 제조업자가 기술의 업데이트나 업그레이드 권한을 가지는 한 제조물책임이 인정되며 개발 위험의 항변은 적용되지 않음 • 안전기준 준수 관련 입증에 현저히 곤란하거나 과도한 비용이 발생하는 경우 결함에 대한 입증책임은 전환되어야 함
7	과실책임 및 주의의무	<ul style="list-style-type: none"> • 운영자는 적합한 시스템의 선택, 시스템 모니터링, 시스템 유지관리에 관한 주의의무를 부담함 • 제조업자는 운영자가 위의 의무를 준수할 수 있도록 제품을 효과적으로 설계·설명·판매해야 하고, 유통 후 모니터링해야 함
8	감독자 책임	<ul style="list-style-type: none"> • 인간 보조자를 고용하는 것과 기능적으로 동등한 방식으로 사용된 자동화 기술로 피해가 발생한 경우 해당 기술에 대한 운영자책임은 인간 보조자 활용에 적용되는 감독자 책임 체계와 일치해야 함 • 자동화 기술의 작동을 평가할 때 인간 보조자의 행위와 동일하게 평가해야 하며, 자동화 기술이 인간 보조자를 능가하는 경우에는 유사한 기술의 성능을 기준으로 작동의 적절성을 평가해야 함
9	로깅 바이 디자인	<ul style="list-style-type: none"> • 제조업자는 자동화 기술의 운영 관련 정보를 생성·기록하는 수단을 장차해야 함 • 데이터 및 영업비밀 보호 관련 규정을 준수하여 정보를 기록해야 함 • 정보가 로깅되지 않았거나 피해자의 정보접근권을 불합리하게 제한하는 경우 해당 정보로 입증하고자 하는 요건사실은 존재하는 것으로 추정함 • 위 추정 결과 운영자가 배상책임을 부담할 경우 제조업자에게 구상을 청구할 수 있음
10	안전기준	<ul style="list-style-type: none"> • 안전기준 위반 시 인과관계, 과실, 결함의 존재 등이 입증된 것으로 봄
11	인과관계 입증책임	<ul style="list-style-type: none"> • 원칙적으로 피해자가 손해 원인을 입증해야 함 • 기술이 손해 발생에 미치는 영향, 다른 원인에 의해 발생했을 가능성, 기술 관련 알려진 결함, 정보 비대칭, 기록된 정보에 대한 접근성 및 이해 가능성, 피해의 종류 및 정도를 고려하여 필요시 인과관계 입증 부담을 완화할 수 있음

〈표 IV-4〉 계속

항목	쟁점	핵심 사항
12	과실 입증책임	• 디지털 신기술 관련 주의의무의 수준을 명확히 하고, 과실의 입증이 지나치게 어렵거나 입증에 과도한 비용이 드는 경우 입증책임을 전환해야 함
13	피해자 측 과실	• 피해자 측 과실은 손해배상액 감액 사유가 됨
14	상업 및 기술 단위	• 서로 구분되는 상업 및 기술 단위를 공급하는 복수의 공급자가 존재하고, 이들이 공급한 요소 중 어느 하나가 피해를 야기했다는 점은 인정되지만 구체적으로 어떤 요소가 피해의 원인인지가 불분명한 경우, 모든 공급자가 연대하여 책임을 부담해야 함 • 상업 및 기술 단위는 공동마케팅 여부, 각 요소의 기술적 상호 의존성 및 상호작용 정도, 독점 여부 등을 고려하여 판단해야 함
15	공동불법행위자 간 책임 배분	• 공동불법행위자들은 연대책임을 부담하되, 과실비율이 확인되는 경우 그에 따라 구상 청구가 가능함
16	데이터 관련 손해	• 데이터에 대한 침해는 계약상 책임이 있는 경우, 다른 재산권에 대한 침해가 발생하는 경우, 해당 침해 행위가 형사 범죄행위에 해당하거나 관련 법규 위반행위에 해당하는 경우, 그 손해에 대한 책임이 인정됨
17	보험	• 사고의 빈도나 심도가 높고, 책임자가 피해자에게 개별적으로 배상을 할 가능성이 낮은 경우 의무보험이 적합할 수 있음
18	보상기금	• 책임법과 보험에 따라 피해 보상이 이루어지지 않는 경우 보충적인 피해자 보호 수단으로 활용될 수 있음

자료: EG-NTF(2019), pp. 32~63의 내용을 요약하여 정리함. 번역은 황현아(2024a), pp. 136~142에 따름

2) 2020년 EU 집행위원회 보고서 발표 및 EU 의회 결의

가) EU 집행위원회 보고서

2020년 2월 EU 집행위원회는 AI 백서를 발간하면서, 그 부속 보고서로 AI 책임법제에 관한 보고서인 “Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics”(이하 ‘EC(2020)’라 함)를 EU 의회에 제출하였다. 동 보고서는 AI, 사물인터넷(IoT), 로봇틱스 등 신기술이 가져올 안전(Safety)과 책임(Liability) 관련 이슈를 제시한다.

EC(2020)는 EU의 책임법제를 EU 회원국에 공통적으로 적용되는 책임법제인 제조물책임과 회원국에서 개별적으로 정하는 책임법제인 불법행위책임 및 기타 엄격책임으로 구분한다. EU 회원국 내에서 발생한 사고에 대해서는 공통 법제인 제조물책임과 개별 법제인 불법행위책임이 상호보완적으로 작용하며 피해자를 구제하고 있다. AI 사고에 대해서도

이러한 이중의 책임법제가 적용된다.

그런데 AI, IoT, 로봇릭스 등 관련 사고의 경우 연결성(Connectivity), 자율성(Autonomy), 데이터 의존성(Data dependency), 학습 능력(Self-learning), 복잡한 공급망, 사후 업데이트 가능성, 알고리즘의 불투명성(Opacity), 사이버 위협 등 기존의 사고와는 구별되는 특성이 있고, 그 결과 기존의 책임법제로는 대응하기 어려운 측면이 있다. 이에 EC(2020)는 안전 측면에서는 설계 단계에서부터 안전 및 보안을 고려하도록 하는 Safety by design, Security by design을 제안하는 한편, 책임과 관련하여서는 책임주체 확대, 엄격책임 도입, 입증책임 전환 등을 고려할 필요가 있다고 제안한다.

책임법제 관련 제안 사항을 구체적으로 살펴보면, 제조물책임의 경우 적용 범위에 소프트웨어 등을 포함하고 '시장 출시' 개념을 AI 특성에 맞게 개정하는 방안을 제안한다. 국가별 불법행위책임의 경우 AI 사고에 적용 시 책임의 요건을 입증하기 어렵거나 입증에 지나치게 과다한 비용이 소요될 우려가 있으므로, 안전기준 위반이 확인되는 경우 고의·과실 및 인과관계의 존재를 추정하는 방식을 고려할 필요가 있다고 제안한다. 한편 불특정 다수의 생명·신체에 피해를 끼칠 수 있는 특정 분야의 고위험 AI에 대해서는 엄격책임 부과 및 필요성 및 보험가입의 필요성에 대해 적극적인 검토가 필요하다고 제안하고 있다.

EC(2020)가 AI 사고 책임법제로 엄격책임 및 의무보험 도입과 같은 적극적인 입법적 조치를 제안한 것은 AI 사고의 피해자를 기존의 다른 사고 피해자와 동등하게 보호해야 한다는 원칙에 입각한 것이다. AI 사고는 기존의 다른 사고에 비해 책임주체가 불분명하고 고의·과실 및 인과관계 입증도 어려울 것으로 예상된다. 그러나 이러한 사고의 특성으로 인해 AI 사고 피해자가 손해배상에서 불이익을 받게 되는 것은 그 자체로 타당하지 않다. 또한 AI 기술에 대한 사회적 수용성 저하 및 AI 기술 보급 장애를 야기할 수도 있다. 한편 책임법제를 명확히 하는 것은 AI 개발자 및 사업자에게도 도움이 된다. AI 사고에 대한 책임의 요건 및 범위가 불분명할 경우 불확실성이 높아져 투자를 저해하고 정보 비용 및 보험 비용의 증가를 야기할 수 있기 때문이다. 이러한 점에서 EC(2020)는 AI 사고에 대한 합리적인 피해 구제 체계를 마련하는 것이 피해자 보호뿐 아니라 기술 및 산업 발전, 나아가 AI 생태계 구축에도 기여할 수 있다고 보았고, 이를 위해 입증책임 전환, 엄격책임 부과 및 의무보험 도입 등의 방안이 필요하다고 본 것이다.

