NEWSLETTER

April 2023

디지털금융팀

Digital Finance Team

CONTACT

변호사 고환경

T: 02.2191.3057
E: hwankyoung.ko@leeko.com

변호사 이정명

T: 02.6386.0730
E: chloe.lee@leeko.com

변호사 이일신

T: 02.772.5982
E: ilshin.lee@leeko.com

변호사 김규성

T: 02.6386.0798
E: gyusung.kim@leeko.com

금융위원회, 「AI 기반 신용평가모형 검증체계」 및 「금융분야 AI 보안 가이드라인」 마련

금융위원회는 2023년 4월 18일 신뢰받는 AI 활용 환경을 구축하기 위한 「AI 기반 신용평가모형 검증체계」와 「금융분야 AI 보안 가이드라인」을 발표하였습니다. 이는 금융위원회가 지난 2022년 8월 발표한 「금융 분야 인공지능 활용 활성화 및 신뢰확보 방안」의 세부과제 이행을 위한 후속조치입니다.

최근 AI 방법론을 이용한 신용평가모형의 개발 · 활용이 활발히 이루어지고 있으며, AI 서비스 활용 확대와 더불어 개인정보 유출, 학습 데이터 조작 등 다양한 보안위협이 발생할 우려가 증가하고 있습니다. 「AI 기반 신용평가모형 검증체계」 및 「금융분야 AI 보안 가이드라인」은 이와 같은 AI 활용의 특수성을 고려하고, 새로운 보안 위험에 대비하기 위하여 마련되었습니다. 이하에서 구체적인 내용에 대해 설명드리겠습니다.

1. AI 기반 신용평가모형 검증체계

■ 추진 배경

AI 신용평가모형은 다양한 비금융 · 비정형 데이터를 평가항목으로 반영할 수 있어 예측력 · 변별력이 우수하다는 장점이 있으나, 복잡한 알고리즘을 사용하기 때문에 평가결과에 대한 직관적인 해석과 설명이 다소 어렵다는 한계도 존재합니다. 이에 AI 신용평가모형의 신뢰성을 확보하고 소비자 보호를 강화하기 위해 신용정보회사(이하 **CB사**)에 대한 「AI 기반 신용평가모형 검증체계」가 마련되었습니다.

■ 주요 내용

(1) 데이터 관리

신용평가에 활용하는 비금융 · 비정형 데이터를 신뢰성 높은 출처로부터 수집하고 데이터의 일관성 · 정확성 등을 주기적으로 확인하는 등 CB사가 적절한 관리체계를 구축하였는지 점검합니다.

(2) 모형 선택

알고리즘 선정 목적, 변수 선정 과정 등 모형 개발의 상세 과정을 확인하여, CB사가 다양한 AI 알고리즘의 특징과 장단점을 고려하여 신용평가에 최적화된 모형을 선택했는지 모형 선정 과정을 점검합니다.

(3) 설명 가능성

CB사가 자동화평가 결과에 대한 설명의무*를 준수하여 금융소비자에게 신용평가모형 및 신용평가 결과 등을 충분히 설명할 수 있는지 확인합니다.

* 신용정보의 이용 및 보호에 관한 법률 제36조의2에 따라 개인인 신용정보주체는 자동화된 신용평가에 대하여 평과 결과, 평가 기준, 평가에 이용되는 정보의 내용 등에 대하여 설명을 요구 가능

(4) 모형 성능

과적합(overfitting)* 가능성 점검, 학습·검증·테스트 데이터의 유사성 확인 등 AI 모형에 특화된 성능 확인 방법을 통해 모형의 변별력·안정성 등 통계적 유의성을 점검합니다.

* 모형이 학습 데이터에 과도하게 의존하는 것을 의미

■ 향후 계획

개인신용평가체계 검증위원회는 「AI 기반 신용평가모형 검증체계」를 활용하여 2023년 중 AI 신용평가모형을 개발 · 활용하고 있는 개인사업자CB사를 검증하고, 향후 개인CB사 등의 AI 신용평가모형에 대해서도 검증을 실시할 계획입니다.

2. 금융분야 AI 보안 가이드라인

■ 추진 배경

AI 서비스의 보안성을 충분히 확보하지 못하는 경우, AI 서비스가 오작동하거나 악의적인 공격에 노출될 가능성이 있어, AI 개발과정에서 현장 실무자가 유의해야 할 구체적 보안위협과 대응방안을 다루는 세부적인 안내서가 필요하다는 지적이 제기되어 왔습니다. 이에 금융 AI 서비스 개발 실무자가 활용할 수 있는 개발단계별 세부 보안 안내서인 「금융분야 AI 보안 가이드라인」이 마련되었습니다.

■ 주요 내용

(1) AI 모델 개발단계별 주요 보안 고려사항

■ 학습 데이터 수집

신뢰성 높은 출처로부터 학습 데이터를 수집하고, 데이터 출처 및 수집 시점 등을 파악할 수 있는 데이터 관리체계를 구축해야 하며, 데이터 관련 공격 · 장애 발생시 그 원인을 파악하여야 합니다.

