

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)
		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.