금융분야 AI
개발·활용 안내서

금융위원회 금융감독원 한국신용정보원 금융결제원 금융보안원
FINANCIAL SECURITY INSTITUTE

은행연합회 금융투자협회 생명보험협회 손해보험협회 여신금융협회

Contents

금융분야 AI 개발·활용 안내서

Chapter 01 개 요

Chapter 01
개 요

1. 발간 배경 및 목적

현재 AI 기술은 전 산업에 걸쳐 다양한 분야에서 활용되고 있으며 특히 금융 분야에서 금융포용과 생산성 향상을 목적으로 활발한 도입이 이루어지고 있다. 그간 국내 금융산업은 금융규제의 혁신과 함께 핀테크, 빅테크 등 비금융회사들이 새롭게 진입함에 따라 양적인 성장을 이루어 왔으며, 앞으로는 AI와 같은 기술적 혁신을 기반으로 한 질적인 변화가 기대되고 있다. 그러나 AI 기술은 높은 수준의 생산성 향상이라는 장점에도 불구하고, 금융소외계층의 포용이나 차별적 요소 등 금융소비자 보호를 위해 고려해야 할 점이 존재하며 안전성에 대한 사회적 신뢰도 충분하지 못한 상황이다.

이러한 문제의식을 바탕으로 AI 활용 활성화를 도모하는 동시에 AI 서비스의 신뢰를 제고할 수 있는 「금융분야 AI 가이드라인」을 지난 2021년 7월 발표한 바 있다. 해당 가이드라인은 AI 서비스의 신뢰 제고에 필요한 최소한의 준칙을 제시하는데 주요 목적이 있었으며, 발표 이후 금융업권 전파와 가이드라인 이행을 위해 약 1년여 간의 준비기간 동안 금융회사 실무자들의 의견을 수렴하였다.

본 안내서는 금융업권별 협회를 중심으로 한 AI 서비스 관련 금융회사 실무자의 의견을 바탕으로, AI 서비스의 신뢰 제고에 필요한 최소한의 준칙을 제시했던 기존의 「금융분야 AI 가이드라인」에서 나아가 금융회사의 AI 서비스의 개발 · 운영 과정에서의 규제 불확실성을 해소하는 한편, 인공지능 활용 과정에서 발생 가능한 리스크를 예방 · 관리하는 데 목적이 있다.

안내서는 「금융분야 AI 가이드라인」에서 제시된 1) 목적과 적용 범위, 2) 거버넌스의 구축, 3) 기획 · 설계 단계, 4) 개발 단계, 5) 평가 · 검증 단계, 6) 도입 · 운영 · 모니터링 단계, 7) AI 업무위탁에 대한 특례 등 7가지 항목을 중심으로 금융회사에 공통으로 적용되는 1) 목적과 적용 범위, 2) 거버넌스의 구축, 7) AI 업무위탁에 대한 특례 등 3가지 항목과 5대 서비스별로 차이를 보이는 3) 기획 · 설계 단계, 4) 개발 단계, 5) 평가 · 검증 단계, 6) 도입 · 운영 · 모니터링 단계 등 4가지 항목을 구분하여 설명하였으며, 각 서비스 간의 특징이 안내서 상에 명확히 드러날 수 있도록 서술하였다.

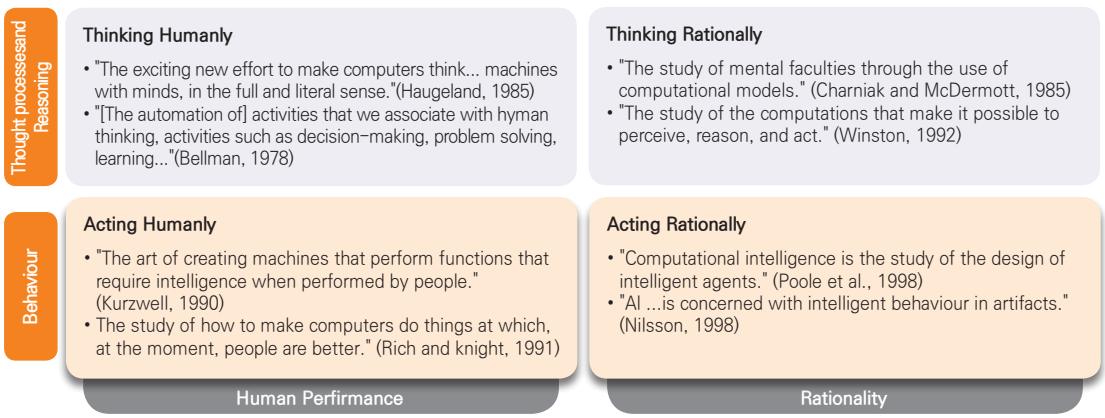
본 안내서는 「금융분야 AI 가이드라인」을 구체화한 세부 안내서로서 금융회사가 실제 AI 서비스를 개발 · 운영하는데 있어 손쉽게 적용할 수 있도록 체크리스트 형태로 구성하는 한편, 체크리스트 항목별로 판단기준들을 제시하고 충족 여부를 점검할 수 있도록 하여 실무적으로 활용하기 용이할 것으로 예상된다. 또한 향후 AI 기술의 발전, 금융규제의 변화, 금융산업 내 AI 기술의 수용도 등과 시행과정에서의 보완점을 반영하여 상시적인 개선 · 보완이 진행될 예정이다.

2. AI 개요

AI의 여러 정의를 살펴보면, Bellman(1978)은 "사람의 생각과 관련된 활동(의사결정, 문제해결, 학습 등)을 지능화하는 것"이라 정의했으며, Nilsson(1990)은 "인공물이 지능적인 행위를 하도록 하는 것", Wilson(1992)은 "인지하고 추론하고 행동할 수 있도록 하는 컴퓨팅"이라 정의했다. AI(Artificial Intelligence)는 인간의 지적능력(학습능력, 추론능력, 지각능력)을 인공적으로 구현하는 컴퓨터과학의 기술이라고 할 수 있다. 그러나 인간의 지적능력이란 매우 모호하고 추상적인 개념이어서 기계의 지적수준을 설계함에 있어 어느 수준까지 만족하면 AI라고 부를 수 있는지 경계를 명확하게 정해놓기도 어렵다. 이처럼 개념이 불명확하다는 것은 그만큼 규제의 대상을 특정하기 어렵다는 이야기이기도 하다.

한편, Russell and Norvig(2020)은 AI를 크게 4가지로 구분했는데, 1) 생각하는 것과 행동하는 것, 2) 인간처럼 또는 이성적으로 중 어떤 것에 중점을 둘지에 대한 2가지 기준에 따라 '인간처럼 생각(Thinking Humanly)', '인간처럼 행동(Acting Humanly)', '이성적으로 생각(Thinking Rationally)', '이성적으로 행동(Acting Rationally)' 으로 구분했다. 그러나 이러한 구분에도 '인간처럼', '이성적으로'라는 개념은 여전히 명확히 구분 짓기 모호한 부분이 있다.

〈 AI 구분 〉



자료 : Russell and Norvig(2020)

AI 알고리즘이 과거에는 사전에 다수의 규칙을 입력해 놓고 이에 따라 판단하도록 하는 '규칙 기반(Rule based)의 알고리즘'이 다수였다고 하면, 최근에는 데이터를 기초로 기계가 스스로 규칙을 찾아내도록 하는 '기계학습(Machine learning)'기반으로 구현되고 있다.

이러한 기계학습은 크게 3가지 방법론으로 나눠 볼 수 있다. 지도학습(Supervised learning)은 기계가 입력변수를 토대로 출력변수를 가장 잘 예측할 수 있는 모델을 생성하는 방법론이다. 지도학습에서는 데이터로부터 어떻게 규칙을 찾아내야 할지에 대한 알고리즘은 인간이 직접 구현(지도)해야 하며, 이때, 규칙을 찾아내기 위해 사용되는 데이터를 학습데이터(training data)라고 한다.

비지도학습(Unsupervised learning)은 입력정보만 주어지거나 입력정보에 대한 결과가 명확하지 않은 경우 입력정보만으로 기계가 스스로 학습해서 집단에 대한 규칙을 찾아내는 것이다. 비지도학습의 대표적인 예가 군집화(Clustering)로, 집단의 특성을 나타내는 데이터로부터 비슷한 특성을 갖는 집단들을 찾아낸다.

강화학습(Reinforcement learning)은 현재 상태(state)를 기반으로 선택 가능한 행동(action)들을 연속적으로 취했을 때 최종적으로 보상(reward)을 받게 되는데, 이 보상을 가장 극대화할 수 있는 방법을 찾도록 학습하는 방식이다. 강화학습은 입출력 쌍으로 데이터가 주어지지 않으며, 잘못된 행동에 대해서 명시적으로 정정이 일어나지 않는 점에서 지도학습과 구분된다. 다만 행동들의 선택으로 얼마나 좋은 결과가 나왔느냐에 따라 보상이 차등화되도록 하여 기계가 주어진 상태에서 최선의 행동을 할 수 있도록 학습된다.

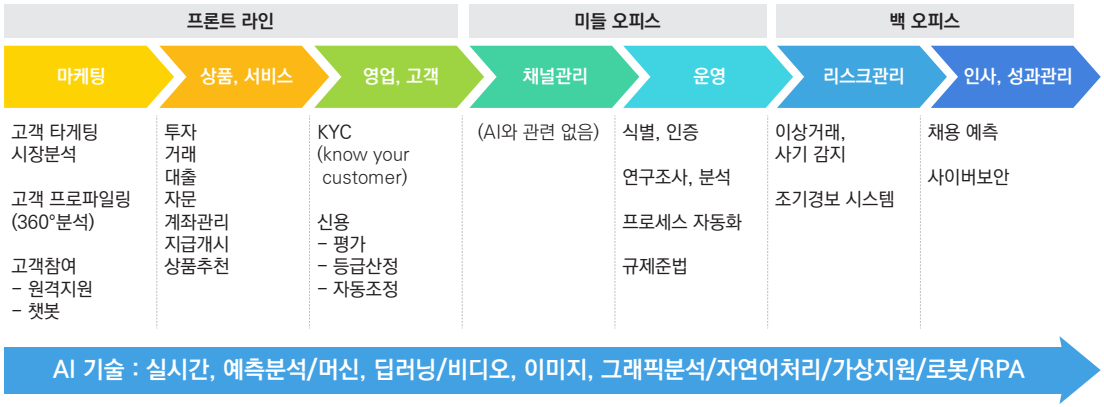
AI의 주요 응용분야를 살펴보면, 데이터를 활용해 지식을 생성하고 이를 토대로 새로운 지식을 추론하는 학습 및 추론, 사람의 언어를 이해하고 이를 모델화하여 활용하는 언어이해, 시각 정보를 이해하고 분석하여 활용하는 시각인식, 인간의 감정이나 주변의 사건 · 사고를 종합적으로 인지해 행동하도록 하는 상황인식으로 나누어 볼 수 있다.

3. 금융권 AI 서비스 특성 및 동향

금융업의 경우 신뢰할 수 있는 양질의 데이터가 집중되어 있으며, 수리적 또는 규칙 기반의 의사결정을 한다는 업무적 특성으로 인해 AI 활용이 빠르게 확산할 수 있는 분야이다. 은행의 경우 신용평가나 재무리스크 측정에서, 증권업의 경우 포트폴리오 구성이나 프로그램 트레이딩 등의 업무에서 인공신경망 모델 등이 꾸준히 사용되어 왔다. 특히 금융회사가 보유하고 있는 고객 데이터는 개인 신상정보, 소득정보, 금융거래 정보 등 금융의사결정에 직접 영향을 미치는 개인신용정보를 다수 포함하고 있어 금융업의 AI 활용을 촉진하는 양질의 토양으로 작용하고 있다.

국내외 금융회사와 핀테크, 빅테크 기업들은 금융사업에 AI를 다양하게 활용하고 또 실험하고 있다. 대표적 분야로는 빅데이터 분석에 기반한 금융시장 분석 및 전망, 금융상품 추천, 담보물의 가격결정, 리스크 특정 및 관리, 투자전략 수립, 포트폴리오 리밸런싱, 고빈도 트레이딩, 자연어 처리, 고객응대, 인증, 사이버 보안, 이상거래 탐지, 업무자동화 등이 있다. AI 기반 업무는 AI의 학습능력이 발달함에 따라 예측 및 분석의 정밀도 같은 업무수행 능력도 함께 증대될 것이기 때문에 변화하는 환경에서 금융회사의 핵심 경쟁력이 될 것으로 예상된다.

