

인공지능(AI)의 환각(Hallucination)과 보험산업

요약

인공지능의 활용은 기업 경영과 국가 경쟁력 강화를 위한 필수 조건으로 자리매김했으나, 생성형 인공지능의 구조적 한계로 인해 환각, 편향, 침해 등 지금까지 경험하지 못했던 새로운 유형의 리스크를 초래하고 있음. 인공지능의 불완전성은 허위 판례 인용 등으로 인한 법적 분쟁으로 이어지고 있으며, 보험산업은 손해를 급증에 대응하여 면책특약을 도입하고 있음. 반면, 인공지능 리스크는 사이버보험에 이어 새로운 시장 개척의 기회로 부각되고 있으며, Deloitte는 관련 보험시장 규모가 2032년 48억 달러에 이를 것으로 전망함

1. 인공지능(AI) 시대의 도래

- 인공지능은 산업 생산 방식을 근본적으로 재편하며, 지능형 자동화와 데이터 기반 의사결정을 통해 사회·경제 전반의 패러다임 전환을 가속화하고 있음
 - Bain & Company에 따르면 미국 기업의 95%, Mckinsey 조사에서는 79%의 기업이 생성형 인공지능(AI)을 업무에 활용하고 있음(Bain & Company(2025), "Survey: Generative AI's Uptake Is Unprecedented Despite Roadblocks"; McKinsey(2025), "The state of AI in 2025")
- 인공지능의 활용은 새로운 기술 도입의 수준을 넘어, 이미 기업 경영과 국가 경쟁력 강화를 위한 필수 조건으로 자리매김함

2. 인공지능의 환각, 편향, 침해

- 인공지능은 산업과 경제뿐만 아니라 일상생활 전반을 고도화하고 있으나, 환각(Hallucination), 편향(Bias), 침해(Infringement) 등 지금까지 경험하지 못한 새로운 리스크를 초래하고 있음
 - 인공지능의 환각으로 생성된 허위 판례 인용(Walmart Inc. v. Morgan & Morgan), 저작권 침해(Graham v. Prince), 의료 AI 챗봇의 자살 유도(Raine v. OpenAI) 등 관련 사고가 지속적으로 발생하고 있음

- 이외에도 에어캐나다 챗봇의 환불 규정 안내 오류(Moffatt v. Air Canada), 의료 오진, 알고리즘에 의한 사회적 차별, 산업용 로봇의 오작동, 딥페이크를 포함한 사이버 범죄 등 각종 인공지능 관련 리스크에 노출되어 있음
 - 인공지능 결과물을 검증 없이 활용할 경우 기업의 법적 분쟁 가능성이 증가하며, 이는 특히 배상책임보험의 손해를 악화를 초래하는 리스크 요인으로 부각됨
- 환각, 편향, 침해 등의 문제는 학습 데이터의 불균형과 인공지능의 확률 기반 언어 생성 모델의 구조적 특성에 기인하는 본질적 한계로 분석됨
- 존재하지 않는 정보, 사실과 다른 내용, 부정확한 진술을 생성하는 환각은 단순한 운영상 오류가 아니라 생성형 모델의 통계적 추론 방식 등 기술적 특성에 내재된 한계에 기인하는 것으로 분석됨
 - 또한, 인공지능의 처리 과정과 작동 원리를 정확히 파악하기 어려운 불투명성(Black box)과 자동화된 의사결정 구조는 오류의 원인 규명을 더욱 어렵게 하고 있음
- 현재 인공지능 관련 사고의 책임 귀속에 대한 논의가 지속되는 가운데, 전문직업 영역에서는 사용자인 전문직업인에게 엄격한 검증 의무를 부과하는 추세임
- 의료, 법률 등 전문직업 영역에서 인공지능은 의사결정 지원 도구로 간주되며, 인공지능 결과물을 검증 없이 채택하여 발생한 사고에 대해서는 최종 판단 책임(Last Mile) 원칙에 따라 해당 전문직업인에게 주된 책임을 부과하는 추세임
- 전문가의 검증 의무와는 별도로 인공지능 모델 자체의 설계 결함이나 학습 데이터의 중대한 편향이 사고의 근본 원인으로 작용할 경우, 개발자 또는 운영자 역시 제조물책임이나 과실책임으로부터 자유로울 수 없음
- 인공지능 제공자는 '인공지능은 오류를 범할 수 있다'와 같은 면책 조항을 명시하여 환각 등 통계적 학습 시스템에 내재된 위험에 대한 책임 부담을 완화하고 있으나, 의료 시 챗봇의 자살 유도와 같은 중대한 인명 피해 사고의 경우 일반적인 면책 고지만으로는 책임을 배제하는 데 한계가 있음

3. 보험산업의 영향과 대응

- 인공지능 산출물의 검증 없는 사용은 배상책임보험의 손해를 악화시키고 있으며, 이에 대응하여 AIG 등 주요 보험회사들은 인공지능 관련 리스크를 관리하기 위해 면책특약을 도입하고 있음