■ 학습 데이터 전(前)처리

이상치(outlier)를 확인 · 처리하는 등 수집한 데이터를 학습에 적합한 형태로 가공하고, 적대적 예제* 생성 · 학습 등을 통하여 AI 모델을 적대적 공격으로부터 보호하는 등 AI 모델의 품질과 보안성을 높여야 합니다.

* AI 모형이 잘못된 예측을 하도록 의도적으로 변조한 데이터

■ AI 모델 설계 · 학습

잠재적 공격자가 AI 모델에 대한 정보를 쉽게 유추할 수 없도록 지나치게 단순한 설계를 지양하고, AI 모델을 세부 변형하는 보안기법 등을 적용하여 적대적 공격 등을 쉽게 수행할 수 없도록 AI 모델을 구성하여야 합니다.

■ **AI 모델 검증 · 평가**

AI 모델을 대상으로 선제적인 적대적 공격을 수행하여 AI 모델이 공격을 탐지 · 방어할 수 있는지 확인하고, AI 모델의 입 · 출력 횟수를 제한하여 잠재적 공격자가 AI 모델에 대한 정보를 수집하기 어렵게 하였는지 평가하여야 합니다. 또한, AI 모델을 통하여 개인정보가 출력되는 경우, 개인정보가 타인에게 노출되지 않는지 검증하여야 합니다.

(2) AI 챗봇 서비스에 대한 보안성 체크리스트

금융분야에서 AI가 가장 활발히 사용되는 분야 중 하나인 챗봇서비스에 대해서는 보안성 확보에 필요한 사항들을 체크리스트 형태로 구체화하였습니다.

※ 체크리스트 문항 예시

분류	소분류	연번	보안성 체크리스트
공통	입력제한	6	개인정보를 처리하지 않는 챗봇의 경우 입력창에 개인정보를 입력하지 않도록 이용자에게 사전에 안내하는가?

3. 시사점

신용평가는 금융소비자의 금융거래 조건 등에 영향을 미치며 금융회사의 리스크 관리에도 밀접한 관련이 있어 투명성과 공정성을 확보할 필요성이 더욱 강조됩니다. 이에 AI를 활용하여 신용평가모형을 구축하고 있는 CB사 및 신용평가모형을 개발하는 사업자들로서는 「AI 기반 신용평가모형 검증체계」에 대응하여 내부 시스템 및 절차를 점검하고 개선할 필요가 있습니다.

한편, 금융분야는 데이터가 풍부하게 집적되어 있어 빅데이터 · AI 활용의 잠재력이 매우 높지만, 개인신용정보 등 민감정보가 다수 포함되어 있고 정보 유출, 알고리즘 오작동 등의 사고가 경제적 손실로 직접적으로 이어질 수 있어 보안의 중요성도 점점 커지고 있습니다.

금융위원회는 새로운 보안위협 · 대응기법 등을 고려하여 본 가이드라인을 지속적으로 개선 · 보완할 예정임을 밝히고 있으므로 그 개정 사항을 모니터링하면서 AI 보안 위협에 선제적으로 대응할 필요가 있습니다.

이번 조치 이외에도 「금융 분야 인공지능 활용 활성화 및 신뢰 확보 방안」의 다른 세부과제(금융 AI 데이터 라이브러리 구축, 설명가능한 AI 요건 안내서 마련, 금융 AI 테스트베드 구축 등)와 관련한 여러 후속조치가 예정되어 있는바, 금융분야 AI 관련 동향을 유심히 지켜볼 필요가 있겠습니다.

법무법인(유) 광장의 디지털금융팀은 AI에 대한 높은 이해도를 바탕으로 AI와 관련된 법제도 개선 등을 위한 활동 및 대내외 세미나 및 연구, 고객에 대한 법률 자문 등 폭넓은 경험을 가지고 있습니다. 특히 법무법인(유) 광장의 디지털금융팀에는 신용평가업 및 금융 보안에 관한 고도의 전문성을 갖춘 변호사, 전문위원들이 대거 포진하여 있는바, AI와 결합된 신용평가모형 개발 및 보안체계 구축에 관하여, 금융당국의 요구사항을 정확히 반영한 신속한 자문 및 컴플라이언스 점검이 가능합니다. 위 내용 및 금융분야의 AI 이슈와 관련하여 법률전문가의 도움이 필요하신 경우 언제든지 법무법인(유) 광장의 디지털금융팀으로 연락주시기 바랍니다.

이 뉴스레터의 내용에 관해 궁금한 사항이 있으시면 상단 연락처로 연락 주시기 바랍니다.

이 뉴스레터는 일반적인 정보 제공만을 목적으로 발행된 것으로서, 법무법인(유) 광장의 공식적인 견해나 법률의견이 아님을 알려드립니다. 법무법인(유) 광장에서 발송하는 뉴스레터를 원하지 않으시면 [수신거부]를 클릭해 주십시오.

뉴스레터 더 보기