〈 은행업의 AI 활용분야 및 수준 〉



자료 : Deloitte(2019)

AI 기술을 챗봇에 활용하고 있는 대다수의 금융회사는 고객용 챗봇뿐만 아니라 직원용 챗봇도 함께 개발하여 활용하는 경우가 많다. 대고객 챗봇의 경우 비대면 금융거래가 증가하고 모바일 채널이 활성화 됨에 따라 고객의 질문에 상담원이 아닌 AI가 답변하는 비율이 크게 증가하고 있다. 특히 금융산업의 경우 고객의 문의가 다른 산업영역에 비해 비교적 정형화 되어 있어 대고객 챗봇 도입이 여타 산업 대비 용이한 측면도 있다.

전문인력의 개입 없이도 자산관리 자문 및 투자일임 등의 서비스를 제공하는 알고리즘인 로보어드바이저는 그 이용자 수가 꾸준히 늘고 있으며, 업종별로는 은행, 서비스 유형별로는 상품추천 서비스에서 가장 많이 활용되고 있다.

〈 국내 로보어드바이저 기반 서비스 계약자 〉

(단위 : 명)

		2017년 말	2018년 말	2019년 말	2020년 말	2021년 6월
업종별	증권사	2,604	6,023	6,928	6,384	1,098
	자산운용	32	17	2,036	21,662	29,094
	자문일임	143	1,002	4,982	63,216	134,474
	은행	35,928	50,828	121,404	187,400	214,811
서비스 유형별	일임	162	108	2,203	21,810	42,367
	자문	2,617	6,934	11,689	69,452	122,299
	무료추천	35,928	50,828	121,404	187,400	214,811

주 : 1) 테스트베드를 통과한 회사만을 대상으로 집계
2) 은행은 무료추천 서비스만 제공하고 있음.

자료 : 로보어드바이저 테스트베드 센터

심사 및 평가 분야의 경우 국내 은행에서 AI 도입이 가장 활발하게 추진되고 있는 분야로 금리승인, 한도의 세부 조정, 관련 오차 확인 등에서 적극적으로 AI 기술을 활용할 것으로 기대된다. 특히, 개인대출 부문에서는 기존 평가모형으로는 금융서비스를 받기 어려웠던 씬파일러(데이터가 부족한 신용이력 부족 차주)를 대상으로 AI 기반의 신용평가모형 개발이 이루어지고 있다. AI 모형에 이동통신 납부, 학자금 상환율, 세금납부 내역 등 상환율이나 부도율과 직접적인 상관관계를 입증할 수는 없으나 '관계의 유의성'을 평가하여 설명변수로 포함하는 방식으로 AI 적용 모형을 대안적으로 활용하고 있다.

금융권에서는 오래전부터 FDS 분야에서 AI를 활용해 왔으며, 자금세탁 방지 및 부정대출 탐지 부문에서도 비정상 패턴을 모델링하여 이상거래를 판별하고 있다. 과거에는 규칙 기반으로 사람에 의해 조정된 변수와 가이드에 의해 결과가 산출되었다면, 최근에는 딥러닝 알고리즘을 활용하여 고객 행동을 스코어링 하고 위험등급에 따라 조치를 취하는 방식이 시도되고 있다.

한편, 금융권 내 AI 활용이 활발해지면서 AI 활용과 관련된 리스크도 점차 커지고 있다. 최근 AI 활용 과정에서 방대한 자료와 복잡한 딥러닝 알고리즘이 적용되면서 다수의 변수가 상호작용하여 결과값이 산출됨에 따라 입력값과 결과값 간 인과관계를 알기 어려워지는 AI의 블랙박스 현상이 심화되고 있다. 또한, AI를 학습시키는 과정에서 편향이 반영된 데이터를 사용함에 따라 AI가 편향을 학습하고 특정 집단에 대해 불리한 결과값을 산출하여 차별을 체계화할 수 있다는 우려도 있다. 이에 따라, AI 활용 과정에서 AI 오작동이나 편향 등으로 사고나 불이익이 발생하더라도 그 원인을 파악하여 책임소재를 규명하고 손해배상 등 적절한 책임을 묻기가 어려울 수 있다는 지적이 제기되고 있다.

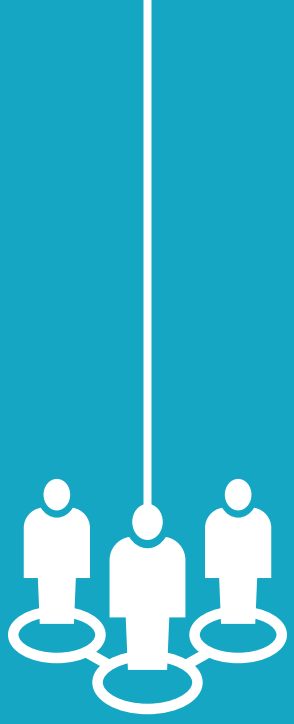
4. 안내서 개발과정 및 활용대상

2022년 1월부터 금융위원회, 금융감독원 및 각 금융업권 협회로 구성된 워킹그룹을 구성 · 운영하였다. 이 과정에서 실제 금융권에서 활용되는 AI 서비스를 조사하여 이 중 가장 많이 활용되는 5대 서비스(신용평가 및 여신심사, 이상거래 탐지, 챗봇, 맞춤형 상품추천, 로보어드바이저)를 도출하였으며, 각 서비스별 특징이 반영될 수 있는 〈금융분야 AI 개발 · 활용 안내서〉 마련을 추진하였다.

이를 위해 「금융분야 AI 가이드라인」의 7가지 항목 중 각 서비스 특징과 무관하게 금융회사가 공통으로 준수해야 할 영역을 식별하고 이에 따라 1) 목적과 적용 범위, 2) 거버넌스의 구축, 7) AI 업무위탁에 대한 특례 등 3개 항목은 공통 영역으로 분류하여 AI 서비스를 개발 · 운영하고자 하는 모든 금융회사가 준수할 사항으로 확인하였으며 해당 내용에 대한 세부 내용을 작성하였다. 또한 3) 기획 · 설계 단계, 4) 개발 단계, 5) 평가 · 검증 단계, 6) 도입 · 운영 · 모니터링 단계 등 4개 항목의 경우에는 서비스별 특징이 드러날 수 있도록 서술하였다. 안내서 초안은 4월부터 5대 서비스별로 각 금융업권 협회(신용평가 및 여신심사 – 은행연합회, 이상거래 탐지 – 여신협회, 챗봇 – 생보협회, 맞춤형 상품추천 – 손보협회, 로보어드바이저 – 금투협회)를 간사로, 업권별 주요 금융회사를 참여기관으로 하여 작업이 진행되었으며, 유관기관과 금융회사의 의견 수렴을 거쳐 7월에 안내서 최종본이 완성되었다.

본 안내서는 「금융분야 AI 가이드라인」의 "목적과 적용 범위"에 언급된 바와 같이, 금융서비스 및 금융상품의 제공을 위한 업무에 AI 시스템을 직 · 간접적으로 활용(금융회사 내부 직원관리, 단순 업무 효율화 등 AI 시스템 활용으로 고객에 미치는 영향이 없는 경우를 제외)하거나 활용하고자 하는 금융회사나 금융연관 서비스 제공을 위한 업무에 AI 시스템을 직 · 간접적으로 활용하거나 활용하고자 하는 비금융회사가 주요 활용대상이라 할 수 있다. 또한 금융회사 등에게 업무를 위탁받아 해당 서비스를 개발 · 운영할 필요가 있는 외부 기관의 경우에도 본 안내서의 주요 이해관계자가 될 수 있다.

금융분야 AI 개발·활용 안내서



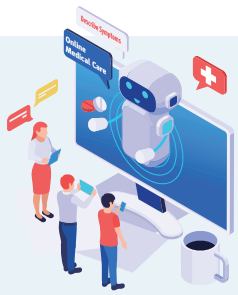
Chapter 02

공통 부문 안내

(목적 · 적용범위, 거버넌스, 위탁)

Chapter 02

AI 개발·활용 안내서 (공통 영역)



1. 목적과 적용 범위

기본원칙 1-1

본 안내서의 적용 범위는 AI 활성화 및 금융서비스에 대한 고객신뢰 확보라는 〈금융분야 AI 가이드라인〉의 기본 취지를 훼손하지 않는 범위 내에서 기관 내 AI 시스템별 특성을 종합적으로 고려하여 조정해야 한다.

점검항목

〈금융분야 AI 가이드라인〉의 기본 취지와 기관 내 AI 시스템별 특성을 종합적으로 고려하여 본 안내서의 적용 범위를 결정하였는가?

필요성

본 안내서의 적용 범위를 명확한 기준에 따라 결정하였는지 확인해야 한다.

체크리스트

1) 안내서 적용대상으로 결정된 AI 시스템에 대해 활용목적과 원칙, 핵심 개념들이 내부 업무지침 등에 명확하게 정의되어 있는가?
YES ☐ | NO ☐

2) 안내서 적용 범위 결정시, AI 시스템이 활용된 서비스의 특성, 고객 특성, 고객 수 등 해당 AI 시스템의 특성을 종합적으로 고려하였는가?
YES ☐ | NO ☐

3) 안내서 적용 범위 결정시, AI 활성화, 금융서비스에 대한 고객신뢰 확보 등 가이드라인의 기본 취지를 훼손하지 않았는지 확인하였는가?
YES ☐ | NO ☐

4) 기타 안내서 적용 범위를 조정 · 결정한 합리적인 사유가 존재하는가?
YES ☐ | NO ☐

준수사례 (예시)

- A카드사는 AI 서비스 활용 목적과 고객 영향, AI 서비스의 특성, 고객 특성, 고객 수 등을 검토한 결과, 운영중인 AI 시스템 중 개인사업자 신용평가모형만이 안내서 적용 범위에 해당된다고 판단하였다.
- B은행은 고객에 대한 대출심사시 AI를 적용한 신용평가모형을 활용하고 있다. 해당 AI 시스템 개발 · 운영 관련 인력들은 〈금융분야 AI 가이드라인〉의 기본 취지와 현재 운영 중인 AI 시스템의 특성을 고려하여, 〈금융분야 AI 가이드라인〉과 이에 따른 안내서 준수 대상임을 인지하고 있으며, 활용 목적과 핵심 개념들이 정의되어 있는 업무 매뉴얼을 작성 · 관리하고 있다. 만약 AI가 의사결정을 대체하지 않고 단순히 보조지표로만 활용되는 경우, 일부 체크리스트를 "N/A"로 적용한다.
- C은행은 현재 운영중인 고객 대상의 챗봇 서비스를 가이드라인과 안내서의 적용 범위에 속하는 것으로 판단하였으며, 챗봇 서비스 활용범위를 특정 앱의 기능과 업무 안내에 한정하고, 개발 · 활용 관련 내부 업무 매뉴얼을 작성하여 운영하고 있다.

기본원칙 1-2

AI 시스템을 현재 운영하고 있거나 도입 예정인 금융회사 등은 〈금융분야 AI 가이드라인〉과 관련 안내서를 참고할 수 있다.

점검항목

가이드라인과 관련 안내서 준수를 위한 내부 업무지침이 마련되어 있는가?

필요성

가이드라인과 관련 안내서의 내용을 준수하기 위해 기관 내에서 별도로 마련한 내부 업무지침이 필요하다.

체크리스트

1) 가이드라인과 관련 안내서 준수를 위해 내부 업무지침을 마련하였는가?

- 기존 내부 업무지침 외에 AI 시스템 도입 · 운영시 추가로 가이드라인과 관련 안내서 준수를 위한 지침이 마련되어 있으며 문서화가 이루어져 있는지 확인한다.

YES ☐ | NO ☐

준수사례 (예시)

- A은행은 가이드라인과 관련 안내서 준수를 위해 AI 시스템 운영 · 관리 지침에 가이드라인 및 안내서 준수 관련 사항을 명시하였으며 이를 문서화하여 내규에 따라 문서 관리번호 등으로 관리하고, 내용 변경시 수시로 개정할 수 있도록 하고 있다.

