Ⅲ AI 활용 관련 주요 이슈

1. AI 활용에 따른 다양한 위험 발생

○ 보험산업 내 AI 활용은 생산성 및 소비자경험을 제고하는 등 긍정적 효과가 기대되지만, 설명가능성, 신뢰성, 편향성, 개인정보, 사이버 리스크의 이슈 측면에서 다양한 문제가 발생할 가능성이 있음

- 인공지능의 기반이 되는 딥러닝은 매우 정교한 프로그램이지만 사용자는 결과물이 도출된 원인에 대해 설명하거나 이해하는 것이 불가능할 수 있으며, 사고 발생 시 원인 해명과 책임 판단이 어려울 수 있음
- 머신러닝은 부여되는 데이터의 종류와 양에 따라 모델이 정해지기 때문에 정확도를 100%를 보장하기 어려우며, 때때로 실수할 수가 있음
- AI 훈련을 위한 데이터가 편향적인 성향을 지닐 경우, 이에 따라 나오는 결과물이 특정 대상에 대한 편견과 차별로 나타날 수 있음
- AI는 데이터 간의 관계 분석하여 익명 데이터를 식별할 가능성도 존재하여 이로 인해 개인정보의 침해 및 기밀 정보의 유출이 나타날 수 있음[32)]
- AI를 활용해 악성코드를 작성하거나 피싱 페이지를 만들어 해킹 혹은 악성코드를 유포시키거나, AI를 활용해 제작된 악성소프트웨어(멀웨어)는 기존 안티바이러스 모델의 탐지를 피할 수 있게 하는 등[33)] 사이버 리스크를 유발함

○ 기존의 AI 활용에 따른 이슈와 더불어 생성형 AI의 등장은 할루시네이션(환각), 저작권 이슈, 악의적 사용에 따른 오정보의 생성과 확산 등 피해가 더욱 확대될 가능성도 있음

- 대규모 언어모델 사용 등 복잡한 모델 사용은 학습데이터의 패턴을 때때로 과도하게 일반화하여 실제로 존재하지 않은 새로운 패턴과 이로 인한 결과를 도출할 수 있으며 결과를 마치 진실인 것처럼 도출(환각)하여 혼란을 야기할 수 있음
- 광범위한 데이터를 활용함에 따른 사용 데이터의 저작권 침해 문제가 발생할 수 있음[34)]

---

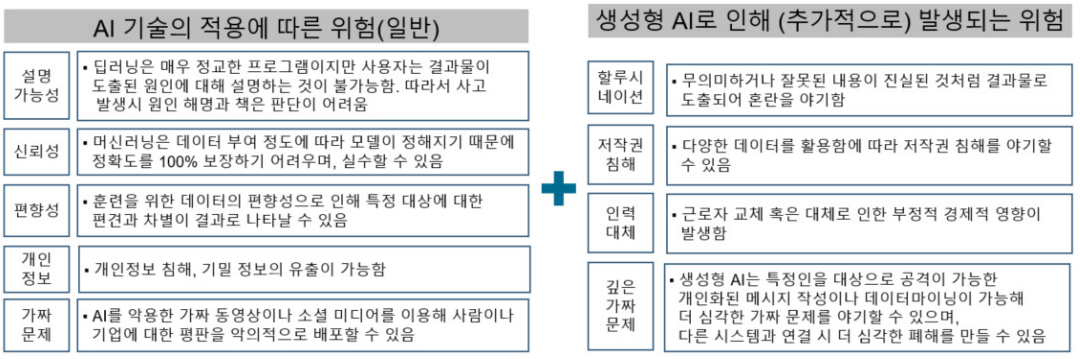
32) 생성형 AI의 활용은 이러한 데이터 간 관계식별을 할 가능성을 높여 개인의 식별 나아가 개인정보 침해 가능성을 높일 것으로 예상됨 (김동겸 2023)

33) 인공지능신문(2023), “인공지능 악용한 사이버공격 증가!… 랜섬웨어 공격과 데이터 절도, 메일 공격은 464% 증가”

34) ’23년 12월 미국 뉴욕타임즈(NYT)는 OpenAI와 Microsoft사에 대해 NYT의 콘텐츠를 무단으로 학습데이터로 사용하여 그 결과 자사 브랜드를 훼손하였다고 제소함

- 그동안 AI를 이용해 만들어진 가짜 동영상을 악의적으로 배포하여 발생한 가짜문제가 생성형 AI를 거치면서 더 심각한 가짜이슈를 발생시킬 수 있음
- 생성형 AI가 개인정보를 수집해 특정인을 대상으로 개인화된 메시지 작성이 가능한 기능을 이용해 이를 이용한 사이버공격이 가능하고, 데이터 마이닝 기능도 보유하고 있어 이를 이용한 사이버 리스크가 심화될 가능성도 높아지고 있음
- 이 외 생성형 AI 활용 확대에 따른 근로자의 교체 혹은 대체로 인해 사회경제에 부정적 영향이 발생할 수도 있음[35]

