2026.09.21. 제39호

KiRi 보험법 리뷰

| 포커스 |

「금융분야 인공지능 가이드라인」 개정과 보험산업의 대응 과제

손민숙 연구원

요 약

○ 2026년 6월 「금융분야 인공지능 가이드라인」이 개정 및 시행되고 금융감독원이 이에 맞춰 「금융분야 AI RMF」를 공개하여 금융분야 AI 규율의 패러다임이 '개별 인공지능 서비스의 개발 및 활용에 관한 원칙 제시'에서 '전사적 거버넌스 및 위험관리 체계 구축 요구'로 전환됨에 따라 금융회사들은 인공지능의 도입 여부, 의사결정의 정도, 보험계약자에게 미치는 영향 등을 종합적으로 고려하여 인공지능 위험관리 거버넌스 및 내부통제 체계를 마련해야 하는 상황임

○ 개정된 「금융분야 인공지능 가이드라인」은 '금융 AI 7대 원칙'을 제시함으로써 금융회사가 인공지능의 활용 시 준수하여야 할 기본 방향과 위험관리의 공통 기준을 마련하고, 「금융분야 AI RMF」는 금융회사가 인공지능 관련 위험을 체계적으로 관리할 수 있도록 ① 거버넌스, ② 위험평가, ③ 위험통제로 구성된 위험 기반 접근방법의 종합 평가 체계를 표준안으로 제시함

○ 금융회사들은 「금융분야 인공지능 가이드라인」상 의무와 「인공지능기본법」상 의무가 별개의 것으로 중복 적용된다는 점, 「인공지능기본법」상 고영향 인공지능과 「금융분야 인공지능 가이드라인」상 고위험 인공지능의 구별이 필요하다는 점에 유의할 필요가 있으며, 우선적으로 인공지능을 활용하는 업무가 「인공지능기본법」상 '대출 심사 등' 고영향 인공지능에 해당하는 경우인지 판단이 필요하나 현행 규정상 그 구체적인 범위는 아직 불명확함

○ 현재 보험회사들이 업무별 위험평가와 통제를 수행하기 위해 구축한 거버넌스 체계 형태로는 ① 인공지능 위험관리를 전담하는 의사결정기구와 조직을 별도로 두는 전담형, ② 기존 조직 내에 전담 기능을 두고 관련 통제기능과 연계하는 연계형, ③ 별도 전담 기능 없이 기존 위험관리·내부통제 체계에 편입하는 통합형이 있으며, 개별 보험회사는 인적·물적 자원과 인공지능 활용 범위 및 업무별 위험 수준 등을 고려하여 적합한 방식을 선택할 수 있음

○ 한편 인공지능 위험관리는 모델의 성능 관리나 사전적 통제에 그치지 않고, 소비자에 대한 영향 관리 및 운영 회복탄력성까지 포괄할 필요가 있음

1. 서론

○ 금융위원회는 2026. 6. 18. 기존의 금융분야 인공지능 가이드라인을 통합 개정하여 「금융분야 인공지능 가이드라인」 최종본을 공개[1]하였으며, 2026. 6. 22. 금융감독원의 「금융분야 인공지능 위험관리 프레임워크(이하 '금융분야 AI RMF[2]'라 함)」 및 금융보안원의 「금융분야 인공지능 보안 안내서」가 함께 배포됨[3][4]

- 프런티어 인공지능의 등장, 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 '인공지능기본법'이라 함)」 시행 등 기술 양상 및 규제 환경이 변화함에 따라 기존 가이드라인을 정비할 필요성이 증대됨
- 개정된 「금융분야 인공지능 가이드라인」은 미토스 대응, 금융권 망분리에 따른 보안 강화 등 최근 이슈를 추가 반영한 최종안으로 기존에 시행 중이던 3개의 가이드라인[5]을 통합한 것임

○ 「금융분야 인공지능 가이드라인」의 개정에 따라 금융 관련 인공지능 규율의 패러다임이 '개별 인공지능 서비스의 개발 및 활용 등에 관한 원칙 제시'에서[6] '전사적 거버넌스, 위험평가, 위험 등급별 통제 및 사후 모니터링을 포괄하는 위험관리 체계 구축 요구'로 전환됨

- 기존에도 금융업권이 3대 가이드라인을 기준으로 인공지능 내부통제 시스템을 구축해 왔다는 점을 고려할 때, 개정된 「금융분야 인공지능 가이드라인」의 실질적인 영향력은 상당할 것으로 예상되며, 자율규제로 운영됨에 따라 직접적인 강제력은 없으나 금융감독원 검사 및 현장 조사 시 내부통제 적정성 평가의 판단기준으로 활용될 것으로 예측됨[7]
- 보험업에서는 인공지능이 보험상품의 추천·설명, 보험인수 및 보험료 산정, 보험금 심사·지급, 보험사기 탐지, 계약관리 등 다양한 업무에 활용될 수 있고, 각 업무의 성격과 인공지능의 활용 방식에 따라 보험계약자의 권리·의무 및 경제적 이해관계에 미치는 영향의 정도가 달라질 수 있음
- 따라서 인공지능 위험관리는 인공지능의 도입 여부나 모델의 성능 관리에 그치지 않고, 인공지능이 임직원의 판단 및 의사결정에 관여하는 정도와 그 결과가 보험계약자에게 미치는 영향의 정도를 함께 고려하여 통제 수준을 차등화할 필요가 있음