기본원칙 1-2

점검항목

가이드라인과 관련 안내서 준수를 위한 교육을 시행하고 있는가?

필요성

가이드라인과 관련 안내서의 내용을 실천하도록 유도하기 위한 교육이 필요하다.

체크리스트

1) 가이드라인과 관련 안내서에 관한 교육이 적절하게 운영·관리되고 있는가?

- AI 시스템 운영 조직과 사내 교육 담당 조직의 책임자가 가이드라인과 관련 안내서에 대한 교육 계획을 수립·운영하고 있는지 확인한다.
- 금융회사 내 신규 인력의 교육 여부, 기존 교육 대상자의 재교육 필요기한 관리, 주기적인 교육 시행 등을 확인한다.

YES ☐ | NO ☐

2) 교육 내용에 AI 시스템의 올바른 사용을 유도하기 위한 설명이 제공되는가?

- AI 윤리원칙, AI 시스템 도입 목적과 활용 방안, 대고객 설명 방식, 고객정보의 이용 범위 등 가이드라인과 안내서의 주요 점검항목이 교육 내용에 포함되어 있는지 확인한다.

YES ☐ | NO ☐

준수사례 (예시)

- A카드사의 사내 교육 담당 부서에서는 매년 1회 이상 AI 시스템을 운영하고 있는 담당 인력들에게 가이드라인과 안내서에 대한 교육을 실시하고 있으며, 해당 교육 내용에는 업무 수행 시 가이드라인과 안내서를 준수하는 요령에 대한 구체적인 내용이 포함된다.

2. 거버넌스의 구축

기본원칙 2-1

AI 활용에 관한 윤리원칙과 기준을 수립해야 하며, 금융회사 등은 AI 윤리원칙과 기준에 맞는 조직 관리를 위하여 AI 윤리위원회를 별도로 설치할 수 있다.

점검항목

조직이 추구하는 가치와 주된 AI 활용 맥락 등을 고려하여 AI 활용에 관한 윤리원칙과 기준을 수립하였는가?

필요성

조직 내 AI 윤리원칙과 AI 활용 목적, 상황 등을 종합적으로 고려하고, AI 단계별 구성원의 역할을 정의함으로써 조직이 추구하는 가치에 맞는 AI 활용에 관한 윤리원칙과 기준을 수립하기 위한 점검이 필요하다.

※ 사람이 중심이 되는 인공지능(AI) 윤리기준(과기정통부, '20.12.23.)
- 3대 기본원칙 : ①인간 존엄성 원칙 ②사회의 공공선 원칙 ③기술의 합목적성 원칙
- 10대 핵심요건 : ①인권보장 ②프라이버시 보호 ③다양성 존중 ④침해금지 ⑤공공성 ⑥연대성 ⑦데이터 관리 ⑧책임성 ⑨안전성 ⑩ 투명성

체크리스트

1) AI 활용에 관한 윤리원칙을 수립하였는가?
- 조직 내 AI 윤리원칙을 수립하여 내부 업무지침 등에 반영하였는지 확인한다.

YES □ | NO □

2) AI 윤리위원회 설치 필요성을 검토하고 관련 지침을 마련하였는가?
- 위험평가 결과를 반영하여 AI 윤리위원회 설치 필요성을 검토하였는지 확인한다.
- AI 윤리위원회 설치 필요성이 인정된 경우, 위원회 운영 규정과 의결사항 반영절차 등 관련 지침이 마련되었는지 확인한다.

YES □ | NO □

준수사례 (예시)

- A카드사는 조직이 추구하는 가치에 맞는 윤리원칙과 기준 수립을 위해 조직의 윤리원칙과 AI 활용목적, 상황 등을 종합적으로 검토하고, 단계별(❶기획/설계, ❷개발, ❸평가 및 검증, ❹도입·운영 및 모니터링) 구성원의 역할을 정의하고 있다.

  예시 AI 활용 목적, 상황 등을 고려하여 조직의 윤리원칙의 합리적인 적용, AI 활용 단계별 구성원 역할 정의를 위한 규정 마련 등

- B은행은 AI 시스템 잠재적 위험 평가 결과에 따라 AI 윤리위원회 설치 필요성을 검토하였으며, 검토 결과에 따라 필요성이 인정되어 AI 윤리위원회를 별도로 설치하고 운영 중이다.

- C증권회사는 FDS AI 시스템에 대한 위험 평가결과 개인권리에 중대한 위험을 초래하지 않아 AI 윤리위원회 설치가 불필요하다고 검토된 경우에는 AI 윤리위원회를 별도로 설치하지 않을 수 있음

기본원칙 2-2

금융회사 등은 AI 시스템의 전 과정에 걸쳐 AI 활용에 따라 나타날 수 있는 잠재적 위험을 인식 · 평가하고, 이를 관리 · 최소화하는 방안을 검토하는 등 AI 활용으로 인한 잠재적 위험을 관리하는데 필요한 위험관리정책을 마련해야 한다.

점검항목

금융회사 등은 AI 활용 전 과정에 걸쳐 나타날 수 있는 잠재적 위험을 평가하고, 이를 관리하기 위한 정책을 마련하였는가?

필요성

AI 시스템의 위험을 관리하기 위한 내부의 평가 · 관리정책 마련 여부 확인이 필요하다.

체크리스트

1) 금융소비자에 대한 영향력, 위험 요소 등을 고려하여 시스템의 위험 수준을 평가할 수 있도록 기준을 마련하였는가?

- 마련된 평가기준을 통해 AI 시스템의 위험 수준이 명확히 구분될 수 있는지 확인한다.

YES ☐ | NO ☐

2) 위험 수준 평가 기준과 수준별 준수사항 등이 포함한 위험관리 정책이 마련되어 있는가?

- 위험관리정책에 금융소비자에게 미치는 영향, 위험 요소 등을 평가하여 고위험·중위험·저위험 등으로 분류하고, 위험 수준별 안내서 준수사항을 정의하였는지 확인한다.
- 위험관리 정책의 제 · 개정 시, CRO 등 책임있는 업무수행이 가능한 자 또는 내부 위원회(관련 부서장 3인 이상 등)의 승인 절차가 마련되어 있는지 확인한다.

YES ☐ | NO ☐

3) 시스템 개발 · 운영 단계별 구성원의 역할 등을 정의하였는가?

- 기획/설계, 개발, 평가 및 검증, 도입·운영 · 모니터링 등 각 단계별 구성원의 역할이 명확히 정의되어 있는지 확인한다.

YES ☐ | NO ☐

4) 위험관리정책을 연 1회 이상 주기적으로 검토하여 최신성을 유지하고 있는가?

- 최소 연 1회 이상 정기적인 위험관리정책을 검토할 수 있는 내부 절차가 마련되어 있는지 검토한다.

YES ☐ | NO ☐

준수사례 (예시)

- A은행은 AI 시스템의 위험 수준을 평가할 수 있는 기준을 마련하고, 개인 권리에 중대한 위험을 초래할 수 있는지에 대한 평가 결과를 바탕으로 위험 수준별 관리하기 위한 위험관리정책을 마련하였다. 평가는 매년 1회 이상 주기적으로 수행한다.
- B은행은 아래와 같은 기준으로 신용평가·여신심사 시스템의 위험수준을 판단한다.

  - 고위험 : AI가 신용평가·여신심사의 의사결정을 전면적으로 대체하는 경우
  - 중위험 : 의사결정과정에 사람이 개입하지만, AI시스템이 고객에게 부정적 영향을 미칠 가능성이 있는 경우
  - 저위험 : 의사결정과정에 사람이 개입하며, 고객에게 유리한 방향으로만 AI시스템이 활용되는 경우

기본원칙 2-3

금융회사 등이 개인에 대한 부당한 차별 등 개인의 권익과 안전, 자유에 대한 중대한 위험을 초래할 수 있는 서비스(고위험 서비스)에 대해 AI 시스템을 활용하는 경우, 적절한 내부통제 · 승인 절차를 마련하고 승인 책임자를 지정한다.

점검항목

고위험 서비스에 AI를 활용하기 위한 승인 절차와 승인책임자 지정 등 적절한 내부통제가 마련되어 있는가?

※ 기관 내 현재 운영중이거나 도입 예정인 AI 시스템 중 전부 또는 일부가 고위험 서비스에 해당되는 경우, 점검항목의 확인이 필요하며, 모든 AI 시스템이 고위험 서비스에 해당되지 않은 경우, 본 점검항목은 생략 가능하다.

필요성

고위험 서비스의 경우, 금융소비자에게 중대한 위험을 초래할 수 있으므로, 이에 대한 적절한 위험관리가 이루어지고 있는지 확인할 필요가 있다.

체크리스트

※ 기관 내 AI 시스템 중 고위험 서비스에 해당되는 경우가 복수인 경우, 해당 서비스별로 각각 체크리스트를 점검한다.

1) 해당 AI 서비스가 고위험 서비스에 해당되는가?

- "고위험 AI 서비스"는 금융회사 등이 개인에 대한 부당한 차별 등 개인의 권익과 안전, 자유에 대한 중대한 위험을 초래할 수 있는 경우에 한한다.

  예시 챗봇 등을 통한 고객 응대에 있어 소수자/취약계층 차별 발언, 신용평가/대출심사/보험심사에 있어 비금융이력자, 특정 성별과 연령대 등의 부당한 차별 등

YES ☐ | NO ☐ | N/A ☐

2) 고위험 서비스의 경우, 적절한 내부통제가 마련되어 있는가?

- 고위험 서비스를 도입하는 경우, CRO 등 책임있는 업무수행이 가능한 자 또는 관련 위원회가 검토 후 승인할 수 있는 내부절차가 마련되어 있는지를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 고위험 서비스의 경우, 주요 내용 변경 시 내부통제 · 승인 절차를 거치는가?

- 주요 내용 변경시, 고위험을 통제하기 위한 내부통제 · 승인 절차를 준수하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A보험사는 개인의 권리에 중대한 위험을 초래할 수 있는 고위험 서비스를 관리하기 위해, 승인 책임자 지정 또는 관련 위원회를 구성하여 내부통제 및 승인 절차를 준수하고 있으며 고위험 서비스의 주요 내용 변경 시 내부통제 절차를 따라 수행한다.

3. AI 업무위탁에 대한 특례

기본원칙 3-1

금융회사 등(위탁기관)은 AI 시스템의 개발·운영 등을 외부기관에 위탁할 경우, 수탁기관이 위탁기관의 AI 윤리원칙 및 위험관리정책을 준수할 수 있는 AI 시스템 개발·운영 위험관리지침을 마련하도록 하여 금융회사 등이 직접 AI 시스템을 개발·운영하는 경우에 비해 위험이 확대되지 않도록 한다.

점검항목

금융회사 등이 AI 시스템의 개발 · 운영 등을 외부에 위탁하는 경우, 수탁기관이 금융회사 등의 AI 윤리원칙 및 위험관리정책을 준수하게 하기 위한 내부 위험관리지침을 마련하였는가?

※ 금융회사 등이 현재 운영중이거나 도입 예정인 AI 시스템 중 전부 또는 일부의 개발 · 운영을 외부에 위탁하는 경우, 본 점검항목의 확인이 필요하며, AI 시스템의 개발 · 운영의 전 과정을 내부에서 수행하고 외부에 위탁하지 않는 경우, 본 점검항목은 생략할 수 있다.

필요성

AI 시스템 개발 · 운영에 따른 위험이 확대되지 않도록 수탁기관이 금융회사 등의 AI 윤리원칙 및 위험관리정책을 준수하고 있는지 확인할 필요가 있다.

체크리스트

1) AI 시스템의 개발 · 운영 업무 수탁기관을 선정하기 위한 합리적인 기준이 마련되어 있는가?
- 내부 업무지침 등을 통해 수탁기관 선정에 대한 합리적인 기준과 방법이 마련되어 있어야 한다.
- 수탁기관을 공정하고 합리적인 기준에 의해 선정했음을 확인할 수 있는 문서를 검토한다.