나) EU 의회 결의안

EU 의회는 2020년 10월 AI에 대한 책임 체계를 정립하기 위한 입법 결의안 EP(2020)를 채택하였다.¹⁰⁰⁾ 동 결의안은 (i) 고위험 AI 시스템 운영자에게 엄격책임(Strict liability)을 부과하고, (ii) 고위험 AI 시스템 운영자의 보험가입을 의무화하며, (iii) 피해자가 인과관계를 입증하기 어렵다는 점이 인정되는 경우 인과관계를 추정할 수 있는 근거를 마련하고, (iv) 고위험 AI 시스템에 대해 작동 이력, 결정 과정, 학습데이터 등을 기록하고 저장할 의무를 부과하며, 법원이 증거 공개를 명령할 수 있도록 하는 것을 주된 내용으로 한다.¹⁰¹⁾ 이 결의안은 뒤에서 살펴볼 AILD의 입법적 근거가 되었다.

동 결의안은 EU AI Act의 고위험 AI 개념을 전제로 고위험 AI 운영자에게 엄격책임 및 보험가입의무를 부과한다는 점에서 EG-NTF(2019) 및 EC(2020)와 동일한 입장을 취하고 있다. 즉, 고위험 AI에 대해서는 불법행위책임의 일반 원칙인 과실책임 원칙을 수정하여 보다 무거운 책임인 엄격책임을 부과하고, 고위험이 아닌 AI에 대해서는 과실책임 원칙을 적용하는 것이다. 이러한 책임 체계는 새로운 것은 아니며, EU 회원국은 물론 우리나라 및 주요국들은 모두 중요한 위험원에 대해 별도의 책임법제를 마련하여 엄격책임을 적용하고 있다. 실제 EP(2020)가 제안한 방안은 자동차 사고에 대한 운전자책임 및 의무보험제도와 매우 유사하다.

결의안 발표 당시만 해도 동 결의안에서 정한 고위험 AI 시스템 요건을 충족하는 사례는 많지 않았기 때문에 이러한 의무보험제도는 장래 기술 발전에 대비하는 성격이 강했다. 다만 불확실성을 해소하고, 보험료가 지나치게 높게 정해지는 것을 막기 위해 EU 차원의 사전적 조치가 필요하다고 보았다. 특히 신기술 위험을 담보하는데 필연적으로 수반되는 데이터 부족 및 보험료 산출 곤란 등의 문제를 극복하고, AI 운영자들에게 합리적인 비용으로 적절한 보상을 제공할 수 있는 보험상품을 제공하기 위해서는 EU 집행위원회가 보험업계와 긴밀히 협력할 필요가 있고, 공공 기금을 활용한 피해 보상 공백 해소 방안은 적절하지 않다고 보았다.¹⁰²⁾

100) 정식 명칭은 Civil liability regime for artificial intelligence: European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL))임. 위 결의안은 INL(Own-Initiative Legislative Resolution)에 해당하되, INL은 EU 의회가 EU 집행위원회에 법안의 제안을 요청하는 성격을 가지며, 법적 구속력은 없음

101) 참고로 김진우(2021b)에는 EP(2020) 결의안에 대한 번역문이 별첨으로 수록되어 있음

102) EC(2020), p. 7

3) 2022년 PLD 개정안 및 AILD 제정안 초안 발표 및 이후 경과

EU 집행위원회는 2022년 9월 PLD 개정안 및 AILD 초안을 패키지 형태로 제안하였다. 앞서 EC(2020)에서 본 바와 같이 AI 사고에 대한 EU의 책임법제는 회원국에 공통적으로 적용되는 PLD와 회원국 개별법에 따르는 불법행위법으로 구성되고, 양자는 중첩적으로 적용되며 상호보완적인 기능을 한다. EU 집행위원회가 PLD 개정안과 AILD 제정안 초안을 함께 발의한 것은, EU 공통 법제인 PLD를 AI 사고에 적합하게 수정·보완하는 한편 회원국 개별 법제인 불법행위법에 대해서도 EU 차원의 일관된 기준을 만들기 위한 조치라고 볼 수 있다. 즉 PLD 개정을 통해 기존 제조물책임 법리에 따라 AI 사고 피해자가 구제될 수 있도록 하는 한편 AILD를 통해 일반 불법행위책임을 통해서도 피해자가 구제될 수 있도록 하는 것이다. 이는 'AI 사고에 대해서는 엄격책임과 과실책임이 병용되어야 한다'고 본 EG-NTF(2019)의 입장과의 연결된다.

개정 PLD 초안은 제조물책임의 대상인 제조물의 범위에 소프트웨어가 포함된다는 점, 소프트웨어 업데이트 실패나 사이버보안상의 문제도 결함에 해당한다는 점, 재조합된 제품도 제조물에 해당한다는 점, 안전기준 위반의 경우 결함의 존재를 추정한다는 점 등을 주요 내용으로 하였다. 제안 이후 PLD 개정 논의는 신속하게 진행되어 일부 추가적인 수정·보완이 이루어졌고, 수정안에 대해 2023년 EU 의회와 이사회의 잠정 합의가 이루어졌다.¹⁰³⁾ 이후 2024년 10월 최종 개정안이 공식 채택되어 11월 18일 관보에 게재되었으며 12월 8일 개정 PLD가 정식 발표되었다. 회원국들은 2026년 12월 9일까지 PLD의 취지에 따라 자국의 제조물책임법을 개정할 의무를 부담한다.

한편 AILD 초안은 EP(2020)의 제안에 비해 상당히 완화된 형태로 마련되었다. 고위험 AI 시스템 운영자에 대한 엄격책임이나 의무보험제도는 반영되지 않았고, 입증책임 완화 및 전환에 관한 사항만 반영되었다. 이처럼 완화된 내용으로 초안이 마련되었음에도 불구하고 AILD 초안은 제안 시부터 강한 반발에 부딪혔다. 주된 반대 의견은 기존 책임법제의 수정·보완으로 AI 사고에 대응할 수 있고, 새로운 책임법제를 도입하는 것은 복잡성만 추가하는 입법 과잉이라는 점이었다. 반면 AILD는 반쪽짜리 접근이며 EP(2020)에서 제안된 엄격책임 도입 방안을 보다 적극적으로 검토했어야 한다는 의견도 제기되었다.¹⁰⁴⁾ 초안이 발표된 지 2년이 경과한 2024년 12월, EU 의회의 법제사법위원회(JURI) 논의를 거쳐

103) 개정 PLD의 내용은 뒤에서 구체적으로 살펴봄

104) Hacker(2023), p. 1; European Parliamentary Research Service(EPRS)(2024), p. 3

2025년 1월 공청회가 개최되었으나, 2025년 2월 EU 집행위원회는 AILD를 공식적으로 철회할 예정이라고 밝혔다.¹⁰⁵⁾ 철회 배경으로는 프랑스 주도의 반대 연합, EU 기관들 간 내부 의견의 충돌, 드라기 보고서 이후 EU 경쟁력 확보를 우선하는 정책 기조 등이 영향을 미친 것으로 분석되고 있다.¹⁰⁶⁾ 한편, 애초부터 AILD 제정 기반이 불안정하여 철회 결정은 놀라운 일이 아니며, EU 회원국들의 기존 책임법제 체계를 통해 AI 사고에 대응하는 것이 가능하다는 의견도 제시되고 있다.¹⁰⁷⁾

다. 개정 PLD 주요 내용

1) 책임 범위 확대

개정 PLD는 제조물, 결합, 책임주체의 범위를 확장함으로써 책임 범위를 확대하였다.

가) 제조물에 AI 시스템 등 소프트웨어 포함

개정 PLD는 제조물의 범위에 소프트웨어가 포함된다는 점을 명시하였다(PLD Article 4(1)). 이때 소프트웨어에는 AI 시스템이 포함되며, 독립형 소프트웨어, 내장 소프트웨어, 장치에 저장된 소프트웨어, 통신 네트워크나 클라우드 기술을 통해 접근되는 소프트웨어 또는 소프트웨어 서비스 모델을 통해 제공되는 소프트웨어 등 사용 방식을 불문한다. 개정 PLD는 AI Act에서 정하는 AI 시스템 공급자는 소프트웨어 제조업자로서 제조물책임의 주체가 된다는 점을 전문에서 명시하였다(PLD 전문: (13)). 다만 미디어 파일이나 전자책 같은 디지털 파일의 콘텐츠나 소프트웨어의 소스코드는 제조물에 해당하지 않는다(PLD 전문: (13)).