---

1) 금융위원회 보도자료(2026. 6. 18.), "정부와 금융권이 함께 금융권 AX를 가속화해 나가겠습니다. - 책임 있는 혁신, 금융권 AX 본격 추진을 위한 한 걸음을 내딛는다"
2) AI Risk Management Framework in Financial Sector
3) 금융권 AI 플랫폼 공지사항(2026. 6. 18.), "금융분야 인공지능 가이드라인 개정 시행 안내 및 안내데스크 운영 일정 공지"
4) 정부(관계부처 합동)는 이와 별도로 2026. 8. 21. 「대한민국 인공지능 윤리원칙」을 확정하고, 과학기술정보통신부가 2026. 8. 24. 이를 공표하여 '이용자'인 기업 또한 책임 있는 인공지능의 활용 주체로 명시하였는데 본고에서는 해당 내용을 다루지 않으며, 「금융분야 인공지능 보안 안내서」의 구체적인 내용도 제외함
5) 금융분야 AI 운영 가이드라인(2021. 7.), 금융분야 AI 개발·활용 안내서(2022. 8.), 금융분야 AI 보안 가이드라인(2023. 4.)
6) 백영화(2021. 8. 23.), 「금융분야 인공지능 가이드라인의 주요 내용」, 『KIRI 보험법리뷰』, 제12호, 보험연구원
7) 법률신문(2026. 7. 7.), "금융분야 AI 통합 가이드라인, RMF 및 보안안내서의 주요내용과 시사점"

2. 「금융분야 인공지능 가이드라인」과 「금융분야 AI RMF」의 주요 내용

가. 금융분야 인공지능 가이드라인

○ 「금융분야 인공지능 가이드라인」은 금융분야에서 인공지능을 활용할 때 준수하여야 할 '7대 원칙'을 마련하고, 이를 통해 금융회사가 인공지능의 기획·개발·도입·운영에 이르는 전 과정에서 준수하여야 할 책임 있는 인공지능 활용의 기본 방향과 위험관리의 공통 기준을 제시하고 있음

- 금융 AI 7대 원칙은 ① 거버넌스 원칙, ② 합법성 원칙, ③ 보조수단성 원칙, ④ 신뢰성 원칙, ⑤ 금융 안정성 원칙, ⑥ 신의성실 원칙, ⑦ 보안성 원칙으로 주요 내용은 〈표 1〉과 같음
- 「금융분야 인공지능 가이드라인」은 원칙별로 참고가 될 수 있는 예시와 준수 사례, 체크리스트 등을 제공함으로써 금융회사가 각 원칙을 실제 업무와 인공지능 활용 과정에 구체적으로 적용할 수 있도록 실무적인 판단기준을 제시함

〈표 1〉 금융 AI 7대 원칙의 주요 내용

원칙	주요 내용
① 거버넌스 원칙	• 의사결정기구(예: AI 윤리위원회 등) 및 위험관리 전담 조직의 구성 • 위험관리 규정 및 지침 등 내부 규정 및 업무 매뉴얼 마련 • 위험 인식·측정, 위험 경감, 잔여 위험평가, 위험 등급 산정 등 종합적인 위험평가 체계 구축 → 금융분야 AI RMF로 구체화 • 위험 수준별 차등화된 통제 절차 마련 및 이행
② 합법성 원칙	• 인공지능기본법 등 적용되는 법규 및 개별법 이슈 사전 파악·검토 - 예: 대출 심사·신용 평가 → 신용정보법, 이상 거래 탐지 → 특정금융거래정보법, 투자 권유·자문 서비스 → 자본시장법·금융소비자보호법 - 외부 위탁 AI도 법령 준수 여부를 직접 확인 - 해외 서비스 제공 시 국외 규제 적용 가능성 검토 • 내부 규정·절차 마련 및 주기적인 점검·개선 현행화
③ 보조수단성 원칙	• 최종 책임을 해당 금융회사 등의 임직원이 수행할 수 있도록 책임 수행 체계 구축 • 인적 개입 원칙의 적용 및 운영 - AI 운영 전 단계에 걸쳐 임직원의 개입이 필요한 상황을 차등화 - 고위험 AI의 경우 의사결정에 보조적인 용도로만 사용 가능(Human-in-the-Loop) - 초고성능 AI(예: 미토스 시스템) 등 고위험 AI가 최종 의사결정을 하지 않도록 설계·운영 → 긴급정지 기능(Kill switch) 등의 수단으로 즉시 개입 • 업무 담당자 및 감독자 등을 대상으로 정기적인 교육 실시 → 자동화 편향 방지
④ 신뢰성 원칙	• 모델 성능 관리(생성형 AI의 환각 현상 완화 등) • 데이터 품질 확보, 공정성·편향성 점검, 설명가능성 확보
⑤ 금융 안정성 원칙	• 금융안정 평가·관리 - 금융안정위원회(FSB)의 AI 취약점을 참고하여 금융안정 위험 요인 고려 • 백업 모형·긴급 정지 기능 등 안전장치 마련 • 제3자 IT 리스크관리 • 감독당국 정보 공유 및 보고 의무

〈표 1〉 계속

원칙	주요 내용
⑥ 신의성실 원칙	• 로보어드바이저, 금융상품 비교·추천 등에서의 이해상충 방지 – 금융소비자보호법 제10조 및 동법 시행령 제5조 제4항 • AI 활용 사실을 일반 소비자에게 사전 고지, 피해 대응 절차 마련을 통해 소비자보호
⑦ 보안성 원칙	• AI 특화 보안 위협(데이터 오염, 프롬프트 인젝션, 탈옥 공격 등) 식별 및 관리 • AI 특화 공격 탐지 및 대응 • AI 자산(데이터, 모델 파라미터 등) 보호 및 관리 • 외부 모델 및 데이터 보안·신뢰성 검증 • 기존 IT 보안 체계의 AI 확장 적용 → 금융분야 인공지능 보안 실무 안내서(금융보안원)