YES □ | NO □ | N/A □

2) 수탁기관의 위험관리지침이 금융회사 등의 AI 윤리원칙 및 위험관리정책에 관한 사항을 규정하고 있는지 여부를 사전에 확인하도록 하고 있는가?
- 내부 업무지침 등에 수탁기관의 위험관리지침을 사전에 확인하는 절차가 마련되어 있어야 한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A은행은 금융기관의 업무위탁 등에 관한 규정(금융위원회 고시 제2019-12호)을 사내 업무지침 등에 반영하여 준수하고 있으며, 추가로 수탁기관이 지켜야 할 AI 윤리원칙과 위험관리정책을 업무지침에 명시하여 이를 준수하도록 하고 있다.
- B보험회사는 수탁기관 선정에 대한 내부 업무규정을 준수하여, 자체 입찰 공고를 통해 공정하게 예비 수탁기관을 평가 · 검토 후 선정하고 있다.

기본원칙 3-2

금융회사 등은 외부기관에 의한 AI 시스템 개발·운영이 위험관리지침에 따라 이루어졌는지 주기적인 보고·점검 체계를 구축·운영하고, 고위험 서비스에 대해서는 AI 개발·운영계획 등에 대한 금융회사 등의 사전확인, 소비자 피해 발생시 조치 및 보고 절차 마련 등 엄격한 사전 점검이 이루어질 수 있도록 한다.

점검항목

수탁기관의 AI 시스템 개발 · 운영이 위험관리지침에 따라 이루어졌는지 주기적인 보고 · 점검 체계를 구축하였는가?

※ 금융회사 등이 현재 운영중이거나 도입 예정인 AI 시스템 중 전부 또는 일부의 개발 · 운영을 외부에 위탁하는 경우, 본 점검항목의 확인이 필요하며, AI 시스템의 개발 · 운영의 전 과정을 내부에서 수행하고 외부에 위탁하지 않는 경우, 본 점검항목은 생략할 수 있다.

필요성

수탁기관에 대한 위험관리 체계가 충분히 마련되어 있는지 확인할 필요가 있다.

체크리스트

1) 수탁기관이 금융회사 등의 AI 윤리원칙 및 위험관리 정책을 준수하고 있는지 점검하기 위한 담당자를 지정하였는가?

- 수탁기관 관리를 위한 내부 업무지침에 해당 내용이 반영되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 금융회사 등은 수탁기관이 윤리원칙 및 위험관리 준수 여부를 보고하도록 하는 체계를 구축하고 있는가?

- 수탁기관 관리를 위한 내부 업무지침에 해당 내용이 반영되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 내부 업무규정에 따라 수탁기관별 담당자를 지정하여 수탁기관의 AI 시스템 개발 · 운영 내용을 점검하고 있다.
- B증권사는 AI 시스템을 외부에 위탁하는 경우, 내부 업무규정을 통해 주기적인 보고 · 점검 체계를 마련하여 운영하고 있다.

기본원칙 3-2

점검항목

고위험 서비스에 대해서는 AI 개발 · 운영계획 등에 대한 금융회사 등의 사전확인, 소비자 피해 발생 시 조치 및 보고 절차 등 엄격한 사전 점검이 이루어질 수 있는 방안을 마련하였는가?

※ 기관 내 현재 운영중이거나 도입 예정인 AI 시스템 중 전부 또는 일부의 개발 · 운영을 외부기관에 위탁하는 경우, 본 점검항목의 확인이 필요하며, 모든 AI 시스템의 개발 · 운영을 외부기관에 위탁하지 않는 경우, 본 점검항목은 생략할 수 있다. 또한 외부기관에 위탁하는 경우라 할지라도 고위험 서비스에 해당하지 않는 경우, 본 점검항목은 생략할 수 있다.

필요성

수탁기관에서 수행하는 고위험 서비스에 대한 금융회사 등의 사전 점검절차가 충분히 마련되어 있는지 확인할 필요가 있다.

※ AI 시스템의 개발 · 운영을 외부기관에 위탁하지 않는 경우, AI 업무위탁에 대한 특례와 관련된 본 점검항목은 생략할 수 있다.

체크리스트

1) 고위험 서비스의 AI 개발 · 운영계획에 대한 사전 점검절차가 마련되어 있는가?

- 수탁기관 관리를 위한 내부 업무지침에 해당 내용이 반영되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 고위험 서비스에 대한 소비자 피해 발생 시, 조치 및 보고 절차 마련 등 엄격한 사전 점검이 가능한 방안이 마련되었는가?

- 수탁기관 관리를 위한 내부 업무지침에 해당 내용이 반영되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A카드사는 고위험 서비스의 소비자 피해 가능성을 사전 검토하기 위해 수탁기관의 AI 시스템 개발 · 운영계획을 자체 체크리스트를 활용하여 점검하고 있다.
- B보험회사는 수탁기관의 고위험 서비스 AI 개발 · 운영계획 사전 검토 절차가 상세하게 명시된 내부 업무규정을 마련하여 이를 준수하고 있다.

기본원칙 3-3

금융회사 등과 수탁기관은 AI 시스템 개발·운영 등에 따라 소비자 피해가 발생한 경우 손해배상 지연 등을 방지하기 위한 명확한 책임 조항 및 손해배상 처리 절차 등을 마련한다.

점검항목

금융회사 등과 수탁기관은 소비자 피해방지를 위한 조치를 마련하였는가?

※ 금융회사 등이 현재 운영중이거나 도입 예정인 AI 시스템 중 전부 또는 일부의 개발 · 운영을 외부에 위탁하는 경우, 본 점검항목의 확인이 필요하며, AI 시스템의 개발 · 운영의 전 과정을 내부에서 수행하고 외부에 위탁하지 않는 경우, 본 점검항목은 생략할 수 있다.

필요성

금융소비자 피해방지를 위한 점검절차가 충분히 마련되어 있는지 확인할 필요가 있다.

※ AI 시스템의 개발 · 운영을 외부기관에 위탁하지 않는 경우, 본 점검항목은 생략할 수 있다.

체크리스트

1) 소비자가 피해 발생 등에 대한 이의제기를 할 수 있는 장치가 금융회사 등과 수탁기관에 있는가?

- 소비자 보호와 책임 조항, 손해배상 처리 등의 내부 업무지침과 수탁기관과의 계약서에 소비자 이의제기 절차가 충분히 반영되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 금융회사 등은 AI 개발 · 운영 등에 따른 소비자 피해 방지를 위해 내부 업무지침 등에 AI 시스템에 대한 주기적인 모니터링을 하도록 규정되어 있는가?

- 내부 업무지침과 수탁기관과의 계약서에 AI 시스템에 대한 주기적인 모니터링 절차와 관련된 내용이 충분히 반영되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 금융회사 등과 수탁기관은 AI 개발 · 운영 등에 따른 소비자 피해 발생 시, 소비자 피해 구제방안을 마련하고 시행할 수 있는 담당자가 지정되어 있는가?

- 소비자 보호와 책임조항, 손해배상 처리 등의 내부 업무지침과 수탁기관과의 계약서에 소비자 피해 구제방안과 관련한 수탁기관의 담당자 지정 관련 내용이 충분히 반영되어 있는지 확인한다.

YES □ | NO □ | N/A □

4) 금융회사 등과 수탁기관은 소비자에 대한 손해배상에 대한 명확한 책임 조항 및 손해배상 처리 절차 등을 명확히 규정하고 있는가?

- 소비자 보호와 책임조항, 손해배상 처리 등의 내부 업무지침에 수탁기관과의 손해배상 처리 내용이 충분히 반영되어 있는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A증권사는 고객 창구, 콜센터 등을 통해 소비자의 피해 발생 사실을 확인 및 처리하는 업무체계를 구축하고 있다.
- B은행은 금융소비자에게 피해가 발생한 경우, 내부 업무규정을 통해 이의제기 및 손배배상 절차를 마련해놓고 있으며, 주기적인 모니터링을 통해 소비자 피해가 발생하지 않도록 노력하고 있다.

Chapter 03

5대 서비스별 안내

(기획 · 설계 → 개발 → 평가 · 검증 → 도입 · 운영 · 모니터링)

Chapter 03

AI 개발·활용 안내서 (5대 서비스별)

1. 기본원칙 및 점검항목

가. 기획·설계 단계

기본원칙 가-1

AI 시스템의 활용 목적이 윤리원칙에 부합하는지 검토하고, 활용 맥락을 고려하여 AI 활용으로 나타날 수 있는 사회적, 경제적, 문화적 영향 및 잠재적 피해 가능성을 평가하여야 한다.

가-1-1

점검항목

AI를 활용하는 목적이 명확하게 정의되고, 윤리원칙 부합 여부와 AI 활용에 따른 영향도와 잠재적 피해 가능성을 점검하였는가?

필요성

AI 시스템의 목적이 윤리원칙에서 벗어날 경우, 고객에게 부정적인 영향을 끼칠 수 있으므로, AI 시스템의 기획 및 설계 단계에서 활용 목적의 윤리원칙 부합 여부와 활용으로 인한 영향 및 피해 가능성을 평가해야 한다.

체크리스트

1) AI 도입시, 기존 업무에 대한 영향도 분석 등을 통해 도입 타당성 검토가 이루어졌는가?
YES ☐ | NO ☐ | N/A ☐

2) AI 활용 목적과 업무범위, 역할 등이 명확하게 정의되어 있는가?
YES ☐ | NO ☐ | N/A ☐

3) AI 활용 목적이 해당 기관의 AI 윤리원칙에 부합하는가?
YES ☐ | NO ☐ | N/A ☐

4) AI 시스템의 기획 및 설계 단계에서 고객에게 미치는 영향, 위험수준, 잠재적 피해 가능성 등이 고려되었는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

기본원칙 가-2

AI 시스템이 인간의 의사결정을 전면적으로 대체하거나, 중요 의사결정을 대체하는 경우, 금융회사 등은 AI 시스템을 효과적으로 감독, 통제하고 책임성을 유지할 수 있도록 AI 시스템을 설계한다.

가-2-1

점검항목

AI 시스템이 인간의 의사결정을 전면적으로 대체 또는 중요 의사결정을 대체하는 경우, 감독 · 통제 절차가 마련되어 있는가?

※ AI 시스템 : 특정 목표가 주어진 상태에서 데이터를 획득하여 환경을 인식하고, 획득된 데이터를 해석하며, 지식을 추론하거나 정보를 처리하고, 해당 목표를 달성하기 위한 최선의 행동을 결정함으로써 물리적 또는 디지털 차원에서 작동하는 인간이 설계한 소프트웨어 또는 하드웨어 시스템

필요성

AI 시스템에 대해 감독 및 통제가 되지 않을 경우, AI 시스템의 자동화된 의사결정으로 인한 부작용이 발생할 가능성이 있고, 이에 따른 후속조치에 어려움을 겪을 수 있으므로, 금융회사의 AI 시스템에 감독 · 통제 절차가 마련되어 있는지 확인할 필요가 있다.

※ 단, 금융회사 등은 AI 시스템의 특성을 고려하여 탄력적으로 활용할 수 있으며 AI 시스템이 중요 의사결정을 대체하지 않고, 인간의 개입비중이 높은 경우에는 본 점검항목을 생략할 수 있다.

체크리스트

1) 의사결정을 AI가 대체하는 경우 발생할 수 있는 잠재적 위험을 확인하고, 이에 대한 감독 · 통제 절차가 수립되어 있는가?
YES ☐ | NO ☐ | N/A ☐

2) 의사결정을 AI가 대체하는 경우, 인간의 개입이 필요한 경우에 대한 관련 절차가 정의되어 있는가?
YES ☐ | NO ☐ | N/A ☐

3) 의사결정을 AI가 대체하는 경우, 관리자, 사용자 및 기타 이해관계자가 해당 의사결정 과정에 대해 해석 및 추적이 가능하도록 설계되어 있는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

나. 개발 단계

기본원칙 나-1

AI 시스템에 활용되는 학습데이터의 출처, 품질, 편향성 등을 조사·검증하고 주기적인 데이터 갱신 등 데이터 품질 개선을 위한 방법을 검토한다.

나-1-1

점검항목

AI 시스템에 이용되는 학습데이터의 출처, 품질 등을 검증하고 개선필요시 조치를 취하였는가?

필요성

AI 모형에 학습데이터를 사용하는 경우, 학습데이터 품질은 AI 모형의 결과와 성능에 영향을 미칠 수 있는 요소이므로 학습데이터 수집·가공 절차에 대해 상세하게 확인할 필요가 있다.