〈그림 Ⅲ-1〉 AI 활용에 따라 발생 가능한 위험



자료: 저자가 작성함

○ **보험산업 측면에서 살펴보면, AI가 모집 및 보상 등에 전면으로 활용되는 과정에서 불법행위가 발생할 경우 책임소재에 대한 이슈가 발생할 수도 있음[36]**

- 예컨대 AI 설계사에 의한 보험모집 과정에서 불완전 판매가 발생하거나 AI 지급심사로 인해 부당하게 보험금 지급을 거절할 경우, AI 개발사업자와 이를 이용해 판매한 설계사 중 누구에게 책임을 물을 것인지에 대한 문제가 발생 가능함
  - 실제 미국 StateFarm, Cigna, UnitedHealthGroup 등 보험회사는 동사가 운영하는 AI 기반 자동화 시스템이 인종 소수자와 고령층 고객에게 보험금 지급을 거부하는 결과를 낳았다는 이유로 집단소송이 진행 중임[37]
- 또한 초개인화된 개인별 맞춤형 상품이 기존의 보험법리 적용이 가능한지에 대한 이슈도 존재하며 계약자 위험도를 세분화할 경우 보험 소외계층을 양산하거나 차별을 야기할 가능성도 있음

---

35) MIT의 인공지능연구소(CASIL)에 따르면 인공지능에 의한 인력 대체는 당분간 어려울 것으로 예상된다고 발표했으나 향후 고도화된 생성형 AI가 광범위하게 적용될 경우, 우려하는 고용문제가 야기될 가능성도 간과할 수 없음

36) 황현아(2024)

37) Bloomberg Law(2023. 11)

○ 한편, AI 활용으로 인해 발생할 수 있는 다양한 이슈들을 AI 기술로의 대응하려는 노력[38]이 제시되고 있으며 대표적인 예로 설명가능한 AI가 제안되고 있음

- 설명가능한 AI(Explainable AI; XAI)란 AI의 판단 결과에 대한 이유를 이해할 수 있도록 제공하는 기술을 의미함
  - 예컨대 대출심사나 신용평가 모델에 AI를 적용했을 때 도출되는 결과의 이유를 설명하기 위해 입력된 특성(입력 데이터)의 결과 도출에 대한 기여도를 계산하여 제공함[39]
- 환자치료를 위한 의사결정의 투명성과 추적 가능성이 필수적인 의료분야, 투명한 대출과 신용 승인, 그리고 자산관리, 상품 추천 등에 소비자의 신뢰가 중요한 금융분야, 그리고 범죄 예측과 학습데이터의 편향을 방지가 중요한 형사사법 분야에 가장 활발하게 사용될 것으로 전망됨[40]
  - 금융업의 경우 AI의 안정성과 신뢰성 확보, AI 사용에 따른 컴플라이언스 준수[41]를 위해 특히 요구되며, AI를 활용한 데이터 분석 시 인사이트 추출 및 해석을 위해서도 XAI의 필요성이 제시되고 있음

2. AI 활용에 따른 위험에 대응하는 주요국 규제

○ 각국의 금융감독기관은 이러한 AI 활용 확대로 인해 발생할 수 있는 소비자 피해를 방지하고 금융기관의 책임있는 AI 기술 활용을 위해 다양한 대응책을 마련해 가고 있으며 초기에는 연성규범을 중심으로 제시되었으나 최근에는 경성규범[42]으로 전환하는 모습을 보이고 있음[43]

- 초기 AI와 관련된 규제는 대부분 연성규제로 OECD의 'AI 권고안'('19년), 유네스코의 'AI 윤리 권고안'('21년)이 대표적 사례이며 이들은 공통적으로 AI의 윤리적 사용과 AI가 지향해야 할 가치와 관리 원칙을 제시하고 있음
- 경성 규범적 접근의 가장 가까운 사례는 EU가 발의한 '21년 인공지능법(AI Act)으로, 동 법은 '23년 12월 집행위원회 및 의회가 법안 내용에 합의했으며, 의회 및 회원국들의 정식 승인절차를 밟아 '26년경에 발효될 것으로 예상됨[44]

---

38) 가짜문제의 가장 대표적인 사례인 딥페이크를 탐지하기 위한 AI 조작 검증 기술(인텔의 '페이크 캐쳐', 미 방위고등연구계획국(DARPA)의 '세마포(Semafor)', 센티널의 '센티널AI', 구글의 '신스ID')이 개발되고 있으며, 99% 딥페이크를 탐지하는 AI 기술(EMD 기술)도 개발되었으며, NVIDA Research는 아바타 핑거프린팅(누군가가 다른 사람의 동의 없이 AI로 애니메이션 된 닮은꼴을 사용하는지 감지하는 프로그램)을 개발하기도 함