자료: 「금융분야 인공지능 가이드라인」을 토대로 저자 작성함

나. 금융분야 AI 위험관리 프레임워크(금융분야 AI RMF)

○ 금융감독원이 「금융분야 인공지능 가이드라인」의 개정에 맞춰 마련한 「금융분야 AI RMF」는 금융 AI 7대 원칙 중 거버넌스 원칙을 구체화한 것으로 금융회사가 인공지능 관련 위험을 체계적으로 관리할 수 있도록 위험 기반 접근방법의 종합 평가 체계를 표준안으로 제시하고 있으며, ① 거버넌스, ② 위험평가, ③ 위험 통제 세 가지 영역으로 구성됨

- 최고 의사결정기구 및 위험관리 전담 조직은 ① 금융 AI 7대 원칙에 대해 위험평가 항목 및 기준 등을 사전에 정의하고 ② 인식·측정한 위험 항목별로 경감계획을 수립·이행하고 각각의 잔여 위험을 확인한 후 ③ 항목별 점검 등을 통해 잔여 위험의 적정성을 최종 평가하여 ④ 최종 위험 등급을 결정하여야 함
- 금융 AI 7대 원칙을 기준으로 ① 거버넌스 원칙, ③ 보조수단성 원칙, ⑤ 금융 안정성 원칙은 정성적 평가 항목으로, ② 합법성 원칙, ④ 신뢰성 원칙, ⑥ 신의성실 원칙, ⑦ 보안성 원칙은 정량적 평가 항목으로 활용하여 점수화함
  - 위험 분류 기준은 회사의 위험 선호도 등을 고려하여 결정하되 저위험(25점 미만), 중위험(25점 이상 50점 미만), 고위험(50점 이상) 등으로 구분하며, 「인공지능기본법」상 '고영향 인공지능'에 해당하는 경우 위험 점수에 따른 등급 분류와 관계없이 '고위험'으로 자동 분류하여 통제를 강화하여야 함
  - 75점 이상인 경우 및 금융안정성을 훼손할 우려가 있는 경우, 금융회사 및 금융소비자에게 중대한 위협을 미칠 우려가 있는 경우에는 인공지능 서비스의 출시 여부를 재검토하는 등 보다 강화된 조치를 취할 수 있음

〈표 2〉 금융분야 AI RMF의 주요 내용

영역	주요 내용
① 거버넌스	• AI 위험관리 등을 위한 의사결정기구(예: AI 윤리위원회, AI 위험관리위원회) 설치(인공지능기본법 제28조) → AI 관련 내규 제·개정, 위험관리·소비자보호 정책 수립, 고위험 AI 서비스 승인 등 AI 관련 중요사항 심의·의결 • AI 기획·개발 추진 조직과 독립된 위험관리 전담 조직 설치 → AI 관련 업무 전반 통제·관리 • AI 위험관리 규정 등 내규 마련 및 AI 관련 업무 전반 규정 → 위험의 종류, 위험 인식·측정·평가·통제 방안, 허용 가능 위험 수준, 고위험 AI 통제 방안 등 명시 • 책무구조도를 도입한 경우, AI 관련 내부통제 및 위험관리 등에 관한 책무를 반영할 것을 고려
② 위험평가	• 위험 기반 접근방법(Risk-based approach)에 따른 4단계 절차 마련 : 위험 인식·측정 → 위험 경감 계획 수립·이행 → 잔여 위험평가 → 위험 등급 산정 • 원칙별 정량 평가 항목 배점 예시: 합법성(20%), 신뢰성(30%), 신의성실(20%), 보안성(30%) → 16개의 세부 항목으로 점수화 • 위험 등급 산정: 25점 미만(저위험), 25점 이상 50점 미만(중위험), 50점 이상(고위험), 75점 이상(초고위험, 서비스 출시 여부 재검토) • 인공지능기본법상 고영향 AI: 위험 점수와 무관하게 고위험 등급 분류
③ 위험 통제	• 위험 수준별 관리 차등화 - 저위험: 승인 절차 간소화, 작성 문서 축소 등 완화된 통제 및 관리 - 고위험: 최고 AI 의사결정기구의 사전 승인·사후 검증, 제3자 평가 검증, 운영 모니터링 강화 등 추가 통제 및 관리 • 고위험 AI 중 금융 안정성 훼손, 금융회사 및 개인에 중대한 위협 가능성이 있는 경우 → 서비스 출시 여부 재검토 • 모니터링·사후 관리, 문서화 및 교육, 감독당국 정보 공유 및 시스템 위험관리 → AI 모형 오작동 대비 백업 모형, 비상 정지(kill switch) 장치, 회생 계획 등 마련

자료: 「금융분야 AI RMF」를 토대로 저자 작성함

○ **금융회사가 위험 기반 접근방법을 실질적으로 구현하기 위해서는 개별 인공지능 모델의 특성만을 기준으로 하기보다 인공지능이 활용되는 업무와 의사결정의 성격 및 그에 수반되는 위험을 파악하고, 이에 상응하는 통제 수준을 정할 필요가 있음**