체크리스트

1) 학습데이터의 출처와 안정적인 데이터 수집 여부를 점검하였는가?
YES ☐ | NO ☐ | N/A ☐

2) 학습데이터의 품질 확보를 위해 데이터의 대표성·정합성을 체크하였는가?
YES ☐ | NO ☐ | N/A ☐

3) 재학습을 수행할 경우, 학습데이터를 갱신하여 데이터의 최신성을 확보하였는가?
YES ☐ | NO ☐ | N/A ☐

4) 학습데이터의 출처, 사전처리, 가공 등의 주요 과정을 문서화하였는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

나-1-2

점검항목

AI 시스템에 이용되는 학습데이터 또는 모형의 편향여부를 개발단계에서 테스트하고, 이를 완화하기 위한 적절한 조치를 취했는가?

필요성

데이터 편향과 차별을 최소화하여 공정하게 적용하기 위해 편향 완화 방안이 필요하다.

체크리스트

1) AI 판단기준에 차별적인 요소가 들어가지 않도록 사전에 점검하였는가?
YES □ | NO □ | N/A □

2) AI 모형의 편향성 판단지표를 선정하여 개발 단계별 편향수준을 테스트하고, 편향을 완화하기 위한 적절한 조치를 취하였는가?
YES □ | NO □ | N/A □

기본원칙 나-2

개인정보보호법 제23조 및 시행령 제18조에 따른 민감정보 등을 AI 시스템에 활용하는 경우, 개인정보 보호를 위한 사전 동의 획득 또는 비식별조치 등 안전조치를 취하고, 해당 정보 활용의 필요성을 평가하고, 데이터 처리 과정에서 해당 정보의 재식별, 유출, 악용 가능성이 없도록 한다.

나-2-1

점검항목

AI 시스템에 개인정보·민감정보를 활용하는 경우, 해당 정보의 필요성을 평가하고 안전조치를 수행하였는가?

필요성

금융회사는 관련법령(개인정보보호법) 준수와 고객의 개인정보·민감정보를 보호하기 위해 노력해야 한다.

체크리스트

1) AI 시스템에서 개인정보·민감정보의 활용 필요성을 점검하였는가?
YES □ | NO □ | N/A □

2) AI 시스템에서 활용하는 개인정보·민감정보에 대해 안전조치를 취하였는가?
YES □ | NO □ | N/A □

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

기본원칙
나-3

고객에 대한 설명의무가 있는 금융서비스나 고위험 서비스의 경우, AI 시스템의 개발단계에서부터 설명가능성을 고려하고, 가용한 설명가능한 AI 기술 등을 확인하여 이를 도입하기 위한 노력을 기울인다.

나-3-1

점검항목

AI 시스템의 설명가능성을 고려하고, 설명가능한 AI 기술 도입방안을 검토 또는 대안을 마련하였는가?

필요성

금융회사는 금융소비자보호법과 신용정보법 등에 따라 해당 서비스에 대해 고객에 대한 설명의무를 이행해야 하며, AI 서비스를 활용하는 경우, 개발 단계에서부터 설명가능성을 고려하여 고객에게 설명가능한 조치가 이루어질 수 있도록 해야 한다.

※ 본 점검항목은 관련 법령 등에 따라 고객에 대한 설명의무가 있는 금융소비자나 고위험 서비스에 AI 시스템을 활용하는 경우에만 해당되며, AI 모델이 중요 의사결정을 대체하지 않고, 인간의 개입 비중이 높은 경우에는 점검항목을 완화 적용할 수 있다.

체크리스트

1) AI 시스템 개발 과정에서 설명가능성을 확보하기 위해 노력하였는가?
YES ☐ | NO ☐ | N/A ☐

2) AI 시스템 개발 과정에서 설명가능한 AI 기술 적용이 어려운 경우, 대안을 마련하였는가?
YES ☐ | NO ☐ | N/A ☐

3) AI 시스템 개발 과정에서 AI 시스템의 결과를 고객에게 설명하기 위한 절차를 고려하였는가?
※ 본 체크리스트는 관련 법령상에 고객에 대한 설명의무가 없는 경우에 한해 생략 가능하다.
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

다. 평가 · 검증 단계

기본원칙 **다-1**

AI 시스템의 적절한 성능 목표 수준 및 성능 측정 지표를 선정 · 관리해야 한다.

다-1-1

점검항목

AI를 활용한 시스템별 성능 지표 선정 및 목표 수준을 설정하고 충족여부를 확인하였는가?

필요성

AI 시스템의 성능이 일정 수준에 미달할 경우, 윤리원칙 위반, 오류 발생에 따른 고객 영향 및 피해 발생 등의 부작용이 생길 수 있으므로 금융회사는 적절한 목표수준을 설정하여 AI 시스템의 성능을 측정 및 관리하여야 한다.

체크리스트

1) AI 기반 모형의 성능을 평가하기 위한 지표를 선정하고 있는가?
YES ☐ | NO ☐ | N/A ☐

2) 선정한 성능 평가 지표에 따라 목표 수준의 달성 여부를 점검하고 미달하였을 경우 조치하고 있는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

기본원칙 **다-2**

AI 시스템의 적절한 공정성 목표 수준 및 공정성 판단 지표를 선정, 관리한다. 선정된 공정성 판단지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 수 있는 기술적, 관리적 측면에서 노력하여야 한다.

다-2-1

점검항목

불합리한 차별이 나타나지 않도록 공정성 판단기준을 설정하고 충족여부를 평가하고 개선하였는가?

필요성

AI 시스템이 불합리한 요소에 따라 차별적 결과를 도출할 경우, 윤리원칙 위반, 부당한 금융거래 거절 등의 부작용이 생길 수 있으므로 금융회사 등은 AI 시스템의 공정성을 측정 및 관리하여야 한다.

체크리스트

1) 공정성 목표 수준 및 공정성 판단 지표를 선정, 관리하는가?
YES ☐ | NO ☐ | N/A ☐

2) 공정성 판단지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 기술적, 관리적 방안을 검토하여 조치하는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

기본원칙
다-3

금융회사 등은 관련 법령 등에 따라 고객에 대한 설명의무가 있는 금융서비스 등에 AI 시스템을 활용하는 경우 또는 고위험 서비스에 AI 시스템을 활용하는 경우 설명가능 AI 기술 등 적절한 AI 기술을 투명하게 적용하여 맥락에 맞는 설명이 도출되는지 여부를 확인하고, AI 시스템의 안정성·신뢰성 등을 훼손하지 않는 범위 내에서 설명가능성을 합리적인 수준으로 개선하기 위해 노력하여야 한다.

다-3-1

점검항목

AI 시스템이 학습한 모형이 상황에 맞게 설명 가능한지 확인하고, 설명가능성을 적법한(또는 합리적인) 수준으로 개선하고자 노력하였는가?

필요성

고객에 대한 설명의무가 있는 금융서비스의 경우 적법한 수준의 설명가능성 확보가 필요하며, 설명의무가 없는 시스템인 경우에도 AI 시스템의 투명성을 확보하고 오류 및 왜곡을 최소화하기 위해 맥락에 맞는 합리적 수준의 설명가능성을 확보하여야 한다.

체크리스트

1) 고객에 대한 설명의무, 금융서비스의 위험 수준 등을 고려하여 고객 및 이해관계자 등 설명 대상자들에 대해 AI 시스템 설명가능성 수준을 평가할 수 있는 기준 및 절차가 수립되어 있는가?
YES ☐ | NO ☐ | N/A ☐

2) AI 시스템의 설명가능성이 기준이 적법한(또는 합리적인) 수준에 도달하지 못할 경우 개선 방안(재학습이나 알고리즘 수정 등)을 검토하여 조치하고 있는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

라. 도입 · 운영 · 모니터링 단계

기본원칙 라-1

금융회사 등은 대고객 AI 시스템 운영시 고객에 AI 이용 여부, 설명 · 이의 제기권 등 관련 법령에 따른 소비자의 권리, 이의신청, 민원제기 방식 등 AI 시스템 성격에 맞추어 적절한 권리구제 방안을 고지해야 한다.

라-1-1

점검항목

AI시스템 제공에 있어 AI 이용여부, 설명·이의제기권 등 고객의 권리 및 이의신청·민원제기 방법 등 소비자로서의 권리구제 방안을 마련하여 고지하였는가?

필요성

AI 시스템의 활용에 따른 결과는 금융회사의 책임이므로 고객의 설명요구권과 이의제기권에 상응하는 방안을 고지하고, 조치 가능한 장치를 마련해야 한다.

체크리스트

1) AI 시스템 적용 사실과 범위에 대해 고객에게 안내가 이뤄지고 있는가?
YES ☐ | NO ☐ | N/A ☐

2) AI 시스템으로 인한 고객의 불편 및 불이익에 대한 구제 방안이 마련되어 있는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

기본원칙 라-2

AI 시스템의 성능을 주기적으로 모니터링하고 데이터 재학습 필요성 검토 등 성능 개선 가능성을 확인한다.

라-2-1

점검항목

AI 시스템의 성능을 주기적으로 모니터링하며 성능 개선이 필요한지의 여부를 확인하고 있는가?

필요성

AI 시스템의 성능은 신뢰성을 판단하는 중요한 기준이며, 성능이 개발시점 대비 하락하지 않도록 주기적으로 점검 및 개선하여야 한다.

체크리스트

1) AI 시스템 성능이 안정적으로 유지되는지 확인할 수 있는 모니터링 절차를 마련하였는가?
YES ☐ | NO ☐ | N/A ☐

2) 유의미한 성능 하락 및 모집단 특성 변화에 따라 모형을 변경 또는 개선 필요시 의사결정 단계나 절차를 사전에 정의하고 있는가?
YES ☐ | NO ☐ | N/A ☐

3) AI 재학습 및 모형 개선 절차를 수립·이행하고 있는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

기본원칙
라-3

AI 시스템에 고객 또는 제3자에 의한 데이터 오염 공격, 적대적 공격 등 오용, 악용 가능성을 최소화할 수 있는 방안을 도입하기 위해 노력한다. 오픈소스 기반 AI 개발 프레임워크 등 AI 개발 환경의 보안 취약성에 관해 상시적으로 통지를 받는 절차를 반영하고 최선의 보안시스템을 구축하기 위해 노력하여야 한다.

라-3-1

점검항목

AI 시스템 및 AI 시스템에서 사용하는 데이터의 오용·악용 가능성을 최소화하였는가?

필요성

고객 정보와 권리를 보호하기 위해 AI 시스템에 고객 또는 제3자에 의한 데이터 오염 공격, 적대적 공격 등 오용·악용 가능성이 있는지 여부를 검토하여 최선의 보안 시스템을 구축했는지 확인하여야 한다.

체크리스트

1) 보안 시스템 구축 등을 통해 AI 시스템에 대한 보안대책을 수립하여 적대적 공격 등 오용·악용 가능성을 최소화하고 있는가?
YES ☐ | NO ☐ | N/A ☐

2) 보안 시스템 구축 등을 통해 고객 또는 제3자에 의한 데이터 오염 공격 등을 통한 데이터의 오용·악용 가능성을 최소화하고 있는가?
YES ☐ | NO ☐ | N/A ☐

체크리스트

3) 오픈소스(프레임워크, 라이브러리 등) 보안 취약성 관리를 위한 체계를 수립하여 AI 시스템의 보안성을 강화하고 있는가?
YES ☐ | NO ☐ | N/A ☐

4) 침해사고 및 재해 등을 예방하기 위한 체계 및 침해사고 또는 재해가 발생했을 때 피해 확산 · 재발 방지와 신속한 복구를 위한 체계를 갖추고 있는가?
YES ☐ | NO ☐ | N/A ☐

※ 체크리스트의 질문에 대해 각 서비스별로 판단기준과 준수사례를 명시함

2. 서비스별 체크리스트

A. 신용평가 및 여신심사

체크리스트 가-1-1-A

AI를 활용하는 목적이 명확하게 정의되고, 윤리원칙 부합 여부와 AI 활용에 따른 영향도와 잠재적 피해 가능성을 점검하였는가?

체크리스트

1) AI 도입 시, 기존 업무에 대한 영향도 분석 등을 통해 도입 타당성 검토가 이루어 졌는가?
- 기존 신용평가 · 여신심사 업무와의 비교 및 도입 시 기대효과 등을 통해 업무 영향도 분석 및 타당성 검토가 이루어졌는지 확인한다.