105) EC(2025) p. 26(집행위원회는 현재 제안된 AILD에 대해 합의 가능성이 없다고 밝히고, 다른 제안을 제출할지 또는 다른 접근 방식을 선택할지 여부를 평가할 예정이라고 하였음)

106) Brian McElligot and Leona Chow(2025); Mason Hayes & Curran(<https://www.mhc.ie/latest/insights/ai-liability>) (2026. 2. 4. 최종 접속)

107) M.N. Duffourc(2025); 동 보고서는 AI 사고 피해 구제는 탐다운 방식보다 바텀업 방식에 의한 접근이 더 효과적일 수 있고, AI 사고에 대해 각국의 법 전통, 판례, 문화적 맥락, 공정성·합리성·비례성 등의 법원칙의 폭넓은 적용이 요구된다고 보면서, "AILD는 기대에 미치지 못했지만, 공정하고 효과적인 AI 책임 체계를 구축하기 위한 여정은 이제 막 시작되었다"고 평가함; Maastricht University(<https://www.maastrichtuniversity.nl/blog/2025/02/withdrawal-ai-liability-directive-critical-reflection-ai-liability-eu>) (2026. 2. 4. 최종 접속)

한편 비상업적 활동 중 개발되거나 제공되는 무료 및 오픈소스 소프트웨어에 대해서는 제조물책임이 적용되지 않는다(PLD Article 2). 무료 및 오픈소스 소프트웨어는 시장에서의 연구와 혁신에 기여하는 것이므로, 시장에 제품을 출시하는 행위와는 구별된다고 본다(PLD 전문: (14)).

나) 결함 여부 판단 시 디지털 신기술 특성 고려

기존의 제조물책임은 제품이 시장에 공급된 시점을 기준으로 제조물의 결함 여부를 판단하였고, 당해 제조물 자체에 결함이 있는지 여부만이 고려대상이었다. 반면 개정 PLD는 결함 여부 판단 시에 AI 시스템을 비롯한 디지털 신기술의 특성을 고려하여 결함 판단 시점 및 고려 대상의 범위를 확대하였다. 즉, 결함 여부를 판단할 때 ① 제품 출시 이후의 학습이나 성능 개선이 제품에 미치는 영향, ② 상호 연계되어 활용할 것이 예상되는 다른 제품이 당해 제품에 미치는 영향, ③ 제품의 출시 및 사용 시점, 제조업자가 제품에 통제권을 유지하는지 여부 및 제조업자가 제품을 더 이상 관리하지 않게 되는 시점, ④ 사이버 보안 요건을 비롯하여 제품 안전 관련 규제 사항, ⑤ 관할 당국에 의한 제품 리콜 등 행정 조치 여부, ⑥ 제품이 적용되는 사용자 그룹의 특수성¹⁰⁸⁾ 등을 고려하여야 한다고 명시하였다(PLD Article 7). 소프트웨어 업데이트나 업그레이드, AI 시스템의 지속적인 학습을 통해 제품에 실질적인 변경이 이루어진 경우, 그 변경 시점에 시장에 제공된 것으로 보아야 한다(PLD 전문 (40)). 이와 같은 결함 판단 기준 및 판단 시점에 관한 규정들은 제조물책임의 인정범위를 보다 확대하는 효과를 가져올 것으로 예상된다. 다만 제품에 대한 리콜이 실시되었다는 점이나 제품에 대한 업데이트나 업그레이드를 포함하여 더 나은 제품이 이미 시장에 출시되었거나 사용되기 시작했다는 점만을 이유로 제품의 결함을 인정해서는 안 된다고 밝힘으로써 결함의 범위가 불합리하게 확대되는 것을 방지하였다(PLD 전문 (34), (35)).

다) 제조물책임주체 확대

개정 PLD는 제조물 책임을 부담하는 책임주체의 범위도 확장하였다. 기존에 책임을 부담하던 제조업자와 준제조업자 외에,¹⁰⁹⁾ 결함이 있는 부품의 제조업자의 경우 해당 부품이

108) 예컨대 생명유지장치의 경우 사용자의 건강상 취약성 등을 고려하여 결함 여부를 엄격하게 판단해야 함(PLD 전문 (30))

제품에 통합되었거나 제품과 상호 연결된 경우에는 경제적 운영자(Economic operator)로서 제조물책임을 부담한다. 이미 시장에 출시되었거나 서비스에 투입된 제품을 현저하게 수정하는 모든 자연인 또는 법인도 제조업자로 간주된다. 또한 유럽 연합 내에서 책임자를 특정할 수 없는 경우, 결합 제품 유통업자나 온라인 플랫폼 제공자도 책임을 부담할 수 있다고 정하였다(PLD Article 8).

2) 입증 부담 완화

개정 PLD는 피해자가 제조물책임 성립의 개연성을 입증하면 법원이 제조업자에게 결합 여부 판단에 필요한 자료를 제출할 수 있도록 하였고, 기술적·과학적 복잡성이 인정되는 경우 간접적 사실만으로도 결합의 존재를 추정할 수 있도록 하였다.

가) 제조업자의 자료제출의무

제조물책임은 제조업자의 고의·과실에 대한 입증을 요구하지 않는다는 점에서 엄격책임의 일종이나, 피해자가 배상을 받기 위해서는 제조물의 ‘결함’ 및 결함과 손해 발생 사이의 ‘인과관계’를 입증해야 한다는 점에서 피해자의 입증 부담은 여전히 남는다. 자동차, 가습기살균제, 담배 등 관련 제조물책임 사건에서 드러난 바와 같이 결함 및 인과관계를 입증하는 것은 쉽지 않다. 첨단기술이 적용된 제조물의 경우 결함 및 인과관계 입증 곤란성은 더욱 커질 것으로 예상된다. 이러한 입증 곤란성이 AI 사고 피해 구제에 현실적인 장애물이 될 수 있다는 점은 EU의 AI 사고 책임법제 논의의 전 과정에서 핵심적인 쟁점으로 지적되어 왔다.

이에 개정 PLD는 소비자 측이 배상 청구의 개연성을 뒷받침할 만한 사실과 증거를 제시한 경우, 제조업자에게 관련 자료를 법원에 제출할 의무를 부과하였다(PLD Article 9(1)). 다만 자료는 청구와 관련된 것으로 필요한 범위 내에서만 제출하도록 하여 영업비밀 및 정보보호와 조화를 이룰 수 있도록 하였다(PLD Article 9(4)).

109) 제조업자는 제품을 개발(Develop), 제조(Manufacture), 또는 생산(Produce)하는 자를 의미함. 준제조업자는 제조업자는 제품을 설계 또는 제조하도록 하였거나, 그 제품에 제조자의 이름, 상표, 또는 그 밖에 그 제품만의 특성을 넣어서 자신이 제조자임을 표시한 자를 의미함

나) 결함 추정 규정 신설

개정 PLD는 기존 제조물책임 법리에 따라 제품의 결함 및 결함과 손해 사이의 인과관계에 대한 입증책임은 피해자가 부담한다고 전제하면서도, 제품의 결함에 대한 추정 규정, 결함과 손해의 인과관계에 대한 추정 규정을 신설하였다(PLD Article 10).

우선, (i) 법원이 제조업자에게 증거 공개를 요청하였음에도 이에 응하지 않은 경우, (ii) 제조업자가 EU 및 회원국 법령에서 명시된 필수적인 제품 안전 기준을 위반한 사실이 인정된 경우, (iii) 합리적으로 예측 가능한 범위 내에서 제품을 이용하던 중에 제품의 명백한 오작동으로 인해 손해가 발생한 경우에는 제품의 결함이 추정된다(결함의 추정). 제품의 결함이 인정되고, 그 결함으로 초래된 피해가 일반적으로 그 결함으로 인해 전형적으로 발생하는 종류의 것이라면 인과관계가 추정된다(인과관계 추정). 특히 AI 등 신기술 적용 제품의 입증 곤란성을 고려하여, '기술적 또는 과학적 복잡성(Technical or scientific complexity)'으로 인해 제품의 결함 및 결함과 손해 사이의 인과관계 입증에 '과도한 어려움(Excessive difficulties)'이 발생하는 경우, 결함 및 인과관계를 추정하도록 하였다(결함 및 인과관계 추정). '기술적 또는 과학적 복잡성' 및 '과도한 어려움'의 의미는 불분명하고 회원국 법원의 판례를 통해 의미가 밝혀질 예정일 것으로 보인다.