- 보험회사는 인공지능 활용 업무를 목록화하고, 각 업무에 대하여 ① 보험 가입·보험료·보험금 지급 등 보험계약자의 권리·의무 및 경제적 이해관계에 미치는 영향, ② 인공지능이 최종 의사결정에 관여하는 정도, ③ 잘못된 판단이나 결과가 발생한 경우 이를 사후적으로 시정·회복할 수 있는 정도, ④ 임직원이 인공지능의 판단을 실질적으로 검토·변경할 수 있는 정도, ⑤ 외부 인공지능 모델·데이터·클라우드 등에 대한 의존도 등을 종합적으로 평가할 필요가 있음
  - 예를 들어, 내부 문서의 검색·요약에 활용되는 인공지능과 보험인수 또는 보험금 지급 판단에 활용되는 인공지능은 동일한 기술이나 모델을 사용하더라도 보험소비자에게 미치는 영향과 인공지능의 의사결정 관여 정도가 다르므로, 동일한 수준의 통제를 적용하는 것은 위험 기반 접근방법의 취지에 부합하지 않을 수 있음
- 이러한 평가 결과에 따라 저위험 업무에 대해서는 승인·검증·문서화 등 관리 절차를 합리적으로 간소화하는 한편, 보험인수·보험료 산정·보험금 지급 등 보험계약자의 권익에 중대한 영향을 미칠 수 있는 업무에 대해서는 독립적인 검증, 실질적인 인적 개입, 지속적인 사후 모니터링, 이의제기 및 재심 절차 등 통제 수준을 강화할 필요가 있음

3. 「금융분야 인공지능 가이드라인」과 「인공지능기본법」의 관계

가. 「금융분야 인공지능 가이드라인」과 「인공지능기본법」의 중복 적용

○ 「금융분야 인공지능 가이드라인」은 금융 관련 인공지능 활용 전반에 폭넓게 적용되며, 「금융분야 AI RMF」와 함께 자율규제로서 각 금융회사들이 업무 특성과 위험 수준에 따라 자율적으로 적용 범위와 수준을 정할 수 있음

- 「금융분야 인공지능 가이드라인」은 업종 및 업무에 관계없이 인공지능을 도입·활용하는 모든 금융회사 및 금융거래에 관여하는 비금융회사들이 적용 대상임
  - 핀테크 기업 등 비금융회사의 경우에도 인공지능 활용 결과가 금융거래 제공에 영향을 줄 수 있는 경우 가이드라인의 적용 대상이 된다는 점에서 「인공지능기본법」보다 적용 대상이 포괄적이며, 적용되는 업무의 범위 또한 넓음[8)]
- 「금융분야 AI RMF」 역시 「인공지능기본법」 제32조 제1항 제2호에 따른 위험관리 체계 구축 의무와는 별개의 제도로 「금융분야 인공지능 가이드라인」과 같이 법적 강제력이 없는 자발적 지침임
  - 「인공지능기본법」 제32조 제1항의 경우 일정 기준 이상인 인공지능시스템의 안정성을 확보하기 위한 것으로서 인공지능 관련 안전사고 모니터링 및 대응이 목적이며, 「인공지능기본법」 제40조 제3항에 따른 중지명령 또는 시정명령 대상인 반면 「금융분야 AI RMF」는 참고 목적의 프레임워크로 자율규제라는 점에 차이가 있음
- 개별 금융회사는 보유한 인적·물적 자원, 업무의 특성, 인공지능 활용 범위 및 서비스 위험 수준 등을 종합적으로 고려하여 자율적으로 가이드라인 적용 범위와 수준을 결정할 수 있음

○ 다만 「인공지능기본법」상 '고영향 인공지능' 및 시행령에서 규율하는 내용에 해당하는 경우 법률 및 시행령상의 의무가 면책되지 않으며, 별도의 법적 의무가 발생하므로 규제가 중복 적용된다는 점에 유의할 필요가 있음

- 인공지능 기본 법령과 「금융분야 인공지능 가이드라인」의 규율 내용이 상충되는 경우 인공지능 기본 법령 및 관련 가이드라인이 우선 적용되며, 인공지능 기본 법령이 적용되지 않거나 규율하지 않는 경우 「금융분야 인공지능 가이드라인」이 적용됨[9)]
- 별도로 '고영향 인공지능'에 해당하는 경우 「인공지능기본법」상 사업자로서의 책무를 이행함과 동시에 추가적으로 거버넌스 원칙 및 금융 안정성 원칙 등 「금융분야 인공지능 가이드라인」의 해당 내용을 적용할 수 있음
  - 「금융분야 AI RMF」 또한 해당 프레임워크를 따르더라도 「인공지능기본법」상 의무가 면책되지 않음
- 「금융분야 인공지능 가이드라인」과 「인공지능기본법」이 중복 적용됨에 따라 금융회사는 동일한 인공지능 활용에 대하여 각 규율체계의 적용 여부와 그에 따른 준수사항을 함께 검토해야 하며, 특히 고영향 인공지능과 고위험 인공지능의 분류 기준과 그에 따른 의무가 서로 다른 데 따른 적용상 불확실성과 규제 준수 부담이 발생할 수 있음[10)]
  - 따라서 규제 부담의 경감은 동일한 사실관계와 위험평가 결과를 관련 법령과 가이드라인 내에서 공통으로 활용할 수 있도록 분류 체계 및 평가·증빙 절차의 정합성을 높이는 데 초점을 둘 필요가 있음

---

8) 대출심사, 신용평가, 챗봇, 금융상품의 비교·추천 등 고객에게 직접적인 영향을 미치는 경우뿐만 아니라 금융회사 내부의 지원 및 관리를 위하여 인공지능시스템을 활용하는 경우에도 적용됨
9) 「금융분야 인공지능 가이드라인」, p. 5
10) 다만 현재 정부의 입장은 중복 규제의 부담 및 인공지능 산업의 진흥, 현실적인 여건 등을 고려해 '고영향 인공지능'의 범위를 최소화하여 제한적으로 해석하고자 함; 대한민국 정책브리핑(2025. 11. 12.), ""AI로 생성된 결과물" 고지해야…AI기본법 시행령 입법예고"