예시 AI 도입 전·후 업무 절차, 담당자 역할 변화, 세부 업무별 AI 적용방식 적용 여부 등

※ 기존 업무가 존재하지 않는 경우, 생략 가능하다.

YES ☐ | NO ☐ | N/A ☐

2) AI 활용 목적과 업무범위, 역할 등이 명확하게 정의되어 있는가?
- 신용평가 · 여신심사를 위한 업무 중 AI 시스템을 활용한 업무 범위, 활용 목적, AI의 역할 등이 명확하게 정의되어 있는지 확인한다.

예시 대안정보를 활용한 개인사업자 신용평가시 AI를 활용, 여신 관리대상(감액 또는 회수 등) 선정시 AI를 활용

YES ☐ | NO ☐ | N/A ☐

3) AI 활용 목적이 해당 기관의 AI 윤리원칙*에 부합하는가?

* 인권보장, 프라이버시 보호, 다양성 존중, 공공성, 책임성, 안전성, 투명성 등

- 도입할 신용평가·여신심사를 위한 AI 시스템이 내부 AI 윤리원칙 등에 부합하는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

4) AI 시스템의 기획 및 설계 단계에서 고객에게 미치는 영향, 위험수준, 잠재적 피해 가능성 등이 고려되었는가?

- 신용평가·여신심사를 위한 AI 시스템이 내부 AI 윤리원칙에 근거한 각 금융회사의 절차에 따라 필요시 AI 모형의 위험수준 평가기준이 마련되어 있는지 확인한다.
- 모형 기획·설계 단계에서 신용정보법 제22조의4(개인 신용평가회사의 행위규칙) 등에 근거하여 신용평가 시, 차별적 요소가 사용되지는 않는지, 또는 정책적으로 사용이 금지된 항목이 반영되어 고객에게 잠재적 피해가능성이 존재하는지 확인한다.
- 모형 기획·설계 단계에서 연체경험 여부, 금융이력 정보 보유수준, (Thin-Filer, Thick Filer) 등 금융거래 특성에 따라 특정 그룹의 고객에게 과도한 혜택이나 피해 가능성이 있는지 점검한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 AI 모형의 기획 및 설계 단계에서 AI 도입 타당성을 검토한 후, 도입이 승인되면 AI 활용목적에 대해 가이드라인에 근거한 AI 윤리원칙에 어긋나는 부분이 없는지 확인하고 있으며, AI 활용으로 인해 발생할 수 있는 잠재적 위험과 고객에게 미치는 영향에 대해 사전 평가를 진행한다.
- B은행은 AI 기반 신용평가 모형 설계 단계에서 모형에 활용될 가능성이 있는 후보변수 선정 시, 개인 신용평가 관련 행정지도 사항, 한국신용정보원 일반신용정보관리규약 등이 충실하게 반영되었는지 확인하고 있다. 또한 이 과정에서 고객의 잠재적 피해 가능성을 점검하여 피해가 우려되는 경우, 후보변수에서 제외하는 절차를 마련해두고 있다.
- C금융지주는 전 계열사의 여신관리 정책을 점검하여 신용평가·여신심사에 AI를 도입하는 경우, 고객에게 발생할 수 있는 영향과 위험수준을 고려하여 AI 기반 모형의 도입을 검토하고 있다.

체크리스트 가-2-1-A

AI 시스템이 인간의 의사결정을 전면적으로 대체 또는 중요 의사결정을 대체하는 경우, 감독 · 통제 절차가 마련되어 있는가?

체크리스트

※ AI시스템이 인간의 의사결정을 대체하는 경우가 없다면, 동 체크리스트는 "N/A" 처리할 수 있다.

1) 의사결정을 AI가 대체하는 경우 발생할 수 있는 잠재적 위험 가능성을 확인하고, 이에 대한 감독 · 통제 절차가 수립되어 있는가?

- AI를 활용하는 경우에도 기존 신용평가 · 여신심사 업무에서 일반적으로 준수하고 있는 내부통제 절차가 수행되고 있는지 확인한다.
- AI 활용에 관한 내부 업무규정에 마련된 위험평가기준과 금융소비자 권리에 중대한 위험이 발생할 가능성을 기준으로 위험평가 결과의 적정성을 확인한다.
- 내부 규정 또는 업무 매뉴얼을 통해 운영 중인 AI 시스템의 위험평가 결과 보고 절차의 마련 여부와 결과 보고 및 승인 여부를 확인한다.
- AI가 기존 신용평가 · 여신심사를 대체하는 경우 발생할 수 있는 잠재적 위험에 대해 예측하여 사전 · 사후 대책이 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 업무별 담당자 지정 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 의사결정을 AI가 대체하는 경우, 인간의 개입이 필요한 경우에 대한 관련 절차가 정의되어 있는가?

- AI 기반의 신용평가 · 여신심사 의사결정에 인간의 개입이 필요한 경우에 대한 업무 매뉴얼 또는 관련 내부절차가 마련되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 의사결정을 AI가 대체하는 경우, 관리자, 사용자 및 기타 이해관계자가 해당 의사결정 과정에 대해 해석 및 추적이 가능하도록 설계되어 있는가?

- AI 시스템과 대출업무 프로세스간 연계사항 확인, AI 시스템의 결과에 대한 설명 방안 설계 여부 등 자동 의사결정에 대해 추적 및 설명방안 마련이 이루어졌는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 AI 기반 모형을 신용평가·여신심사에 활용하는 경우, 위험관리 위원회의 승인을 득하도록 하고 있으며, 모형개발부서와 통제·검증업무 부서를 독립적으로 운영하고 주기적으로 모형에 대한 검증을 진행하며 이에 대한 검증결과를 이사회·경영진에 보고한다.
- B카드사는 AI 신용평가 결과에 중요한 영향을 미치는 정보가 오류인 것으로 확인된 경우, 신용등급조정위원회 또는 여신심사위원회의 심의를 받아 해당정보를 갱신하고 결과를 재산출하는 절차를 마련하였다.

체크리스트 나-1-1-A

AI 시스템에 이용되는 학습데이터의 출처, 품질 등을 검증하고 개선필요시 조치를 취하였는가?

체크리스트

1) 학습데이터의 출처와 안정적인 데이터 수집 여부를 점검하였는가?

- AI 기반 신용평가 · 여신심사 모형의 학습시, 학습데이터 생성에 사용되는 원천데이터의 출처의 신뢰성과 데이터 수집의 안정성을 확인한다.
- 신용평가회사, 신용정보원 등 데이터 공급사와의 공급방식 및 공급내역을 확인한다.
- 개인정보보호법을 준수하여 데이터 수집 시, 제공자에게 수집 사실을 고지하고 동의를 얻어야 한다.
- AI 시스템 개발 시 고객이 일부 데이터에 대해 사용 동의를 철회한 경우 이를 반영할 수 있는 로직을 검토한다.
- 데이터 품질 관리를 위한 거버넌스 조직 또는 담당자 지정 여부를 확인하고, 양질의 데이터가 왜곡없이 제공되는지 점검한다.

YES □ | NO □ | N/A □

2) 학습데이터의 품질 확보를 위해 데이터의 대표성 · 정합성을 체크하였는가?

- 학습데이터가 AI 시스템이 적용될 신용평가의 대상(모집단)을 대표하는지 확인한다. (개발 모집단 선정기준 및 개발대상 샘플 추출기준 점검)
- 데이터 품질관리에 대한 내규 존재 여부를 확인한다.
- 학습데이터 생성에 대한 라벨링 작업 시 원천데이터와 라벨링 데이터의 동기화 여부를 확인한다.

- 사용 데이터의 정합성 검증 여부를 확인한다.
- 학습 데이터에 대한 조직의 품질 관리역량 강화를 위한 교육 및 지원 체계의 확보 여부를 확인한다.

YES □ | NO □ | N/A □

3) 재학습을 수행할 경우, 학습데이터를 갱신하여 데이터의 최신성을 확보하였는가?

- 내부 규정 또는 업무매뉴얼 등을 통해 재학습 및 재배포 절차 마련 여부를 확인한다.
- 재학습 및 재배포 절차가 데이터 최신성 및 적정성 유지에 적합한지 확인한다.

YES □ | NO □ | N/A □

4) 학습데이터의 출처, 사전처리, 가공 등의 주요 과정을 문서화하였는가?

- 데이터의 수집과 처리 업무를 위한 절차로서 수집처리 방법 및 기준에 대한 내부 규정 또는 업무매뉴얼에 반영하였는지 확인한다.
- 신용평가회사나 신용정보원 등 기존 신용정보 인프라 기관 외에 추가로 외부 데이터를 활용(예: 대안 신용정보)하는 경우 데이터 출처에 대한 신뢰성을 확인한다.
- 개별 데이터 처리에 대한 처리 로직기록 등 통한 데이터처리(변환/합성 등)에 대한 기록 보관 및 유지 여부를 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A은행은 AI 모형 개발에 사용되는 학습데이터 구성시, 원천데이터의 출처와 입수 안정성 등을 사전에 점검하고 있으며, 학습데이터의 정합성 점검 (예: 분포 확인, 이상치 값 제외 처리 등)과 최신성 확보 등을 위한 데이터 품질관리절차를 수행하고 있다. 데이터 명세 자료는 문서화하여 내규에 따라 관리하고, 내용 변경시 수시로 개정한다.

체크리스트 나-1-2-A

AI 시스템에 이용되는 학습데이터 또는 모형의 편향여부를 개발단계에서 테스트하고, 이를 완화하기 위한 적절한 조치를 취했는가?

체크리스트

1) AI 판단기준에 차별적인 요소가 들어가지 않도록 사전에 점검하였는가?

- 불합리한 차별을 방지하기 위하여 성별, 학력, 연령, 지역, 종교, 인종 등 불합리한 차별이 발생 가능한 평가요소를 선정하고 평가요소에 반영되었는지 점검한다.

YES □ | NO □ | N/A □

2) AI 모형의 편향성 판단지표를 선정하여 개발 단계별 편향수준을 테스트하고, 편향을 완화하기 위한 적절한 조치를 취하였는가?

- 내부 정책에 따른 편향성 판단지표를 선정하였는지 확인한다.

예시 모집단과 샘플데이터의 연령 · 성별별 구성비 등을 비교하여 유사한 분포를 이루는지 확인

- 개발표본 선정, 라벨링, 모델링 단계 등 세부 단계별로 편향이 존재하는지 점검한다.
- 업무 매뉴얼, 요구사항 문서 등을 통해 데이터 편향 완화 방안 마련 및 적용 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 신용평가모형 개발 시 AI 활용 여부와 관계없이 성별·학력·연령·지역·종교·인종 등 차별적인 항목을 사전에 검토하여 학습데이터 항목에서 배제하고 있으며, 내부적인 편향성 판단지표를 선정하여 AI 모형의 편향 수준을 점검한다.

- B은행은 신용평가모형 개발시 모집단과의 분포차이를 확인하고, PSI지수가 일정 값을 초과할 경우 모델을 다시 점검한다.

체크리스트 나-2-1-A

AI 시스템에 개인정보·민감정보를 활용하는 경우, 해당 정보의 필요성을 평가하고 안전조치를 수행하였는가?

체크리스트

1) AI 시스템에서 개인정보·민감정보의 활용 필요성을 점검하였는가?

- AI 시스템의 활용 목적, 적용대상을 고려하여 개인정보·민감정보 활용의 필요성을 점검하고, 반드시 필요한 경우에만 활용한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템에서 활용하는 개인정보·민감정보에 대해 안전조치를 취하였는가?

- AI 시스템의 개발을 위해 개인정보·민감정보를 수집·활용하는 경우 개인정보보호법 등 관련 법령을 준수하고, 활용 범위와 목적에 대한 명확한 기준을 마련했는지 확인한다.
- AI 시스템 개발 시 관련 정보의 유출, 악용 가능성이 없도록, 기술적·물리적 통제 방안을 마련해야 한다.