라. AILD 초안 주요 내용¹¹⁰⁾

1) 제안 설명

가) 제안 배경(공평한 피해자 보호 및 법적 불확실성 해소)

AI 사고 책임에 대한 법적 불확실성은 EU 회원국 기업들의 AI 활용에 장애 요소로 꼽히고 있다. AI의 복잡성, 자율성, 불투명성과 같은 특성들은 기존의 불법행위책임 요건인 고의·

110) 이하의 내용은 EU 집행위원회가 발표한 AILD 초안 제안서인 "Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non-contractual civil liability rules to artificial intelligence"의 주요 내용을 요약·정리한 것임. 원문을 그대로 번역한 것이 아니고 의미와 가독성을 고려하여 요약·정리한 것이어서 번역문과는 차이가 있음. 원문의 내용에 포함되지 않은 저자의 의견이나 해설은 각주로 기재함. AILD의 경우 최종 법안의 내용보다 그 내용의 도출 과정이 중요한바, 제안 설명 부분의 내용을 충실히 전달하는 데 중점을 두고 요약·정리를 진행함

과실 및 인과관계의 입증을 어렵게 만든다. 이러한 상황에서 EU 회원국 법원에 AI 사고 관련 손해배상청구 소송이 제기되면, 법원은 피해자에게 공평한 보상이 이루어질 수 있도록 사안의 특성에 맞게 기존 법제를 탄력적으로 해석하여 적용하게 될 것이다.¹¹¹⁾ 회원국의 법원들이 이처럼 저마다 개별적인 판단을 하게 되면 기업의 불확실성은 커질 수밖에 없다. 특히 여러 국가에 제품을 공급하는 기업이나 중소기업은 더 큰 영향을 받게 된다. 회원국들이 개별적으로 AI 사고 책임에 대한 입법적 조치를 취할 경우에도 마찬가지로의 위험이 발생한다. 이에 EU 집행위원회는 EU 차원의 통일된 AI 사고 책임법제를 마련함으로써 AI를 개발·활용하고자 하는 기업의 불확실성을 해소하는 한편 AI 사고 피해자들이 일반적인 불법행위 피해자들과 동등한 보호를 받을 수 있도록 보장하기 위해 이 지침을 제안하였다.

AI Act 초안¹¹²⁾은 AI의 안전성을 확보하고 기본권을 보호하기 위한 방안을 제안했다. 안전과 책임(Safety and liability)¹¹³⁾은 동전의 양면과 같아서, 적용되는 시점은 다르지만 상호 간에 강화 작용을 한다. 안전에 관한 규제법¹¹⁴⁾은 위험을 줄일 수 있지만 완전히 제거할 수는 없다. 따라서 위험이 현실화된 경우에는 책임법제가 적용된다. 책임법제는 안전에 관한 규제법을 준수해야 할 경제적 인센티브를 제공함으로써 사고를 예방하는 기능도 한다. AILD와 PLD는 패키지 법제로서, 제품의 결함에 대한 책임은 PLD를 통해 규율하고, 그 밖의 일반적인 불법행위 책임은 AILD를 통해서 규율한다. 이러한 책임법제를 통해 안전에 대한 규제법에도 불구하고 피해가 발생한 경우 피해자가 효과적으로 보상을 받을 수 있도록 한다.

이번 지침은 이해관계자의 법적 불확실성을 낮추고, 특히 중소기업이 국경을 넘어 새로운 시장을 탐색할 수 있도록 함으로써 경제적으로 약 5억 유로~11억 유로의 시장 가치를 창출할 것으로 예상된다. 또한 AI에 대한 사회적 신뢰를 제고하고 AI 사고 관련 사법시스템 접근성을 높이는 등 사회적인 효과도 거둘 것으로 예상된다. 특히 사회적 신뢰 증가는 AI 가치 사슬에 속하는 모든 이해관계자에게 이익이 되고, 건강 및 안전의 증진을 통해 EU 시민 전체에게 이익이 될 것이다.

111) 기존 불법행위법의 요건은 고의·과실, 위법성, 인과관계 등 추상적으로 정해져 있어 법원의 판단의 여지가 상대적으로 큰바, 법원은 개별 사건의 피해자에게 가장 공평한 결과를 도출하기 위해 상당한 재량권을 행사할 수 있음

112) AILD 초안 제안 당시인 2022년 9월에는 AI Act 초안에 대한 논의가 진행 중이었음

113) “Safety and liability”는 규제법인 AI Act와 책임법인 AILD 및 PLD의 연계성을 나타내는 표현으로, 규제법과 책임법의 상호 보완성을 압축적으로 드러냄

114) 원문은 “rules to ensure safety and protect fundamental rights”이며, AI Act를 의미함

나) 법적 근거, 보충성 및 비례성

AILD는 EU가 EU 역내 시장을 형성하고 기능을 보장하기 위한 조치를 취할 수 있도록 하는 TFEU(Treaty on the Functioning of the European Union) Article 114에 법적 근거를 두고 있다. AI 책임법제에 관한 법적 불확실성은 EU 역내 시장의 발전을 저해하며, AI 기반 제품 및 서비스의 국경 간 무역에 장애가 된다. 특히 비계약적 책임인 일반 불법행위책임은 사고가 발생한 국가의 법률이 적용되는바,¹¹⁵⁾ 기업들 입장에서는 관련 책임과 위험을 미리 알고 이에 대해 보험 등을 통해 위험을 관리하는 것이 필요하다. AI 책임법제에 관한 법적 불확실성은 개별 회원국의 독자적 입법으로 인한 법적 분열로 이어질 수 있다. 이미 체코, 이탈리아, 몰타, 폴란드, 포르투갈 등에서 채택된 AI 전략은 책임법제를 명확히 하겠다는 점을 명시하고 있다. 이러한 개별적 입법은 법적 분열을 가속화시킬 것이다. 이처럼 개별 회원국 차원의 대응은 AI 기반 제품 및 서비스 도입에 장애가 되고 법적 분열을 심화시킬 것으로 예상되는바, TFEU Article 114에 따른 조치가 요구된다. EU 차원의 조치는 개별 국가 차원의 조치로 목적을 달성할 수 없는 경우에 보충적으로 적용되어야 하는바, AI 사고 책임법제는 이러한 보충성(Subsidiarity) 요건을 충족하는 것으로 본다.

본 지침은 비례성(Proportionality)¹¹⁶⁾을 준수하기 위해 최소한의 개입으로 목표를 달성하는 방안을 우선적으로 추진한다. 즉, 입증책임 관련 조치를 먼저 채택한다. 현재 회원국 국내법상 일반 불법행위책임 성립 요건에 해당하는 고의·과실, 인과관계 등의 요건을 AI 사고에 대해서도 동일하게 요구하되, 입증 곤란성으로 인해 AI 사고 피해자가 불리해지지 않도록 하기 위해 입증책임에 대한 조치를 마련하는 것이다. 입증책임 관련 조치 중에서도 가장 낮은 수준의 조치인 반증 가능한 추정(Rebuttable presumption)을 채택한다.¹¹⁷⁾ 이러한 반증 가능한 추정은 일반적으로 널리 인정되는 것으로, 원고와 피고의 이익을 균형 있게 조정하는 역할을 수행한다. 또한 본 지침은 AI 관련 제품 및 서비스 제공자의 주의

115) 계약책임의 경우 계약 당사자들이 합의에 의해 준거법(계약의 해석 및 적용에 적용될 법률)이나 관할법원을 정할 수 있는 반면, 일반 불법행위책임의 경우 당사자들이 사전에 준거법이나 관할을 정하는 것이 어려움

116) 법적 조치는 그 목적 달성에 필요한 범위를 넘지 않아야 한다는 의미임

117) 피해자의 입증책임을 완화하는 방법으로는 입증책임 자체를 전환하는 방법(입증책임 전환), 간접사실 증명 시 요건사실이 입증된 것으로 보고 반증을 허용하지 않는 방법(간주 규정), 간접사실 증명 시 요건사실을 추정하되 반대 증거를 통해 그 추정을 번복할 수 있도록 하는 방법(추정 규정) 등이 있음. 피해자의 입증책임을 완화하는 것은 곧 가해자의 지위에 있는 불법행위자의 책임을 보다 쉽게 인정한다는 것을 의미하고, 잠재적 가해자에 해당하는 AI 관련 제품 및 서비스 제공자의 책임을 강화하는 효과가 있음. 이에 EU는 가장 낮은 수준의 입증책임 완화 방안인 반증 가능한 추정을 채택한 것임

의무 수준에 대해서는 기존의 EU 및 회원국 국내법 기준에 따르도록 한다. 이처럼 비례성 원칙을 준수함으로써 AI 시스템 제공자, 운영자 및 이용자가 과도한 위험에 노출되지 않도록 한다. 우선적으로 이러한 조치를 취하여 일정 시간이 경과한 후 그 효과를 평가하여, 엄격책임이나 의무보험 등 더욱 강한 조치가 필요한지 여부를 다시 판단하기로 한다.