- 금융회사 역시 「인공지능기본법」상 고영향 인공지능과 「금융분야 AI RMF」상 고위험 인공지능을 각각 별도의 체계로 관리하기보다, 하나의 인공지능 활용 업무목록(Inventory)과 공통된 위험평가 체계를 기초로 각 규범의 적용 여부와 그에 따른 추가 의무를 식별·관리하는 방식이 효율적일 수 있음

나. 「인공지능기본법」상 고영향 인공지능과 「금융분야 인공지능 가이드라인」상 고위험 인공지능

○ **「인공지능기본법」에 따라 '고영향 인공지능' 또는 이를 이용한 제품 및 서비스를 제공하는 경우 인공지능사업자는 각종 규제를 준수할 의무를 부담하며, 업무에 활용 중인 인공지능이 '고영향 인공지능'에 해당하는 경우 엄격한 규제가 적용됨에 따라 금융회사들은 해당 업무에 활용되는 인공지능이 '고영향 인공지능'에 해당하는지 선제적인 판단이 필요함**

- 「인공지능기본법」은 '고영향 인공지능'을 "사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템으로서 다음 각 목의 어느 하나의 영역에서 활용되는 것"으로 정의함(제2조 제4호)
- 인공지능사업자는 인공지능 또는 이를 이용한 제품·서비스를 제공하는 경우 그 인공지능이 고영향 인공지능에 해당하는지 여부를 사전에 검토하여야 하고(제33조), 고영향 인공지능을 이용한 제품 또는 서비스를 제공하려는 경우 투명성 확보 의무가 부과되며(제31조), 위험관리 방안의 수립·운영 및 문서 작성 및 보관 의무 등 사업자의 책무를 부담할 뿐만 아니라(제34조) 사람의 기본권에 미치는 영향을 평가하기 위하여 노력하여야 함(제35조)
- 「고영향 인공지능 판단 가이드라인」[11]에 따르면 ① 특정 영역에서 활용되고(1단계 조건)[12], ② 사람의 생명, 신체의 안전 및 기본권에 중대한 영향을 미치거나 위험을 초래할 우려가 있는 인공지능시스템에 해당하는 경우(2단계 조건)[13]에 모두 해당되는 경우에만 고영향 인공지능으로 판단할 수 있음

○ **금융분야와 관련해서는 "대출 심사 등 개인의 권리 · 의무 관계에 중대한 영향을 미치는 판단 또는 평가"가 '고영향 인공지능'에 해당할 수 있음이 명시되어 있으며(제2조 제4호 사목), 보험 관련 업무 역시 그 성격에 따라 적용 대상에 포함될 여지가 있으나 현행 규정상 그 구체적인 범위는 명확하지 않음[14]**

- 금융분야에서 '고영향 인공지능'은 ① 금융업권에서 금융거래나 대고객서비스에 인공지능시스템을 활용한 경우, 또는 ② 비금융업권이라도 인공지능시스템의 활용 결과가 금융거래에 영향을 미치는 경우 적용됨[15]
  - 금융회사가 신용정보 등에 근거하여 대출 신청인의 신용 및 상환능력을 평가하여 대출의 승인 여부와 대출 조건을

---

11) 과학기술정보통신부·한국지능정보사회진흥원(2026. 1.), 「고영향 인공지능 판단 가이드라인」

12) 1단계 조건인 '특정 영역'은 에너지, 먹는 물, 보건의료, 공공서비스, 교육 등 10개의 영역으로 「인공지능기본법」 제2조 제4호 각 목에 해당함. 동조 동호 카목은 "그 밖에 사람의 생명·신체의 안전 및 기본권 보호에 중대한 영향을 미치는 영역으로서 대통령령으로 정하는 영역"이라고 하여 하위 법령의 위임을 통한 추가적인 영역을 규정하고 있으나, 시행령은 아직 새로운 영역을 규정하고 있지 않음

13) 2단계 조건 중 '기본권에 대한 위험성' 판단 시 인공지능시스템의 의도된 목적, 기능, 활용 맥락을 다양한 관점에서 고려하며, '위험의 중대성' 판단 시 권익의 속성과 증가된 위험 정도를 종합적으로 고려함. 금융과 관련하여 '대출 심사 등'은 금융소비자의 주거 및 생활 안정 등에 필요한 자금의 원활한 공급을 위해 대출 심사 과정에서 편향이나 차별 등이 발생하지 않도록 함으로써 금융소비자의 평등권, 재산권, 인간다운 생활을 보장할 권리와 관련이 있고(기본권에 대한 위험성), 금융회사의 대출 심사 등 신용공여 행위 관련 업무에 사용되는 인공지능시스템의 의도하지 않은 작동 등으로 인해 금융소비자의 금융거래·계약 체결·유지 등에 중대한 영향을 미칠 우려가 있음(위험의 중대성)

14) 비교법적으로 EU AI Act의 경우, 생명보험 및 건강보험에서 자연인에 대한 위험평가 및 가격 책정을 목적으로 하는 인공지능시스템을 Annex III상 '고위험'으로 분류함(Article 6(2), Annex III 5(c)). 이는 이러한 인공지능시스템이 사람의 생계에 중대한 영향을 미칠 수 있으며, 정당하게 설계·개발·이용되지 않을 경우 금융 배제 및 차별을 초래하는 등 사람의 생명과 건강에 심각한 결과를 초래할 수 있다는 점에 근거함(Recital 58). 다만 Annex III에 해당하는 인공지능시스템이라 하더라도 의사결정의 결과에 실질적인 영향을 미치지 않거나 사람의 건강, 안전 또는 기본권을 해할 중대한 위험을 초래하지 아니하는 등 Article 6(3)의 요건을 충족하는 경우에는 고위험 인공지능에서 제외될 수 있음