- 개인정보·민감정보 파기에 관해 관련 사항이 준수되고 있는지 확인한다.
- 개인정보·민감정보의 수집·학습이 예상되는 경우, 초기 개발 단계에서 개인정보 수집과 생성에 대한 시스템을 점검할 수 있는 내부 절차를 구축했는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

〈 관련 법령에 따른 별도 체크리스트 〉

3) 신용정보법 제34조의3(정보활용 동의등급) 적용 대상기관의 경우 AI 기반 신용평가 · 여신심사 과정에서 수집·이용하는 개인정보·민감정보와 관련하여 정보활용 동의등급을 부여받고 정보활용주체에게 안내하였는가?

- 신용정보법 제34조의3(정보활용 동의등급)에 따라, 신용정보주체에게 미칠 위험요소와 이익을 고려하여 정보활용 동의등급을 부여받았는지 확인한다.
- 신용정보주체에게 정보활용 동의등급을 알리고, 정보활용 동의사항을 글자크기나 줄간격을 확대하는 등의 방법으로 표기해야한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 개인 신용평가 및 여신심사 과정에서 활용되는 기초정보를 점검하여, 민감정보를 보유한 경우 대체할 수 있는 정보는 없는지 필요성을 확인하고, 필요한 경우 해당 항목을 비식별화 조치하여 이용한다.

체크리스트 나-3-1-A

AI 시스템의 설명가능성을 고려하고, 설명가능한 AI 기술 도입방안을 검토 또는 대안을 마련하였는가?

체크리스트

1) AI 시스템 개발 과정에서 설명가능성을 확보하기 위해 노력하였는가?

- AI 서비스 성격을 감안하여 설명 대상 및 설명하는 절차를 검토한다.
- 결과의 오해석 방지를 위해 설명 공유 대상과 범위를 설정한다.
- 설명가능한 AI 기술(예: Surrogate Model, LIME, SHAP 등)을 검토한다.
- 고객이 결과에 대한 설명을 요청하였을 때, 주요 기준 등에 대해 출력 할 수 있도록 개발과정에 반영한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템 개발 과정에서 설명가능한 AI 기술 적용이 어려운 경우, 대안을 마련하였는가?

- 설명가능한 AI 기술 개발 트렌드를 충분히 확인하여 적용이 어려운 타당한 이유와 이에 대한 대안이 충분히 마련되었는지 확인한다.
- 업무 매뉴얼, 요구사항 문서 등에 AI 시스템이 생성한 결과를 설명하기 위한 기법 검토 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) AI 시스템 개발 과정에서 AI 시스템에 의한 결과를 고객에게 설명하기 위한 절차를 고려하였는가?

- 개인신용평가회사와 금융회사(금융위원회의 설치 등에 관한 법률 제38조 각 호에 해당하는 자)의 경우, 신용정보법 제36조의2(자동화평가 결과에 대한 설명 및 이의제기 등)의 준수 여부를 확인한다.

  ※ 개인인 고객은 AI 시스템에 의한 신용평가 · 여신심사 결과에 대해 평가의 결과, 평가의 주요 기준, 평가에 이용된 기초정보 등을 설명해줄 것을 요구할 권리가 있음

- 고객에게 신용평가 · 여신심사의 결과를 설명할 수 있는 절차가 마련되어 있으며 고객에게 적절한 안내가 이루어지고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 AI 신용평가모형 개발과정에서 설명가능한 인공지능 기술을 적용할 수 있을지 우선 검토하고 있으며, 현재 운영중인 개인 신용평가모형의 경우, 평점표 형태의 Surrogate Model을 활용하여 머신러닝의 설명가능성을 확보하고 있다.

- B은행은 고객이 개인신용평가결과에 대해 설명을 요청할 경우, 주요 거래내용 판단정보(대출·신용카드 사용내역 등), 신용도판단정보(연체 등), 신용거래능력 판단정보(연소득 등) 등 신용평가에 영향을 준 기초정보의 개요와 정정·재산출을 요청할 수 있는 방법을 안내한다.

체크리스트 다-1-1-A

AI를 활용한 시스템별 성능 지표 선정 및 목표 수준을 설정하고 충족여부를 확인하였는가?

체크리스트

1) AI 기반 모형의 성능을 평가하기 위한 지표를 선정하고 있는가?

- AI 모형 성능에 대한 유지 목표 및 평가지표 선정하고, 해당 모형이 적정한 성능으로 운영되고 있는지 확인한다.

예시 AR 일정 이상, PSI 일정 미만을 유지하도록 분기별로 점검

YES □ | NO □ | N/A □

2) 선정한 성능 평가 지표에 따라 목표 수준의 달성 여부를 점검하고 미달하였을 경우 조치하고 있는가?

- 성능 평가 결과가 목표 수준에 미달하였을 경우 원인을 분석하여 모형 재개발 여부를 검토하여 필요시 조치한다.
- 특정 학습데이터에 과적합이 되지 않았는지 다양한 테스트를 수행하고, 테스트 데이터에 따른 변동성을 모니터링하고 있는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A은행은 AI를 적용한 신용평가·여신심사 모형을 운영하면서, 모형의 성능을 변별력과 안정성, 신용평점 및 등급의 서열화 등의 측면에서 확인하고 있으며, 변별력은 AR (Accuracy Ratio), K-S 통계량 등을 통해, 안정성은 PSI(Population Stability Index)를 통해 점검한다. 또한 평가지표가 적정기준에 미달하는 경우, 모형의 재개발을 고려한다.
- B은행은 다양한 유형의 데이터 샘플을 이용하여 모형의 성능을 점검하며, 급격한 변별력 하락이 일어나는 경우 과적합으로 판단하고 모형 수정 또는 재학습을 수행한다.

체크리스트 다-2-1-A

불합리한 차별이 나타나지 않도록 공정성 판단기준을 설정하고 충족여부를 평가하고 개선하였는가?

체크리스트

1) 공정성 목표 수준 및 공정성 판단 지표를 선정, 관리하는가?

- 학력, 성별, 연령, 종교 등 불합리한 차별이 발생 가능한 요소를 선정하고, AI 기반 신용평가·여신심사 평가 결과를 확인하여 상기 요소에 따른 불합리한 차별이 발생하고 있지 않은지 점검한다.
- 공정성 목표 수준 및 공정성 판단 지표를 선정 및 관리하고 있는지 확인한다.

예시 데이터 샘플링을 통해 데이터 분포 검증을 수행하는지 확인

YES □ | NO □ | N/A □

2) 공정성 판단지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 기술적, 관리적 방안을 검토하여 조치하는가?

- 편향방지 방법론을 활용하는 등 편향을 완화하기 위한 방안을 검토하였거나 조치하였는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 AI를 적용한 신용평가모형의 결과를 확인하여 성별·학력·연령 등에 따른 차별이 있는지 확인한다. 다른 조건이 유사함에도 학력, 성별, 연령, 종교 등에 따라 신용등급이나 평점에 큰 폭(자체 등급 3단계 이상)의 차이가 있거나, 공정성 목표 수준 또는 판단기준에 미달되는 경우 모형의 수정을 검토한다.

체크리스트 다-3-1-A

AI 시스템이 학습한 모형이 상황에 맞게 설명 가능한지 확인하고, 설명가능성을 적법한(또는 합리적인) 수준으로 개선하고자 노력하였는가?

체크리스트

1) 고객에 대한 설명의무, 금융서비스의 위험 수준 등을 고려하여 고객 및 이해관계자 등 설명 대상자들에 대해 AI 시스템 설명가능성 수준을 평가할 수 있는 기준 및 절차가 수립되어 있는가?

- 내부 규정, 업무매뉴얼 등에 설명가능성 수준에 대한 평가 기준 및 절차가 수립되어 있는지 확인한다.

예시 신용평점 또는 등급 산출결과에 따른 연체 이력, 소득 감소 등 주요 사유를 제공 여부 확인, 개별 샘플 리뷰를 통해 설명가능 인공지능 기술 등이 제공한 주요 사유로 AI 모형의 출력결과가 유의미하게 설명되는지 여부 확인

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템의 설명가능성이 기준이 적법한(또는 합리적인) 수준에 도달하지 못할 경우 개선 방안(재학습이나 알고리즘 수정 등)을 검토하여 조치하고 있는가?

- 평가 기준에 도달하지 못했을 경우 수행했던 개선 방안 검토 및 조치 결과를 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 AI 기반 여신심사 평가모형에 입력된 핵심 데이터들이 결과에 미친 영향력을 백분율(%)로 시각화하고, 증액된 소득을 입력하면 대출 승인으로 변경됨을 제시하는 등 설명가능한 방식을 도입하여 사용하고 있다.

- B은행은 AI 기반 신용평가 모형의 출력결과에 대해 설명가능 인공지능 기술을 적용하여 고객에게 신용평가에 활용된 항목과 해당 항목의 기여도(feature importance) 등을 설명할 수 있는 절차를 마련하였다.

- C은행은 AI 기반 신용평가 모형의 출력결과에 대해 설명가능성 평가결과 합리적인 수준에 미달된 경우, 주요 사유가 설명되지 않음 등의 미흡한 부분을 보완할 수 있는 효과적인 설명가능 인공지능 기술을 재검토하여 설명가능성을 개선하였다.

- D은행은 AI 시스템을 통해 위험 모형 변수를 소비자에게 신속하게 공개하고, 소비자에 대출 결정에 대한 설명을 돕는 모형을 개발하였으며, 이를 통해 고객이 대출금을 상환할 수 있을지를 평가하고 있다. 대출을 거절할 때는 대출이 왜 거절됐는지에 대한 이유를 최근 코드로부터 근거로 제시하고 있다.

체크리스트 라-1-1-A

AI시스템 제공에 있어 AI 이용여부, 설명·이의제기권 등 고객의 권리 및 이의신청·민원제기 방법 등 소비자로서의 권리구제 방안을 마련하여 고지하였는가?

체크리스트

1) AI 시스템 적용 사실과 범위에 대해 고객에게 안내가 이뤄지고 있는가?

- 상품설명서, 홈페이지, 모바일 앱 등의 채널을 통해 AI 시스템 활용 여부와 목적을 쉬운 용어를 사용하여 안내하고 있는지 확인한다.
- 고객의 권리에 대한 사전고지 여부 및 변경 이력을 확인한다.

YES □ | NO □ | N/A □

2) AI 시스템으로 인한 고객의 불편 및 불이익에 대한 구제 방안이 마련되어 있는가?

- 내부 규정 또는 업무 매뉴얼에 신고 및 이의제기 절차가 구체적으로 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 전담 조직 및 인력을 확인한다.
- 홈페이지, 모바일 앱 등의 채널을 통해 고객에게 적절한 권리구제 절차 및 기준 등 사전고지 여부를 확인한다.
- 신용정보의 이용 및 보호에 관한 법률 제36조의2(자동화평가 결과에 대한 설명 및 이의제기 등)를 준수하여 ❶자동화평가 실시여부, 평과결과·기준·기초정보 등에

대한 설명요구권, ❷자동화평가 결과 산출에 유리하다고 판단되는 정보를 제출할 수 있는 정보제출권, ❸기초정보의 정정·삭제, 결과 재산출을 요구 할 수 있는 이의제기 권을 포함한 권리구제방안을 마련하고, 관련 내용을 안내하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 AI가 심사하는 비대면 대출상품을 판매하면서 상품안내 페이지에 AI 시스템을 통해 여신심사를 진행한다는 내용을 고지하고 있으며, 평가 시 주요 평가기준과 기초정보 개요 등에 대해 설명을 요청하거나 활용된 정보를 정정·삭제를 요구할 수 있는 절차를 마련·운영하고 있다.

- B은행은 고객이 대출을 받은 후 대출 조건 등에 불만이 있을 경우 금융회사의 소비자 지원부서나 인터넷 홈페이지에 문의하거나, 금융감독원 등에 도움을 요청할 수 있는 절차를 상품설명서 내에 안내, 운영하고 있다.

체크리스트 라-2-1-A

AI 시스템의 성능을 주기적으로 모니터링하며 성능 개선이 필요한지의 여부를 확인하고 있는가?

체크리스트

1) AI 시스템 성능이 안정적으로 유지되는지 확인할 수 있는 모니터링 절차를 마련하였는가?

- 성능 모니터링 주기, 범위 및 보고 등의 절차가 마련되어 있는지 확인한다.

예시 개발 시점대비 성능이 크게 변동하지 않는지 매 월 지표 확인

- 성능 목표값에 대한 설정을 사전에 정의하고, 목표값 이하로 하락하는 경우, 파악할 수 있는 체계를 마련한다.