한편, 본 제안의 형식으로 법적 효력이 없는 권고(Recommendation), 지침(Directive), 규정(Regulation) 등이 고려되었으나,¹¹⁸⁾ 법적 확실성을 제공하면서도 개별 회원국의 책임 법제와 마찰 없이 통합할 수 있도록 유연성을 제공하기 위해 지침(Directive)의 형식을 채택한다.

다) 관련 연구, 이해관계자 의견수렴 및 입법 영향 평가

EU는 2018년 EG-NTF를 구성하여 AI 책임법제에 대한 준비 작업을 시행하였고, AI 책임 법제 관련 주요 쟁점에 대한 평가 보고서인 EG-NTF(2019)를 제출하였다. EG-NTF(2019)에서 다루지 못한 내용을 보완하기 위해, (i) EU 불법행위법에 대한 비교법적 연구,¹¹⁹⁾ (ii) 책임법제가 소비자의 의사결정, 특히 AI 기반 제품 및 서비스에 대한 신뢰와 수용에 미치는 영향을 다룬 행동경제학 연구,¹²⁰⁾ (iii) AI 사고 피해자의 손해배상 청구의 곤란성의 정도, 기업이 느끼는 AI 책임법제의 불확실성의 정도 및 불확실성이 투자에 미치는 영향, 책임법제의 분열이 EU AI 제품 및 서비스 시장에 미치는 영향, AI 책임법제 마련이 EU 기업들의 AI 기술 채택에 미치는 영향 등에 관한 경제학 연구¹²¹⁾ 등을 수행하였다.

또한 이해관계자 의견수렴을 위해 2021년 10월 18일부터 2022년 1월 10일까지 소비자, 시민단체, 산업별 협회, 중소기업을 포함한 기업, 공공기관 등으로부터 의견을 제출받았고, 21개 회원국 및 제3국에서 233개의 응답이 접수되었다. 응답자들의 다수는 AI 사고 관련 입증책임의 문제, 법적 불확실성 및 법제 분열의 문제들에 공감하였고, EU 차원의 조치가 필요하다는 점을 지지하였다. 시민단체 및 학술기관은 피해자의 입증책임 완화를 위한 EU 차원의 조치 필요성을 적극 지지하였다. 기업들은 책임법제의 불확실성에 대한 우려

118) 권고(Recommendation)는 법적 효력이 없고, 지침(Directive)은 회원국이 지침의 취지에 따라 국내법을 개정할 의무를 부과하고, 규정(Regulation)은 그 자체로 곧바로 회원국에 적용됨. AI 윤리 가이드라인은 권고에 해당하고, PLD는 지침에 해당하며, GDPR 및 AI Act는 규정에 해당함

119) Ernst Karner, Bernhard A. Koch and Mark A. Geistfeld(2021)

120) Kantar(2021)

121) Deloitte(2021)

를 나타내면서도, 혁신을 제한하지 않기 위한 방안에 대한 검토를 요청하는 등 신중한 입장을 취하였다.

EC는 (i) 입증책임 완화 방안, (ii) 입증책임 완화에 더하여 특정한 위험성 요건을 충족하는 경우 엄격책임 및 의무보험제도를 도입하는 방안, (iii) 먼저 (i)의 조치를 취한 후 재검토를 거쳐 (ii) 도입 여부를 결정하는 방안에 대해 입법 영향 평가를 실시하였다. 입법 영향 평가는 효과성, 효율성, 일관성, 비례성을 고려한 다기준 분석(Multi-criteria analysis)에 의하였고, 그 결과 (iii)의 방안이 가장 효과적인 것으로 나타났다.

2) 초안 내용

AILD 초안은 AI에 기인한 손해에 대하여 일정한 경우 인과관계를 추정한다(AILD Article 4). 고위험 AI의 경우 원칙적으로 추정이 적용되고 다만 AI Act에 따라 설계, 개발된 경우에는 추정이 적용되지 않는다. 고위험 AI 이외의 경우에는 인과관계 입증에 지나치게 곤란하다고 법원이 판단한 경우에 한하여 추정이 인정된다. 한편 고위험 AI의 로깅 등에 대해서는 증거 공개 절차가 도입되어, 원고가 손해배상청구의 타당성을 입증하기에 충분한 사실 및 증거를 제시한 경우에는 법원이 고위험 AI 시스템 제공자에게 증거 공개를 명할 수 있도록 하였다(AILD Article 3).¹²²⁾

이처럼 AILD 초안에는 기존 EG-NTF(2019)에서 논의된 내용 중 인과관계 추정 및 증거 공개 절차에 관한 내용이 포함되어 있으나, 핵심이라 할 수 있는 운영자의 엄격책임 및 의무보험에 관한 사항은 포함되지 않았다. AILD 초안의 제안 설명에 따르면 AI 사고 관련 책임 중 엄격책임은 기존 제조물책임 법리에 따라 PLD에서 다루고, AILD는 과실책임을 전제로 마련되었으며, AILD에 AI 운영자의 엄격책임을 도입할 것인지 여부는 추후 논의할 방침이라고 한다.¹²³⁾

122) 이에 관한 보다 상세한 설명은 신봉근(2023), p. 222를 참조 바람

123) European Commission, "Proposal for a AILD", p. 14

5. 시사점

가. AI 사고 피해 구제 논의의 필요성

EU의 AI 사고 피해 구제 관련 논의의 목적은 크게 두 가지로 구분해 볼 수 있다. 첫째, AI 사고 피해자가 다른 사고 피해자보다 피해 보상에 있어서 더 불리한 지위에 놓여서는 안 되고, AI 사고에 대한 책임을 공평하게 배분해야 한다는 것이다. 이는 불법행위법이 추구하는 목적인 피해의 회복 및 공평한 책임 배분이 AI 사고 영역에서도 달성되어야 함을 의미한다. 둘째, EU 역내 시장의 형성 및 유지를 위해 AI 사고에 대한 통일성 있는 법적 규율이 필요하다는 점이다. 회원국마다 개별적으로 법원 판례에 의한 법리 형성 및 입법 조치가 이루어질 경우 EU 역내 시장의 기능 저하를 야기할 수 있다는 점은 EU가 다른 나라에 앞서 선제적으로 포괄적인 AI 책임법제 도입을 추진하게 된 주된 원동력이라고 할 수 있다. EU가 개인정보 보호에 관한 포괄적 법제인 GDPR을 제정하여 역외 기업에도 이를 적용하자 다른 나라들이 모두 EU의 기준에 따라 자국의 개인정보 보호 법제를 마련한 ‘브뤼셀 효과’가 AI 법제 영역에서도 작동하기를 기대한 측면도 있는 것으로 보인다.

AI 사고 시 피해자를 구제하고 공평하게 책임을 배분할 수 있는 방안을 마련하는 것은 전 세계 공통의 과제라 할 수 있다. 따라서 이 문제에 관한 EU의 논의 내용은 우리나라에서도 적극 참고할 필요가 있다. 반면 역내 시장의 형성 및 유지를 위한 선제적·통일적 기준 마련의 필요성은 EU의 특수한 사정이라고 볼 수 있다. 우리나라, 일본 등의 경우 AI 사고 책임에 관한 문제 제기가 지속되고 있음에도 입법적 조치로 나아가는데 보다 신중을 기할 수 있는 것은 이러한 제도적 환경의 차이 때문인 것으로 생각된다. 미국의 경우 연방 차원의 통일된 입법적 조치에 대한 요구가 없는 것은 아니나, 기술 및 산업 발전을 우선하는 정책의 영향으로 입법에 신중을 기하고 있는 것으로 보인다.

요컨대 EU의 논의를 참고하여 AI의 특성 및 그로 인한 AI 사고의 특수성을 파악하고 AI 사고에 대응할 수 있는 피해 구제 체계에 대한 논의를 본격적으로 진행해야 할 것이다. 다만 이는 입법 자체를 서둘러야 한다는 의미는 아니다. ‘세계 최초’ 같은 수사나 ‘피해자 보호’ 같은 이념이 앞설 경우 자승자박의 결과를 낳게 될 우려가 있는바, 논의는 신속히 하되 입법 자체는 신중할 필요가 있을 것이다.

나. AI 사고 책임법제 논의의 전개 과정

EU의 AI 사고 피해 구제 관련 논의는 초기에는 전자인 제도 도입과 같은 급진적인 제도 변화 필요성에 대한 주장이 시선을 끌었으나 후반으로 갈수록 신중론에 무게가 실린 것으로 평가해 볼 수 있다.