15) 금융상품 및 금융서비스의 제공을 위한 업무에 인공지능시스템을 직·간접적으로 활용하거나 활용하고자 하는 금융회사 및 금융상품추천·신용평가 등 금융 연관 서비스 제공을 위한 업무에 인공지능시스템을 직·간접적으로 활용하거나 활용하고자 하는 비금융회사가 여기에 해당함

결정하는 데 직접적으로 관련되는 업무에 한하며, 대출 과정에서의 모든 업무가 '대출 심사'에 포함되는 것은 아님[16]

- 의사결정 과정에서 인공지능 산출 결과를 단순히 보조적·참고적 수단으로 활용한 경우, 즉 최종 의사결정 과정에 실질적인 인적 개입이 있는 경우 '고영향 인공지능'에서 제외될 수 있으며, 형식적으로 인간이 개입하더라도 인공지능의 산출물에 의존하거나 이를 의사결정의 중요한 근거로 활용하는 경우에는 고영향 인공지능에 해당할 수 있음
- 「고영향 인공지능 판단 가이드라인」은 '대출 심사 등'에서 '고영향'의 의미 및 확인 기준을 제시하고 있으나 판단 시 개별 요소 및 맥락적 환경을 종합적으로 고려하도록 하고 있어 금융회사가 이를 활용하고자 할 때 여전히 불명확성이 존재하며,[17] 구체적으로 어떤 유형의 서비스를 '대출 심사 등'으로 포함할 것인지에 대해서는 아직 확정된 바 없음[18]

○ **주의할 것은 「인공지능기본법」상 '고영향 인공지능'과 「금융분야 인공지능 가이드라인」상 '고위험 인공지능'은 포괄하는 범위가 다르므로, 각각의 해당 여부를 구분하여 판단하고 그에 따른 규정을 적용할 필요가 있다는 점임**

- 「금융분야 인공지능 가이드라인」은 「인공지능기본법」상 '고영향 인공지능'의 개념을 수용함과 동시에 「금융분야 AI RMF」의 위험평가 체계를 통해 고위험 서비스로 분류되는 인공지능시스템인 '고위험 인공지능'을 더 포괄적으로 넓게 정의하고 있음
- 즉, 「인공지능기본법」상 '고영향 인공지능'으로 판별되면 「금융분야 인공지능 가이드라인」상 '고위험 인공지능'에 해당되나, '고위험 인공지능'이 반드시 '고영향 인공지능'에 해당되는 것은 아님

4. 보험산업 대응 과제

○ **「금융분야 AI RMF」에 따라 금융회사의 인공지능 위험관리 및 내부통제 체계 구축이 요구되는 가운데, 국내 보험회사의 인공지능 거버넌스 구축은 아직 초기 단계에 있는 것으로 보임**

- 금융감독원 조사에 따르면 2025년 4월 기준 금융회사 118곳 중 은행 5개사(25%), 보험회사 4개사(7.5%), 증권사 1개사(2.7%)만이 인공지능 거버넌스 관련 의사결정기구를 설치하였으며,[19] 일부 금융그룹이나 금융회사가 관련 체계를 구축하고 있으나 최근 전수 조사 자료는 확인되지 않음
- 또한 지속가능경영보고서를 발간한 국내 보험회사 중 일부가 인공지능 거버넌스 관련 내용을 공시하고 있으나,[20] 의사결정기구의 설치나 공시 여부만으로 실제 위험관리 수준을 판단하기에는 한계가 있음
- 보험회사의 대응은 인공지능이 활용되는 업무와 그 과정에서 인공지능이 수행하는 판단·의사결정 기능을 파악하고, 보험계약자의 권익에 미치는 영향, 인공지능의 의사결정 관여 정도, 잘못된 판단의 수정 가능성, 인적 개입의 실효성

16) 즉, 신용공여 행위와 직접적인 관련이 적은 챗봇 서비스, 이상거래탐지서비스(FDS) 및 인공지능시스템의 활용으로 고객에게 미치는 영향이 없는 금융회사 내부 직원 관리, 단순 업무 효율화 등은 '대출 심사'에 포함되지 않음

17) 백연주(2025. 10. 25.), 「인공지능기본법 하위법령(안)의 금융분야 시사점과 개선방향」, 한국금융연구원 금융브리프 논단; 저자는 "대출심사 고영향 판단기준 가이드라인에 대해 ① 기준의 모호성 및 규제범위 확정의 어려움, ② 과기부의 재량적 판단 여지가 크다는 점, ③ 전문위원회 구성 한계로 인한 금융당국과의 조율 부재 가능성, ④ 절차 복잡성 등의 우려가 제기되고 있다"고 설명함

18) 서울경제(2026. 3. 18.), "금융권 AI 의무이행 준비 부족해…CCO가 통제 적극 참여해야"

19) 디지털타임스(2026. 1. 15.), "리스크 커지는 데…금융사 85% AI 위험관리 미흡 "금감원 손본다""

20) 생명보험협회 및 손해보험협회의 정회원사를 기준으로 전자공시시스템(DART)(2026. 8. 3. 기준)을 살펴본 결과 생명보험회사 20개사, 손해보험회사 17개사 중 지속가능경영보고서를 발간하여 자율공시를 실시하고 있는 곳은 생명보험회사 3개사, 손해보험회사 3개사로 총 6개사이며, 그중 인공지능 위험관리 거버넌스와 관련한 내용을 공시하고 있는 곳은 생명보험회사 2개사, 손해보험회사 1개사임