- 인공지능 산출물의 무분별한 채택과 사용은 사이버보험뿐만 아니라 전문직업인배상책임보험(Malpractice liability 및 E&O), 임원배상책임보험(D&O), 영업배상책임보험 등의 손해율을 악화시킬 수 있음
 - 이에 대응하여 미국 ISO는 인공지능의 오류로 인한 손해를 면책하는 특약(Generative AI Exclusion)을 개발하였으며, AIG, Great American 등 주요 보험회사들은 해당 특약의 도입을 준비 중임
- 인공지능 시대의 도래는 보험산업의 손해율을 악화시키는 위협으로 작용함과 동시에 새로운 시장 개척의 기회를 제시하고 있음
- 글로벌 6대 보험시장(미·영·일·독·프·중) 기업을 대상으로 한 설문결과 90% 이상의 기업이 인공지능 관련 보험에 관심을 보였으며, 특히 디지털 성숙도가 높은 미국과 중국 기업들의 인공지능 도입률과 보험가입 수요가 모두 높게 나타남(Geneva Association(2025), “Gen AI Risks for Businesses: Exploring the role for insurance”)
 - Deloitte는 인공지능 리스크 관련 보험시장이 향후 연평균 80%의 높은 성장세를 보여, 2032년에는 48억 달러 규모에 이를 것으로 전망하고 있음(Deloitte(2025), “AI insurance could be a \$4.8B market by 2032”)
- 리스크 특성 측면에서 인공지능 관련 위험은 사이버보험과 유사한 특징이 있음
- 인공지능보험과 사이버보험은 보험사고의 동시다발성, 정보 비대칭성, 지속적 기술진보, 무형자산 중심의 보장 등에서 유사한 특징이 있음
 - 반면, 해킹 등 외부 공격을 주로 보장하는 사이버보험과 달리, 인공지능보험은 모델 자체의 결함이나 환각 등 내부 오작동으로 인해 손해가 발생하며, 표준화된 위험 평가지표가 부재하다는 등의 차이가 있음
- 현재 인공지능의 환각 등 관련 리스크를 보장하는 상품은 도입 초기 단계이며, 사이버보험 등 기존 보험상품을 활용한 확장형 보장과 함께 독립형 상품을 포함한 다양한 상품 개발이 진행되고 있음
- Munich Re는 인공지능 제공자와 기업 사용자를 대상으로 aiSure 솔루션을 제공하고 있으며, Relm Insurance의 경우 인공지능 관련 배상책임뿐만 아니라 당사자손해, 기초배상책임보험과 포괄배상책임보험의 보장공백을 담보하는 D.I.C. Policy와 같은 특화된 상품을 개발·판매하고 있음

권순일 연구위원
sikwon@kiri.or.kr

AI Hallucinations and the Insurance Industry

ABSTRACT

Artificial intelligence has become essential to corporate management and national competitiveness, yet the structural limits of generative AI create new risks—such as hallucinations, bias, and infringement. These failures have led to legal disputes, including fabricated case citations, prompting insurers to introduce exclusion clauses in response to rising loss ratios. At the same time, AI risk represents a new growth frontier beyond cyber insurance, with Deloitte projecting the AI insurance market to reach USD 4.8 billion by 2032.

1. The Advent of the AI Era

Artificial intelligence is fundamentally reshaping production processes across industries, accelerating a broad socio-economic paradigm shift through intelligent automation and data-driven decision-making. According to Bain & Company, around 95% of U.S. firms already use generative AI in their operations, while a McKinsey survey finds that 79% of companies have adopted it. In this sense, AI adoption has moved beyond a matter of technological uptake and has become a core determinant of corporate performance and national competitiveness.

2. Key Risks of Artificial Intelligence: Hallucination, Bias, and Infringement

Artificial intelligence has greatly improved efficiency and convenience across industries, the broader economy, and daily life, but it has also introduced new risks,

including hallucination, bias, and infringement. AI-related failures have already triggered legal and social disputes, such as fabricated court citations, copyright infringement, and incidents involving medical AI chatbots, alongside errors by airline chatbots, medical misdiagnoses, algorithmic discrimination, industrial robot malfunctions, and deepfake-enabled cybercrime.

When AI outputs are used without adequate verification, firms face a heightened risk of legal disputes, contributing to rising loss ratios in liability insurance. These risks reflect structural limitations of AI systems—such as imbalanced training data, probabilistic language generation, and opaque “black box” decision-making—which make errors difficult to anticipate and trace.

In response, professional fields such as medicine and law increasingly treat AI as a decision-support tool and impose strict verification duties on human users. Under the “last mile” principle, professionals who rely on AI without sufficient review tend to bear primary responsibility, while developers and operators may also face product liability or negligence claims when design flaws or data bias are the root cause. Broad liability disclaimers, however, offer limited protection in cases involving serious harm or threats to human life.

3. Impact on the Insurance Industry and Policy Responses

The unverified use of AI-generated outputs has worsened loss ratios in liability insurance. In response, major insurers—including AIG—have introduced AI-related exclusion clauses. The indiscriminate adoption of AI may adversely affect loss experience across multiple liability lines, including cyber insurance, professional liability (malpractice and E&O), directors’ and officers’ liability (D&O), and commercial general liability. Accordingly, the U.S. Insurance Services Office has developed a Generative AI Exclusion, with insurers such as Great American Insurance preparing similar endorsements.

At the same time, AI presents new growth opportunities for the insurance industry. A survey across the six largest insurance markets shows that over 90% of firms are interested in AI-related insurance, with particularly strong demand in digitally mature markets such as the United States and China (Geneva Association, 2025). Deloitte (2025) projects that the AI insurance market will grow at an annual rate of around 80%, reaching USD 4.8 billion by 2032.

AI-related risks share similarities with cyber risk, including systemic loss potential, information asymmetry, rapid technological change, and coverage focused on intangible assets. However, unlike cyber insurance, which primarily covers external attacks, AI insurance addresses losses arising from internal failures such as model defects or hallucinations, amid the absence of standardized risk metrics.

While insurance products explicitly covering AI risks remain at an early stage, insurers are developing both extensions of existing policies and standalone products. Munich Re offers the aiSure solution for AI providers and users, while Relm Insurance provides specialized coverage for AI-related liability, first-party losses, and gaps addressed through difference-in-conditions policies

Soon-il Kwon, Research Fellow
sikwon@kiri.or.kr