논의 초기인 2017년에는 AI 사고에 대한 책임 공백을 방지하기 위해 AI에 법인격을 부여하는 전자인(電子人, e-Person) 제도가 필요하다는 주장이 제기되었고, EU 의회에서 관련 제도 도입에 관한 결의가 이루어지기도 하였다. 그러나 AI 사고 책임법제 논의가 본격화된 2020년 무렵부터는 EG-NTF(2019), EC(2020), EP(2020) 등 EU의 전문가 및 지도자 그룹은 일관되게 AI 사고 책임 문제를 해결하기 위해 전자인 제도를 도입할 필요는 없다고 보았다. 대신 AI 사고 책임법제 문제를 제조물책임법과 일반 불법행위법의 상호 보완을 통해 해결하고자 하였다.

제조물책임법의 경우 결함을 책임의 근거로 삼는 기본 틀은 종전과 동일하게 유지하되, 기술 발전 상황에 맞게 소프트웨어도 제조물에 포함된다는 점을 명시하고, 책임자의 범위를 확대하고, 입증책임을 완화하는 방향으로 개정하는데 합의가 이루어졌다. 반면 일반 불법행위법의 경우 과실책임 원칙을 적용하게 되면 AI 사고 피해자를 구제하기 어렵다고 보았다. 특히 AI Act 제정 과정에서 '위험기반 접근법'이 채택되며, 책임법제 측면에서도 이러한 위험기반 접근법을 반영할 필요성이 대두되었고, 위험성이 높은 AI가 유발한 사고에 대해서는 보다 엄격한 책임 체계가 요구된다는 견해가 설득력을 얻었다. 이에 EU 집행위원회 및 의회는 고위험 AI에 대해서는 엄격책임과 의무보험제도를 도입하고 그 밖의 AI 사고에 대해서는 과실책임인 기존 불법행위책임을 적용하되 피해자의 입증 부담을 줄여주기 위한 방안을 제안하였다. 고위험 AI에 대해 적용되는 엄격책임과 의무보험제도는 현재 자동차 사고에 적용되는 운전자책임 및 자동차책임보험을 모델로 하여 마련되었다.

그러나 고위험 AI에 대해 엄격책임 및 의무보험제도를 도입하는 방안은 다양한 반발에 부딪혔고, 2022년 발표된 AILD 초안에는 고위험 AI에 대한 엄격책임 및 의무보험제도는 포함되지 못하였다. 이후 딥시크 쇼크 및 AI 관련 국가 간 경쟁의 심화 등으로 인해 AI 책임법제에 대한 회의적인 시각이 커졌고, 이미 발효된 AI Act의 시행을 유보해야 한다는 주장도 제기되는 등 분위기가 반전되면서 AILD는 더욱 힘을 잃게 되어 결국 철회되기에 이르렀다.

신기술이 야기하는 사고에 대한 대응 방안을 논의할 때 초기에 급진론이 우세하다 이후

신증론으로 전환되는 것은 익숙한 현상이다. 자율주행차 관련 책임법제 논의도 이와 유사한 과정을 거쳤다. 자율주행차 책임법제 논의 초기에는 자율주행차의 위험성, 제조업자의 책임성, 운전자의 책임 제한 필요성이 강조되면서, 자율주행차 사고에 대해서는 무조건 제조업자가 책임을 부담하도록 해야 한다는 다소 급진적인 견해가 제기되었다. 자율주행차 시스템 오류나 해킹 위험성에 대한 막연한 두려움, 기업이란 이윤만 추구하고 책임은 회피하려는 성향이 있으므로 사전에 책임을 못 박아 두지 않으면 국민 전체에게 손해가 전가될 수 있다는 우려는 이러한 급진적 견해를 확산시키는 양분이 되었다. 그러나 전문가들의 검토, 이해관계자들의 의견 조율 등을 거치는 과정에서 이러한 급진적 대응 방안이 실제 피해자 구제에 효과적이지도 않고, 법리상으로도 문제점이 많으며, 자율주행차 기술 및 산업 발전을 저해할 수 있다는 점이 드러나게 되었다. 자율주행차 사고 책임 문제는 오랜 논의 과정을 거쳐 기존의 운행자책임과 자동차보험을 일부 보완하여 활용하는 것으로 일단락되었으나, 자율주행차의 발전 단계에 따른 재논의 가능성은 여전히 열려있는 상황이다.

AI 책임법제 논의는 우리나라에서도 곧 본격화될 것으로 보인다. 인공지능 기본법이 시행되었고, 전자인 제도 도입이나 AI 책임법에 관한 법안이 국회에 발의되기도 하였다. 논의 과정에서 AI 관련 사업자에게 무과실책임을 부과해야 한다거나, 보험가입을 의무화해야 한다는 주장이 제기될 것으로 보인다. 이에 반대하며 신증론을 제기하거나 기술 개발 및 산업 진흥을 강조하는 것은 AI의 위험성을 외면한 주장으로 치부될 수도 있을 것이다. 이때 EU의 사례, 특히 2017년 전자인 제도 관련 EU 의회 결의부터 2025년 AILD 초안 철회에 이르기까지의 과정에서 이루어진 구체적인 공방의 과정과 내용은 중요한 참고 사례가 될 수 있을 것이다.

다. AI 사고 피해 구제 체계

EU의 PLD는 기존의 제조물 결함으로 인한 사고에 대한 책임이라는 기본 틀을 유지하며 AI 사고에 적합하도록 일부 수정·보완을 하는 것으로 개별적 대응에 해당한다. 반면 AILD는 모든 영역에서 발생하는 AI 사고를 대상으로 하는 포괄적 대응에 해당한다. AI 사고에 대한 사후적 대응은 포괄적 대응과 개별적 대응이 모두 가능하므로, EU에서 포괄적 대응 체계로서 AILD 도입을 추진한 것 자체에 어떠한 문제가 있다고 보기는 어렵다. 문제는 AI 사고에 대해 포괄적 대응 체계를 도입할 수 있는 현실적·제도적 여건이 갖추어졌는지 여부이다.

PLD와 AILD는 처음에는 상호보완적인 패키지 법안으로 출발하였다. 그러나 PLD가 당초 의도한 일정과 내용에 따라 개정이 완료된 것과 달리 AILD는 논의를 거치는 과정에서 제도 도입의 핵심이라 할 수 있는 고위험 인공지능 사업자에 대한 엄격책임 및 보험가입 의무 부과가 법안에서 제외되었고, 결국 AILD 초안은 입증책임 완화로 대폭 축소되어 발표되었으며, 그마저도 철회되었다. 이처럼 PLD와 AILD가 함께 추진되었음에도 불구하고 다른 결과에 이른 것은 개별적 대응과 포괄적 대응에 근본적인 차이가 있기 때문일 것이다.

AI가 도입 초기 단계인 점, AI의 범용성으로 인한 사고 범위 및 피해 범익의 광범위성 등을 고려할 때, 모든 AI 사고에 적용될 수 있는 하나의 포괄적 책임법제 및 보험제도를 마련하는 것은 처음부터 어려운 일이었다고 생각된다. AI는 ‘자율성’, ‘적응성’ 같은 독특하고도 중요한 특성을 가지고, 기술의 ‘범용성’으로 인해 국민의 생활과 산업의 거의 전 영역에서 활용됨에 따라 사고 유형이나 침해되는 법익도 다양하다. 따라서 이러한 독특하고 광범위한 ‘AI 사고’를 하나의 책임법제와 보험제도로 해결하려는 시도는 적어도 AI 시대에 이제 막 진입하는 현 단계에서는 적절치 않아 보인다. 논의가 충분히 성숙되지 않은 상태에서 EU의 AILD 입법이 관철되지 않고 철회된 것이 오히려 다행스럽게 여겨지기도 한다. EU의 입법이 우리나라에 미치는 즉각적이고 강력한 영향을 고려할 때, EU에서 AILD가 통과되었다면 우리나라도 서둘러 인공지능 책임법을 도입해야 한다는 주장이 봇물 터지듯 제기되었을 것이기 때문이다.

라. 소결

EU의 사례는 AI 책임법제 논의의 목적이 무엇인지, 논의가 어떠한 방향으로 전개되는지, 포괄적 책임법제 도입이 현실점에서 가능하고 적절한지 등에 대한 시사점을 제공한다. AI 책임법제 논의를 시작해야 하는 우리나라의 입장에서 EU의 사례는 두 가지 측면에서 중요한 참고 사례가 된다.