및 외부 모델·데이터에 대한 의존도 등을 종합적으로 고려하여 업무별 위험에 상응하도록 통제 수준을 차등화하는 데서 출발할 필요가 있음

○ 미국과 유럽의 해외 주요 보험회사[21]들을 살펴본 결과, 현재 해외 보험회사들이 업무별 위험평가와 통제를 수행하기 위해 구축한 거버넌스 체계 형태는 ① 인공지능 위험관리를 전담하는 의사결정기구와 조직을 별도로 두는 전담형, ② 기존 조직 내에 전담 기능을 두고 관련 통제기능과 연계하는 연계형, ③ 별도 전담 기능 없이 기존 위험관리·내부통제 체계에 편입하는 통합형으로 유형화할 수 있으며, 이는 국내 보험회사의 인공지능 거버넌스 체계 구축 시 다양한 방식이 고려될 수 있음을 시사함(자세한 내용은 〈표 3〉 참고)

- 국내 보험회사들은 인적·물적 자원, 인공지능 활용 범위 및 업무별 위험 수준 등을 고려하여 적합한 방식을 선택하되, 특정한 조직 형태를 추구하기보다는 인공지능의 기획·개발·활용과 검증·통제 간 상호 견제와 독립성을 확보하고, 관련 조직 및 임직원의 역할과 책임을 명확히 하는 것이 중요함
- 특히 보험인수, 보험료 산정, 보험금 심사·지급 등 보험계약자의 권익에 중대한 영향을 미칠 수 있는 업무에 대해서는 독립적인 검증, 실질적인 인적 개입 및 지속적인 사후 모니터링 등을 강화하는 한편, 내부 업무지원 등 상대적으로 위험이 낮은 업무에 대해서는 승인·검증·문서화 등 관리 절차를 합리적으로 간소화하는 등 업무별 위험 수준에 상응하도록 통제 수준을 차등화할 필요가 있음

〈표 3〉 해외 주요 보험회사별 인공지능 위험관리 거버넌스 체계[1)]

유형[1)]	회사명	조직 및 기구 등	거버넌스 관련 주요 내용	기존 체계와의 연계 방식
전담형	AIG (미국)	• AI 자문위원회	• AIG Global AI Policy • AI Advisory Council – Steering Group + Working Group	• AI 자문위원회에 법무·기술 등 참여 • 이사회: 인공지능 전략 및 관련 위험 직접 감독
연계형	Allianz (유럽)	• DAITAB • AI Trust Officer	• Group Data and AI Trust Advisory Board(DAITAB) • AI Trust Officer – Group AI Trust Officer + local AI Trust Officer + Group AI Office • Responsible AI Corporate Rules	• 기존 개인정보보호 위험평가 체계를 RAI (Responsible AI) 영역으로 확장 • 데이터 거버넌스와 AI 거버넌스를 연계 → Privacy Impact Assessment 및 AI Risk Assessment 등 위험평가 절차 운영
	AXA (유럽)	• Group AI Governance Forum • Group GenAI Model Advisory body	• Group AI Governance Forum • Group GenAI Model Advisory body • 7대 RAI 원칙 → 위험 기반 AI 관리 체계 운영 • AI Risk Library • Fairness Compass	• AI 전담 거버넌스 체계+기존 위험관리·준법·내부통제 기능 • 그룹 최고위험관리책임자와 그룹 최고기술·인공지능책임자 공동 주관 위험위원회 → 운영·정보·준법 위험 및 내부통제를 기존 전사적 위험관리체계와 연계
	Zurich (유럽)	• Global Responsible AI Champions Network • AI Assessment Framework	• Global Responsible AI Champions Network • AI 평가 체계를 통해 안전성·투명성·책임성·신뢰성 원칙을 구체화, 모델·제3자·데이터 위험 등에 관한 통제 요건 마련 • 그룹 RAI 총괄 및 글로벌 RAI 챔피언 네트워크를 운영	• AI 관련 위험을 그룹 전사적 위험관리 체계(ERM)의 일부로 관리 • 위험 선호도 및 위험 허용도와 연계하여 통제 • 개인정보·데이터 보호 등 기존 위험관리 체계와 결합하여 운영
	Munich Re (유럽)	• AI Governance and Strategy section, DAA1.3.2	• AI Governance Framework + AI Guiding Principles • AI 거버넌스·전략 전담조직을 제1선 기능으로 운영 • 별도의 AI 거버넌스 지침과 RAI 활용 원칙 → AI 개발·배포 및 활용 관리	• AI의 고유한 위험을 비재무위험 분류 체계에 포함 • AI 개발·배포에 관한 최소 통제요건 → 기존 비재무위험관리 체계에 편입

21) 보험료 및 보험수익, 사업 규모, 총자산, 국제적 기업 순위, 글로벌 영업 범위 등을 고려하여 유럽의 경우 Allianz, AXA, Zurich Insurance Group, Generali, Munich Re를, 미국의 경우 MetLife, Prudential Financial, AIG, Allstate, Travelers를 살펴봄. 해외 주요 보험회사들도 특정한 인공지능 거버넌스 체계를 일률적으로 채택하기보다는 각 회사의 조직 구조와 위험관리 체계를 바탕으로 다양한 거버넌스 방식을 모색하고 있으며, 인공지능의 활용 수준과 위험 환경의 변화에 따라 전담기구를 신설하거나 기존 위험관리 체계와의 연계를 강화하는 등 거버넌스 체계를 지속적으로 발전시키고 있음