39) 금융보안원(2024)

40) 한국IBM(2023)

41) 금융분야 AI 가이드라인에서는 관련법령에 따라 설명의무가 있는 금융서비스 또는 고위험 서비스에 AI 시스템을 활용하는 경우 설명가능한 AI 기술을 도입하기 위한 노력을 기울여야 함을 명시하고 있으며(금융분야 AI 가이드라인 및 주요 검토 필요사항(2021. 7)), 현재 국내에서는 금융보안원이 금융당국과 함께 사용가이드라인을 작성 중임

42) 연성규범(Soft Law)이란 직접적으로 법적 강제력을 갖지 않으나 간접적으로 사회구성원의 행위에 실질적인 영향력을 미치기 위해 만들어진 행위 규범을 말하며, 경성규범(Hard Law)이란 법적 구속력이 있는 규범을 말함

43) 고상원(2023)에 따르면 AI와 관련된 정책 대응 유형은 ① 연성규범, ② 경성규범, ③ 애플리케이션별 일시적 서비스 중지, 또는 전면 중지, 규제 실험을 위한 통제된 환경 촉진, 국제 표준화 노력 및 국제법 지원 등이 포함되나, 본고에서는 ①, ②, ③을 중심으로 살펴봄

- 미국은 '22년 AI 개발 등에 있어 고려해야 할 원칙을 위한 청사진을 발표하고, '23년 AI 사용에 대한 안전, 보안, 신뢰 확보에 대한 규제를 마련함
  - 최근 전미보험감독자협회(NAIC)은 보험회사의 인공지능 활용 관련 가이드라인을 발표하였으며 동 가이드라인은 보험회사의 데이터 및 인공지능 시스템을 제공하는 제3자에 대한 보험회사의 관리방안을 제시하고 있음[45]
- 중국은 생성형 AI 서비스 콘텐츠에 대한 규제를 담고 있는 생성형 AI에 대한 임시 관리 조치(The Provisional Administrative Measures of Generative AI service)를 발표(2023. 7)[46]하였고, 싱가포르는 금융기관의 책임있는 AI 기술 사용을 위한 Veritas Toolkit 2.0[47]을 발표(2023. 6)함
- 일본은 생성형 AI 규제를 위한 특정 법률은 없으나 개인정보보호법에 의거 오픈AI에 허가없는 민감 데이터 수집을 금지(2023. 6)했으며, 생성형 AI 개발 및 제공자에 대한 제3자 감시·인증제도 도입을 검토 중임[48]
- 우리나라의 경우 '20년 과학기술정보통신부가 'AI 윤리 기준'을 마련하였으며, 최근 AI 관련하여 13건의 법안[49]이 발의되어 있는 상태이며, 금융분야와 관련하여 금융위원회가 '21년 금융분야 AI 활용 가이드라인을 발표함

○ **Geneva Association은 AI와 관련된 주요 이슈와 관련된 각국의 규제를 정리하면서 관련하여 보험회사에게 요구되는 사항을 정리하였음(〈그림 Ⅲ-2〉 참조)**

- 투명성과 설명가능성 그리고 데이터와 관련된 각국 규제는 고객에게 객관적인 상품정보를 제공하고 데이터 사용 및 처리에 개방성과 투명성을 전제해야 함을 규정하고 있어 이에 대응해 보험회사는 AI에 공급되는 데이터에 대한 사례에 대한 투명성을 확보하고 데이터 보안을 강화할 것을 권고함
- 편향성, 차별성에 관련된 각국 규제는 개인 데이터의 합법적이고 공정하며 투명한 사용과 처리를 보장하고 인종 및 성차별 금지를 규정하고 있으며, 이를 위해 보험회사는 AI 모델에 불필요한 상관관계가 적용되지 않도록 감시체제를 마련하고 직원 대상 AI 위험 관련 교육프로그램 실시를 권고함
- 사람에 의한 감독과 관련하여 각국의 규제는 자동화로 인해 도출된 결정에 반대할 권리를 제공하고 사내 효과적인 거버넌스 체계를 갖출 것을 규정하고 있으며, 이에 대응해 보험회사도 강력한 거버넌스 및 감독을 설정하여 편향성 발생 가능성을 최소화해야 함을 주지시키고 있음

---

44) EU의 AI Act는 AI를 위험성 정도에 따라 금지, 고위험, 저위험, 허용되는 AI로 구분하고, 고위험 인공지능에 대해 제3자의 적합성 평가를 요구하는 내용을 담고 있으며, 금지된 AI 애플리케이션 위반이나 AI 법 의무 위반 시 벌금을 부과함

45) NAIC(2023. 12)