첫째, AI 책임법제 문제는 하루아침에 결론을 낼 수 있는 것이 아니라 장기적인 검토와 논의, 기술 및 산업 발전 상황에 대한 고려, 국제적 동향과의 조율이 필요하다는 점이다. AI 활용이 확대되고 그에 따른 위험과 불확실성이 커짐에 따라 전격적인 입법을 통해 이러한 위험과 불확실성에 대처하고자 하는 목소리가 높아지고 있다. 물론 불확실성을 해소하기 위한 입법도 필요할 수 있다. 그러나 법에서 다루는 대상의 존재나 특성 자체가 명확하지 않은 상황에서 그에 대한 법제를 선제적으로 명확하게 하려는 노력은 오히려 위험하다.

특히 기존에 개별 영역에서 작동하고 있는 사후 구제 체계를 폐기하고 이를 대체하는 포괄적 피해 구제 체계를 수립하는 방식의 대응은 위험과 불확실성을 줄이기는커녕 오히려 증폭시킬 수 있다.

둘째, EU가 거쳐온 과정을 살펴봄으로써 우리나라가 거쳐야 할 논의의 과정을 보다 압축적으로 진행할 수 있을 것이다. AILD 초안이 비록 철회되기는 하였으나 AILD 초안 작성 및 철회 과정에서 이루어진 다양한 논의들, 특히 AILD 입법론자들이 가졌던 공평한 피해자 구제 필요성에 대한 인식은 주목할 가치가 있다. EU AILD 논의 과정은 우리나라가 거쳐야 하는 시행착오 과정일 수 있으며, 이러한 과정을 거치지 않고 입법이 곧바로 이루어지는 것이 오히려 더 문제일 수 있다. 따라서, EU의 논의 과정과 내용을 충분히 검토하여 우리의 시행착오를 줄이고 사회적 합의에 이르는 과정을 단축하는 데 활용할 필요가 있다.

‘AI가 사고를 일으키면 누가 책임을 져야 하는가, 또 보험에 의한 보상은 어떻게 이루어져야 하는가’라는 질문은 AI 등장 이래 현재까지 계속하여 제기되고 있다. 자율주행차가 오작동하여 보행자를 충돌하면 누가 책임을 져야 하는가, 수술 로봇이 오작동하여 환자가 중태에 빠지면 누가 책임을 져야 하는가, 서빙 로봇이 손님에게 뜨거운 음식을 쏟으면 누가 책임을 져야 하는가, 로보어드바이저가 잘못된 투자 자문을 하여 고객의 재산에 손실이 발생하면 누가 책임을 져야 하는가, AI 설계사가 보험상품을 불완전판매하여 보험계약자에게 손해가 발생하면 누가 책임을 져야 하는가. 이 모든 질문들에 담겨 있는 핵심 쟁점은 동일하다. 인간의 지적 능력과 유사한 지능을 가진 기계가(기계지능), 인간의 개입을 요하지 않는 상태에서(자율성) 스스로 학습 및 개선을 하면서(적응성) 우리 삶의 거의 모든 분야에서(범용성) 일정한 출력물을 제공하는 과정에서(출력물 제공) 손해가 발생하고, 그 손해가 우리 법질서 관점에서 위법한 것으로 평가되는 경우, 그 손해에 대한 책임을 누구에게 귀속시키는 것이 가장 합리적인가 하는 것이다.

이 질문에 대해서는 다양한 답변이 제시되어 왔다. AI에게 법인격을 부여해서 AI가 책임을 져야 한다. AI를 이용하는 사람은 잘못이 없으니 AI를 만든 사람이 책임을 져야 한다. 하드웨어 제조사는 잘못이 없으니 소프트웨어 개발자가 책임을 져야 한다. 국가나 공동체가 책임을 져야 한다 등 질문만큼이나 다양한 답변이 존재한다. 그러나 적어도 현시점에서는 AI 사고 책임 문제에 대한 단순하고 명쾌한 해답을 제시하기는 어렵다고 생각된다. 이는 AI 자체가 새롭고 독특한 위험원이기 때문이기도 하지만, 본래 사고에 대한 책임 문제가 단일한 해결책으로 풀릴 수 있는 문제는 아니기 때문이기도 하다. 일상적인 사고로 여겨지는 자동차 사고의 경우만 하더라도, 자동차 사고 가해자와 피해자 사이에서는 자배법상 운행자책임, 민법상 불법행위책임과 사용자책임이 문제 되고, 차량 결함이 사고 원인인 경우 가해 차량 차주와 자동차 제작사 사이에는 제조물책임 및 계약상 하자담보책임 등이 문제 된다. 따라서 누군가 만약 AI 사고 책임과 보상 문제에 대해 단순 명쾌한 해결책을 제시한다면 오히려 의심해 볼 필요가 있다.

AI 사고 문제에 실질적으로 대응하고 있는 것은 각 분야별 현황 및 특성을 디테일하게 반영하여 마련된 각종 안전기준과 가이드라인들, 그리고 사고가 발생했을 때 피해자에게 신속한 보상을 담당하는 보험제도이다. 대표적인 예라고 할 수 있는 자율주행차의 경우 자율주행차의 제작이나 수입 시 준수해야 하는 안전기준과 도로에서 자율주행차를 운행할 때 지켜야 할 운행기준이 마련되어 있다. 또한 자율주행차도 기존 자동차와 동일하게 차주가 자동차 종합보험에 가입하도록 하여 사고 발생 시 자동차보험으로 피해를 보상하도록 하고, 자율주행차 특약을 별도로 마련하여 해킹사고나 시스템 오작동 사고도 자동차보험에서 보상하도록 하고 있다. 사고조사위원회가 신설되어 자율주행차 사고 조사를 전담하는 등 자율주행차 사고 원인 규명, 통계 수집 노력도 체계적으로 이루어지고 있고, 자율주행 기술 발전 단계 및 활용 상황, 개인정보 보호법 등 관련 제도 변경 상황을 실시간 모니터링하여 필요한 경우 수시로 관련 기준을 업데이트하고 있다. 의료, 금융 등 AI를 활용하고 있는 다른 분야에서도 이와 유사한 노력들이 이루어지고 있을 것으로 생각된다.¹²⁴⁾

AI 사고 피해 구제 체계에 관한 논의가 본격화되면 언젠가는 AI 사고에 대한 포괄적 피해 구제 체계로서 책임법제와 보험제도도 마련될 수 있을 것이며, 이에 대비한 논의를 지금부터 단계적으로 진행해 나갈 필요가 있다. 다만, 이와 별개로 현실에서 발생하는 AI 사고에 대응하기 위해서는 개별 영역별 피해 구제 체계를 보완할 필요가 있다. 이와 관련하여 세 가지 향후 과제를 제시해 보고자 한다.

첫째, AI 사고에 대응할 수 있는 책임보험의 현황을 점검해 볼 필요가 있다. 특히 피지컬 AI의 경우 물리적 실체가 있어 오작동 시 사람의 생명·신체에 직접적인 위해를 야기할 수 있는바, 공공장소에서 활용되는 피지컬 AI에 대해 책임보험이 마련되어 있는지를 점검해 볼 필요가 있다. 해당 영역에서의 사고를 보상하는 기존 보험이 있는 경우에는 해당 보험을 확장하여 AI 사고도 보상할 수 있도록 하고, 기존 보험이 없는 경우에는 신규 보험을 도입할 필요가 있을 것이다. 전자에 해당하는 것으로 자율주행차보험이 있고, 후자에 해당하는 것으로는 최근 법제화된 실외이동로봇보험이 있다.

둘째, 분야별 책임법제의 내용을 AI 사고의 관점에서 재검토해 볼 필요가 있다. 현행 자배법이 자율주행차 사고에 대응할 수 있는지, 현행 의료법 및 의료배상책임 관련 판례의 법리가 진단 및 수술 로봇 관련 사고에 대응할 수 있는지, 현행 금융소비자보호법이 금융 AI에 의한 금융소비자 권익 침해에 대응할 수 있는지 등을 점검해 보는 것이다. 이에 대해서

124) 황현아(2024b), pp. 7~8

는 분야별로 상당한 연구가 이루어지고 있는바, 이러한 연구들을 보험의 관점에서 정리하고 구체적 보상 방안의 마련으로 이어질 수 있도록 할 필요가 있다.

셋째, 장기적인 관점에서 AI 위험에 대한 정의, 분류, 평가를 통해 가계, 기업, 정부 등 경제주체별 AI 위험을 구체화하고, 위험 유형별로 보험을 통한 위험관리 체계를 구축할 필요가 있다. AI 위험을 보험을 통해 담보함으로써 AI 시대 보험산업의 역할을 정립함과 동시에 신성장동력을 마련하고, AI 사업자들이 사고와 책임에 대한 과도한 우려에서 벗어나 적극적으로 AI를 개발·활용할 수 있도록 지원하며, 국민들이 일상생활에서 안심하고 AI를 활용할 수 있도록 해야 할 것이다. AI 사고 책임법제 및 보험제도는 이러한 AI 위험관리 체계의 하나로서 기능하게 될 것이므로, 전체적인 AI 위험관리 체계 수립에 대한 고민이 함께 이루어질 필요가 있다.