〈표 3〉 계속

유형[1]	회사명	조직 및 기구 등	거버넌스 관련 주요 내용	기존 체계와의 연계 방식
연계형	Prudential Financial (미국)	• Global AI Oversight Council	• Ethical Principles for Artificial Intelligence 및 AI Product Policy • 글로벌 AI 감독위원회가 AI 거버넌스 설정, 전사적인 AI 제품의 도입·활용을 모니터링	• 글로벌 AI 감독위원회를 기존 위험·준법 거버넌스 체계의 일부로 운영 • AI 윤리 통제를 기존 업무 관행 및 프로세스에 내재화(AI Ethics by Design)
통합형	Generali (유럽)	공개 명시 없음[3]	• Trustworthy AI Principles • Group Guideline on Artificial Intelligence Governance 제정·시행	• AI 위험을 기존 운영 위험 및 정보통신기술(ICT) 위험관리 체계에 편입하여 평가·관리
	Allstate (미국)	• Risk and Return Committee • ERRM • Audit Committee	• AI Governance Program • 이사회가 AI 전략·거버넌스를 감독하고 위험·수익위원회가 AI 위험·기회·거버넌스를 감독	• AI 거버넌스 프로그램을 기존 전사 위험·수익관리체계(ERRM), 사이버 복원력 체계 및 전사 개인정보보호 프로그램과 결합하여 운영
	Travelers (미국)	• 이사회 • 위험위원회 • 전사적 위험관리	• Responsible Artificial Intelligence Framework • AI·고급분석·모델의 개발 및 활용에 인간의 감독·판단, 개인정보보호·보안, 공정성·비차별, 책임성, 투명성·설명가능성 등이 핵심 원칙 • 이사회가 AI 전략 및 AI 거버넌스 감독	• RAI 체계를 기존 기업 거버넌스의 일부로 운영 • 전사적 위험관리, 언더라이팅·요율 및 모델 거버넌스, 데이터 거버넌스 등 기존 위험관리·통제체계와 연계

주: 1) 각 회사별 홈페이지, 연차보고서, 지속가능경영보고서, 보도자료 등 공개된 자료를 종합하여 유형을 분류하였으나, 세부적인 내부 거버넌스 체계에 대한 대외 공시가 제한적이어서 명확한 구분에는 한계가 존재함
2) 10개사 중 미국 MetLife의 경우 AI의 설계·개발·배포·사용 전 과정에 대한 법률적·전략적 감독을 위한 거버넌스 체계를 운영 중이며, 이사회가 책임 있는 AI 거버넌스 및 위험관리체계를 감독하고 있음을 확인하였으나, 공식자료만으로는 유형의 구분이 명확하지 않아 제외함
3) '공개 명시 없음'으로 표시된 경우라 하더라도 내부적으로 실무 협의체가 있을 가능성을 배제할 수 없음

자료: AIG(2026), "2026 Notice of Annual Meeting and Proxy Statement"; Allianz 홈페이지; AXA 홈페이지; AXA(2026), "2025 Annual Report"; Zurich 홈페이지; Zurich(2026), "Annual Report 2025"; Munich Re 홈페이지; Munich Re(2026), "Annual Report 2025"; Munich Re(2026), "Excerpt/public version of: Non-Financial Risk Management Framework including Information Security Management"; Prudential(2026), "2026 Prudential Financial, INC. Proxy Statement"; Prudential(2024), "Ethical Principles for Artificial Intelligence"; Generali Group 홈페이지; Generali Group(2026), "Annual Integrated Report 2025"; Allstate(2026), "2026 Allstate Proxy Statement"; Travelers 홈페이지; The Travelers Companies, INC(2026), "Notice of 2026 Annual Meeting of Shareholders & Proxy Statement"; Travelers(2026), "2025 Annual Report"

○ **한편 인공지능 위험관리는 모델의 성능 관리나 사전적 통제에 그치지 않고, 소비자에 대한 영향과 인공지능 활용 결과의 신뢰성, 시스템 장애 등 상황에서도 핵심 보험업무를 지속·복구할 수 있는 운영 회복탄력성까지 포괄할 필요가 있음**

- 보험회사는 보험상품 추천, 인수심사, 보험금 지급 등 소비자에게 직접 영향을 미치는 업무에서는 인공지능 활용 사실의 고지뿐만 아니라 판단 결과에 대한 설명, 이의제기 및 임직원에 의한 재검토 등 소비자보호 절차를 갖추는 한편, 시스템 장애 등 위기 상황에 대비하여 인적 개입을 통한 결정의 변경 및 번복, 대체 모델 활용, 이전 상태로의 복구 및 긴급정지 등의 기능과 함께 수작업 전환, 대체 사업자 확보, 서비스 전환·종료 및 사고 대응 계획 등 업무의 지속·복구를 위한 방안을 마련할 필요가 있음

- 감독당국은 싱가포르의 FEAT(Fairness, Ethics, Accountability and Transparency),[22] Veritas[23] 및 MindForge 사례[24] 등을 참고하여 「금융분야 AI RMF」를 실제 업무에 적용할 수 있는 공통 평가방법론과 이행도구 및 보험업무별 적용 사례를 제공하되, 특정 조직 구조나 절차를 획일적으로 요구하기보다는 보험회사가 업무별 위험 수준에 적합한 관리·통제 방식을 선택하고 그 적정성을 설명할 수 있도록 지원할 필요가 있음

22) MAS(2018. 11. 12.), "Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector"
23) MAS(2021. 1. 6.), "Veritas Initiative Addresses Implementation Challenges in the Responsible Use of Artificial Intelligence and Data Analytics"; MAS(2022. 2. 4.), "MAS-led Industry Consortium Publishes Assessment Methodologies for Responsible Use of AI by Financial Institutions"; MAS(2023. 6. 26.), "MAS-led Industry Consortium Releases Toolkit for Responsible Use of AI in the Financial Sector"
24) MAS(2026. 3. 20.), "MAS Partners Industry to Develop AI Risk Management Toolkit for the Financial Sector"; MAS(2026. 1.), "AI Risk Management: Operationalisation Handbook"; MAS(2026. 1.), "AI Risk Management: Implementation Examples"