46) 2023년 8월 시행되었으며 콘텐츠를 생성하는 기술의 허용 범위, 준수 의무, 위반 시 조치 등의 내용을 포함함

47) 동 법안은 금융기관이 공정성, 윤리성, 책임성, 투명성 원칙을 회사의 리스크 거버넌스에 통합해야 함을 명시함

48) 파이낸셜 뉴스(2023. 7), "일 정부, 생성형 AI 개발·제공자 인증제도 창설 검토"

49) 발의된 법안 중 가장 최근에 발의된 '인공지능 책임 및 규제법(안)'은 EU의 AI Act와 유사하게 고위험 AI 사업자에 대한 책임의 일반 원칙에 대해 정하고 있음(황현아 2023)

〈그림 Ⅲ-2〉 AI 활용에 따른 위험 발생 방지를 위한 글로벌 규제와 보험회사 대응

주요 위험	관련 규제		보험사에게 요구되는 사항
투명성 설명가능성 데이터 관련	▪ EU IDD Article 20 ▪ EU GDPR Article 5,13,14, 30 ▪ 미국 Gramm-Leach Bliley Act ▪ 중국 개인정보보호법 조항 5	▪ 고객에게 객관적인 상품정보를 제공 ▪ 데이터 사용 및 처리에 개방성과 투명성 전제 ▪ 데이터 활용 시 적절성, 관련성,정확성 포함 ▪ 데이터 처리 활동의 기록	▪ AI 모델에 공급하는 데이터에 대해 사용 사례별 투명성 확보 ▪ 데이터에 대한 철저한 관리 ▪ 데이터 보안 강화
편향성, 차별 관련	▪ EU Racial Equality Directive, Gender Directive, ▪ EU GDPR Art 5 ▪ 미국 California Consumer Privacy Act(CCPA) 등 ▪ 중국 인터넷 규제 대책 조항 17	▪ 인종 차별, 성차별 금지 ▪ 개인 데이터의 합법적이고 공정하며 투명한 사용 및 처리 보장	▪ AI 모델에서 불필요한 상관관계가 적용되지 않도록 감시하고 방지하는 방안 고안 ▪ AI 모델에 활용되는 평가 항목들의 수를 제한 ▪ 보험사 직원 대상 AI 교육 프로그램 개발
사람에 의한 감독	▪ EU S-II Directive Art 41 ▪ EU GDPR Article 22 ▪ 중국 개인정보보호법 조항 24, 보험법 조항 5	▪ 자동화로 인해 도출된 결정에 반대할 권리 제공 ▪ 건전한 비즈니스를 위해 효과적인 거버넌스 시스템의 필요성	▪ 강력한 거버넌스 및 감독을 설정해 편향성 발생 가능성 완화

자료: Geneva Association(2023)

3. 물리적 망 분리 규제와 AI 활용 부담 이슈

○ ChatGPT와 같은 생성형 AI의 도입과 관련하여 보험업을 비롯한 금융업은 망 분리 규제로 인해 AI 활용에 부담이 적지 않음

- 망 분리 규제는 보안을 위해 금융회사 내부 업무용 시스템과 일반 인터넷 외부 통신망을 분리, 차단 및 접속 금지하는 조치[50]를 말함
- 다양한 디지털 금융서비스 제공에 망 분리 규제가 걸림돌로 작용한다는 의견이 존재함[51]
  - 망 분리 규제를 적용받지 않고 AI 시스템을 이용하기 위해서는 금융회사 전산센터 내부에 위치한 온프레미스 형식 혹은 프라이빗 클라우드 형식이어야 하며, 이때 시스템 구축 비용이 부담으로 작용함

○ 그러나 민감한 개인 금융 정보를 관리하는 금융회사에 있어 개인정보 유출, 악성코드 감염, 해킹 등의 위험도 간과할 수 없어 이에 대한 심도 있는 논의가 필요함

- 금융당국은 SaaS(Software as a Service) 관련 물리적 망 분리 규제의 조건적 면제를 위한 규제 샌드박스를 지정하는 등 망 분리 규제 개선방안을 제시[52]하였으며 단계적 개선을 언급하는 등 향후 규제의 유연화 가능성도 보임[53]

---

50) 2013년 대규모 사이버테러사태로 국내 주요 언론 및 금융사의 전산망이 마비되고 악성코드로 인해 약 3만 2천여 대의 시스템이 감염되는 피해가 발생한 '3. 20 전산망 마비사태' 이후 금융위는 전자금융 감독 규정을 개정하여 물리적 망 분리를 의무화함

51) 서병호(2023); 전자신문(2023), "망 분리 규제, "시대착오적" vs. "대안은 있나""

52) 금융위원회 보도자료(2022), "금융분야 클라우드 및 망 분리 규제 개선방안"

53) 금융권 망 분리 정책 개선 끝장 토론회(2023. 9. 14)