참고문헌

- 김진아(2022), 「인공지능 오류사고 민사책임에 대한 입법론적 검토」, 『재산법연구』, 39(2)
- 김진우(2021a), 「인공지능 시스템의 책임능력」, 『중앙법학』, 제23집 제4호
- _____ (2021b), 「인공지능 운영자책임에 관한 유럽의회의 초안: 평가 및 우리 법에의 시사점」, 『법학논문집』, 45(1), 중앙대학교 법학연구원
- 대런 아세모글루·사이먼 존슨(2023), 『권력과 진보: 기술과 변명을 둘러싼 천년의 쟁투』, 생각의 힘
- 박상철(2024), 「인공지능 기본법의 시행 전 개정 필요성-규제 조항의 체계·축조상 문제점을 중심으로」, 『정보법학』, 28(3), 한국정보법학회
- 신봉근(2023), 「인공지능(AI)의 민사책임 - EU의 입법을 중심으로」, 『민사법학』, 105, 한국민사법학회
- 심소연(2024), 「규제 중심의 유럽연합 인공지능법(EU AI Act)」, 『최신외국입법정보』 2024-4, 국회도서관
- 이해원(2023), 『인공지능과 불법행위책임』, 박영사
- 이현균(2024), 「인공지능 모빌리티 관점에서의 도심항공교통(UAM) 보험제도 검토」, 『교통학연구』, 3-2, 한국교통법학회
- 정진명(2024), 「인공지능의 법인격에 관한 시론」, 『재산법연구』, 40(4)
- 지광운(2024), 「인공지능에 의한 사고발생 위험과 보험제도의 대응방안에 관한 소고」, 『법과 정책연구』, 24-2, 한국법정책학회
- 채은선(2024), 「EU AI법의 주요 내용 및 시사점」, 『디지털 법제 Brief』, 한국지능정보사회진흥원
- 한국지능정보사회진흥원(2025a), 「트럼프 2기 행정부 AI·디지털 정책 동향 분석(1~3월)」, 『주요국 AI·디지털 정책 모니터링 리포트 THE LENS』, 2025-3
- _____ (2025b), 「일본 ‘AI 전략회의·AI 제도 연구회 중간보고서’ 주요 내용-‘세계에서 가장 AI 개발·활용이 쉬운 나라’를 목표로 AI 제도 마련」, 『THE AI REPORT』, 2025-4
- 황현아(2024a), 「인공지능 규제법, 책임법, 그리고 보험」, 『규제법제리뷰』, 24-3, 한국법제연구원

_____(2024b), 「인공지능 시대 보험의 역할과 과제」, 『KIRI 보험법리뷰』, 보험연구원
_____(2026), 『AI와 보험』, 커뮤니케이션북스

Anat Lior(2022), “Insuring Ai: The Role Of Insurance In Artificial Intelligence Regulation”,
Harvard Journal of Law & Technology, Volume 35, Number 2, Spring 2022

Baris Soyer And Andrew Tettenborn(2022), “Artificial intelligence and civil liability-do
we need a new regime?”

Brian McElligot and Leona Chow(2025), “AI Liability From the AI Liability Directive to
the current framework”, Mason Hayes & Curran

CMA(2024), “AI Foundation Models: Technical update report”

_____(2023), “AI Foundation Models: Initial Report”

Deloitte(2021), “Study to support the Commission’s Impact Assessment on liability for
artificial intelligence”

DSIT(2025), “AI Opportunities Action Plan”

_____(2023a), “A pro-innovation approach to AI regulation”

_____(2023b), “Frontier AI Taskforce: first progress report”

_____(2023c), “Frontier AI Taskforce: second progress report”

_____(2023d), “The Bletchley Declaration by Countries Attending the AI Safety Summit,
1-2 November 2023”

EC(2025), “ANNEXES to the COMMUNICATION FROM THE COMMISSION TO THE
EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND
SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Commission
work programme 2025”

_____(2020), “Report on the safety and liability implications of Artificial Intelligence, the
Internet of Things and robotics”

EG-NTF(2019), “Liability for Artificial Intelligence and Other Emerging Digital
Technologies”

Ernst Karner, Bernhard A. Koch and Mark A. Geistfeld(2021), “Comparative Law Study on Civil Liability for Artificial Intelligence”, EC

European Parliament(EP)(2017), “Resolution on Civil Law Rules on Robotics with recommendations to the Commission”

European Parliament(EP)(2020), “Civil liability regime for artificial intelligence”

European Parliamentary Research Service(EPRS)(2024), “Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence- Complementary impact assessment”

HM Government(2021), “National AI Strategy”

ICO(2023), “Guidance on AI and data protection”

Kantar(2021), “Behavioural Study on the link between challenges of Artificial Intelligence for Member States’ civil liability rules and consumer attitudes towards AI-enabled products and services, Final Report”, EC

Michael Faure and Shu Li(2022), “Artificial intelligence and (compulsory) insurance”, JETL, 13(1)

M.N. Duffourc(2025), “The Withdrawal of the AI Liability Directive: A Critical Reflection on AI Liability in the EU”, Maastricht University

OECD(2024), “Defining AI Incidents and Related Terms”

Philipp Hacker(2023), “The European AI Liability Directives-Critique of a Half-Hearted Approach and Lessons for the Future”

UK Government(2023), “Introducing the AI Safety Institute”

White House(2025), “America’s AI Action Plan”

_____ (2016), “Preparing for the Future of Artificial Intelligence”

内閣府(2016), “第5期科学技術基本計画(平成28～平成32年度)”

株式会社読売新聞グループ本社, 日本電信電話株式会社(2024), “生成 AI のあり方に関する共同提言”

総務省, 経済産業省(2024), “AI 事業者ガイドライン案”

AI ネットワーク社会推進会議(2019a), “AI 利活用ガイドライン”

_____ (2019b), “AI 利活用ガイドライン～AI 利活用のためのプロ
ラクティカルリファレンス～”

_____ (2017), “国際的な議論のためのAI 開発ガイドライン案”

AI 原則の実践の在り方に関する検討会, AI ガバナンス・ガイドライン WG(2019), “AI 原則
実践のためのガバナンス・ガイドライン Ver. 1.1”

AI Incident Database, Editors’ guide(<https://incidentdatabase.ai/editors-guide/>)

AI Incident Database, AIID Blog(<https://incidentdatabase.ai/blog/incident-report-2025-april-may/>)

Clyde&Co(2025), “The relaunched UK AI Regulation Bill – a step towards statutory regulation of AI in the UK? Insight Article”(<https://www.clydeco.com/en/insights/2025/03/the-relaunched-uk-ai-regulation-bill-a-step-toward>)

Clyde&Co(2024), “Product Liability legislative reforms: A snapshot of the UK’s position”(<https://www.clydeco.com/en/insights/2024/12/product-liability-legislative-reforms>)

Maastricht University(<https://www.maastrichtuniversity.nl/blog/2025/02/withdrawal-ai-liability-directive-critical-reflection-ai-liability-eu>)

Mason Hayes & Curran(<https://www.mhc.ie/latest/insights/ai-liability>)

White House, LAUNCHING THE GENESIS MISSION(<https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/>)

도서회원 가입안내

회원	연회비	제공자료
법인회원	₩300,000원	- 연구보고서 - 기타보고서 - 연속간행물 · 보험금융연구
특별회원	₩150,000원	
개인회원	₩150,000원	

* 특별회원 가입대상 : 도서관 및 독서진흥법에 의하여 설립된 공공도서관 및 대학도서관



가입 문의

보험연구원 도서회원 담당

전화 : (02)3775-9113 | 팩스 : (02)3775-9102



회비 납입 방법

무통장입금

- 계좌번호 : 국민은행(400401-01-125198) | 예금주: 보험연구원



자료 구입처

서울 : 보험연구원 자료실(02-3775-9113 | lsy@kiri.or.kr)

연구보고서

2026-03

AI 사고 피해 구제의 쟁점과 과제: EU 사례를 중심으로

발행일 2026년 4월

발행인 김헌수

발행처 보험연구원

주소 서울특별시 영등포구 여의대로 70 ONE CENTINEL

인쇄소 경성문화사

ISBN 979-11-93021-97-2
979-11-85691-50-3(세트)

(정가 10,000원)