예시 PSI 일정 미만, KS통계량 일정 이상

- 모니터링 결과 및 조치 이력의 문서화 여부 및 기록 내용을 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 유의미한 성능 하락 및 모집단 특성 변화에 따라 모형을 변경 또는 개선 필요시 의사결정 단계나 절차를 사전에 정의하고 있는가?

- 주기적인 성능평가와 재학습 필요성을 검토하고, 필요 시 모형을 교체하거나 개선을 진행해야 하며 모형 변경 또는 개선을 위한 의사결정 단계와 절차가 마련되어 있는지 확인한다.

예시 신용평가의 모집단 특성이 변경되거나 성능지표가 기준치 이하일 경우 재학습 진행

YES ☐ | NO ☐ | N/A ☐

3) AI 재학습 및 모형 개선 절차를 수립·이행하고 있는가?

- AI 재학습 및 모형 개선 절차 수립과 이행 여부를 확인한다.
- 개선 방안 수립 시 데이터의 최신성, 정확성, 정합성, 비편향성을 고려하였는지 확인한다.
- 학습데이터 변경 이력 기록 여부 및 관리 문서를 확인한다.
- 활용된 데이터의 보관기간 준수 여부(5년, 10년 등)를 확인한다.
- 학습데이터가 PSI(모집단 안정성 지표) 등을 통해 관리되고 있는지 여부를 확인한다.
- 학습데이터 상의 설명변수 안정성 테스트 통과 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- D은행은 신용평가·여신심사 시스템의 성능을 변별력, 안정성 지표를 통해 연 1회마다 주기적인 모니터링을 실시하고 있으며, 성능이 유의미한 수준으로 유지되는지 확인한다. 또한 특정 지표 미달 시 시스템 개발 부서에 모형 개선을 권고하고 있다.
- B은행은 AI 기반 신용평가모형에 대해 안정성 지표(PSI)로 적용 대상 모집단 특성이 유사하게 유지되고 있는지 확인하고 있으며, 데이터 재학습이 이루어지는 경우, 학습데이터 품질 점검을 수행하며, 신규 데이터 비율을 관리한다.
- C은행은 정보보호담당자를 지정하고 시스템의 보안점검을 주기적으로 수행하여 보안 문제로 인한 시스템 성능 저하를 예방하고 있다.

체크리스트 라-3-1-A

AI 시스템 및 AI 시스템에서 사용하는 데이터의 오용·악용 가능성을 최소화하였는가?

체크리스트

1) 보안 시스템 구축 등을 통해 AI 신용평가·여신심사 시스템에 대한 보안대책을 수립하여 적대적 공격 등 오용·악용 가능성을 최소화하고 있는가?

- AI 신용평가·여신심사 시스템 구축 시 시스템, 인프라, 데이터, 네크워크, 이용자 보호 등과 관련하여 다양한 보안 위협 및 대응조치를 포함한 적합한 보안대책을 수립하고 검토하고 있는지 확인한다.
- 신용등급 조작 등 비정상 동작이나 예기치 못한 오류에 대한 대책을 수립하고, AI 신용평가·여신심사 시스템에 대한 접근 권한을 관리하고 있는지 확인한다.
- 신용평가·여신심사의 기초정보 등 개인정보 암호화 이행 및 권한별 접근 통제 조치 여부를 확인한다.
- AI 신용평가·여신심사 분류 성능을 주기적으로 점검하여 성능 저하를 유발하는 공격 여부를 확인한다.
- AI 신용평가·여신심사 시스템을 운용하는 중요 서버에 백신 프로그램을 설치하고 주기적 업데이트 및 악성코드 점검, 실시간 검사 설정을 하고 있는지 확인한다.
- 보안 취약점에 대한 사전·정기 점검 수행 및 취약점 조치 여부를 확인한다.
- AI 신용평가·여신심사 시스템 관련 침해사고 분석 시 필요한 로그에 대하여 보존 및 검토에 대한 정책을 수립하였는지 확인한다.
- AI 신용평가·여신심사 시스템 관련 로그별 보존기간 및 검토 주기를 지정하였는지 확인한다.
- AI 신용평가·여신심사 시스템 테스트 시 개인신용정보가 아닌 임의의 데이터를 생성하여 테스트를 수행하는지 확인한다. 단, 개인신용정보 이용이 불가피한 경우, 책임자의 승인 절차에 따라 가공하여 사용하고 즉시 폐기하는 등 관리대책이 수립·이행되었는지 확인한다.
- 암호화 대상, 사용 암호 알고리즘, 암호키 관리 방안을 포함한 개인정보 암호화 정책을 수립하였는지 확인한다.
- 정보자산 등 운영체제, 소프트웨어 패치 관리정책 및 절차를 수립·이행하고 인터넷 직접 접속을 통한 패치를 제한하고 있는지 확인한다.
- AI 신용평가·여신심사 시스템 개발 과정에서 적대적 공격 등 오용·악용 가능성을 최소화하기 위해 보안성 검증을 실시하고 있는지 확인한다.

YES □ | NO □ | N/A □

2) 보안 시스템 구축 등을 통해 고객 또는 제3자에 의한 데이터 오염 공격 등을 통한 데이터의 오용·악용 가능성을 최소화하고 있는가?

- AI 신용평가·여신심사 시스템에 사용되는 데이터 오용 · 악용을 감지 지표 설정 및 주기적 확인 여부 등 절차와 대처방안 수립 · 운영을 통해 관리하고 있는지 확인한다.
- AI 신용평가·여신심사 시스템 운영 시 비정상적 데이터 변동을 점검하기 위한 변동성 감지 지표 마련 및 운영 여부를 확인한다.

예시 전년 동기 대비 데이터 건수 또는 값 변동률에 대한 기준치(예:00% 감소)를 정하고 기준치 초과 시 사유를 확인한다.

- 잘못된 개인신용정보 주입·조작 등 학습데이터 변조 여부를 주기적으로 확인한다.
- AI 신용평가·여신심사 시스템에 사용되는 데이터의 무결성, 기밀성을 유지하기 위한 관리 규정을 준수 여부를 확인한다.

YES □ | NO □ | N/A □

3) 오픈소스(프레임워크, 라이브러리 등) 보안 취약성 관리를 위한 체계를 수립하여 AI 신용평가·여신심사 시스템의 보안성을 강화하고 있는가?

- AI 신용평가·여신심사 시스템 개발에 필요한 오픈소스 사용 시 라이브러리의 보안 취약점을 사전에 확인 · 관리하고 호환성, 라이선스를 확인한다.
- 오픈소스의 알려진 보안 취약점을 주기적으로 확인하여 업데이트 및 패치를 적용하고 있는지 확인한다.
- 신용평가·여신심사 관련 AI 알고리즘의 취약점 점검 여부를 확인한다.

YES □ | NO □ | N/A □

4) 침해사고 및 재해 등을 예방하기 위한 체계 및 침해사고 또는 재해가 발생했을 때 피해 확산 · 재발 방지와 신속한 복구를 위한 체계를 갖추고 있는가?

- 해킹, 악성코드, 논리폭탄, 메일폭탄, 서비스거부 또는 고출력 전자기파 등의 방법으로 사고 발생 시에 대비한 예방·복구 체계 마련 여부를 확인한다.
- 정상적 보호 · 인증절차를 우회하여 정보통신망에 접근할 수 있도록 하는 프로그램이나 기술적 장치 등에 대한 차단 체계 마련 여부를 확인한다.
- 침해사고 발생 시 기록 및 보고, 신고 및 통지, 비상 연락체계 등 내용이 포함된 침해사고 대응 절차를 수립하였는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A은행은 내부통제 절차를 통해 데이터에 접근 권한이 있는 인력을 최소한으로 유지하며, 고객정보도 비식별화(마스킹)하여 관리한다. 또한 기관 내에 개인정보가 포함된 데이터에 대한 관리지침이 마련되어 있다.

- B은행은 입력변수, 목적변수 및 활용되는 전체 데이터에 대한 안정성을 모니터링 하는 체계를 수립하고, 주기적으로 보고하고, 해당 내역을 관리한다.

- C은행은 오픈소스 라이브러리의 의존성 회피 등 호환성을 검토하며, 보안 취약점 사전점검 후 도입하고 있으며, 라이선스는 별도로 관리하고 있다.

- D은행은 전담 조직을 통해 보안점검을 월 1회 이상 수행하고 있으며, 문제발생시 신속하게 조치될 수 있도록 하고 있다.

B. 이상거래 탐지

체크리스트 가-1-1-B

AI를 활용하는 목적이 명확하게 정의되고, 윤리원칙 부합 여부와 AI 활용에 따른 영향도와 잠재적 피해 가능성을 점검하였는가?

체크리스트

1) AI 도입 시, 기존 업무에 대한 영향도 분석 등을 통해 도입 타당성 검토가 이루어졌는가?

- 기존 이상거래 탐지 업무와의 비교 및 도입 시 기대효과 분석 등을 통해 업무 영향도 분석 및 타당성 검토가 이루어졌는지 확인한다.

예시 AI 도입 전·후 업무 절차, 담당자 역할 변화, 세부 업무별 AI 적용방식 적용 여부 등

※ 기존 업무가 존재하지 않는 경우, 생략 가능하다.

YES ☐ | NO ☐ | N/A ☐

2) AI 활용 목적과 업무범위, 역할 등이 명확하게 정의되어 있는가?

- 이상거래 탐지를 위한 업무 중 AI 시스템을 활용한 업무 범위, 활용 목적, AI의 역할 등이 명확하게 정의되어 있는지 확인한다.

예시 이상거래 탐지 유형과 활용목적 따른 AI 역할의 정의 – 맞춤화, 의사결정, 예측력, 상호작용, 패턴인식 등

YES ☐ | NO ☐ | N/A ☐

3) AI 활용 목적이 해당 기관의 AI 윤리원칙*에 부합하는가?

* 인권보장, 프라이버시 보호, 다양성 존중, 공공성, 책임성, 안전성, 투명성 등

• 도입할 이상거래 탐지 AI 시스템이 내부 AI 윤리원칙 등에 부합하는지 확인한다.
YES ☐ | NO ☐ | N/A ☐

4) AI 시스템의 기획 및 설계 단계에서 고객에게 미치는 영향, 위험수준, 잠재적 피해 가능성 등이 고려되었는가?

• 이상거래 탐지를 위한 AI 시스템이 내부 AI 윤리원칙에 근거한 각 금융회사의 절차에 따라 AI 모형의 위험수준 평가기준이 마련되어 있는지 확인한다.

• 이상거래 탐지 서비스에 오류가 발생하거나, 실제 이상거래가 아님에도 AI 시스템이 이상거래로 판단하는 경우, 고객의 금융생활에 미칠 영향을 확인하여 잠재적 피해 가능성을 고려한다.
YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

• A은행은 이상거래 탐지 업무에 AI를 도입하고 있으며, AI를 활용하는 업무, 역할, 서비스 유형, 속성을 정의하고 활용 목적이 명확한지 검토 후 도입했다.

• B카드는 기존 이상거래 탐지 모형의 연도별 효율성을 분석하여 모형의 적정성 여부를 평가하여 탐지 모형에 AI 활용 시 기대효과를 비교, 검토하여 AI 도입여부를 평가하고 도입 시 이상거래 탐지 범위를 정의하고 서비스 유형 및 역할을 명확히 했다.

• C카드는 전문화 · 첨단화되어 가고 있는 신용카드 범죄에 대응하기 위해 빅데이터 분석이 가능한 AI 머신러닝 기법을 적용한 사고 예측 모형 도입이 타당한지 검토하고, 도입 전·후 업무 절차, 담당자 역할 변화 등을 분석하여 적정하다는 내부 승인절차에 따라 AI 도입을 결정하여 운영하고 있다.

• D은행은 자율적으로 수립한 내부 AI 윤리원칙인 인간존중*, 피해방지**, 공정성 등에 부합하는지를 내부 승인절차를 통해 검토 후 승인되는 경우에 한하여 AI 기술을 적용하고 있다.

* 인간에게 동등하게 부여된 권리를 존중하고, 다양한 민주적 가치와 권리를 보장

** AI로 인해 발생하는 피해를 방지

• E은행은 이상거래 탐지서비스의 오탐률이 지나치게 높으면, 정상 고객들이 거래가 제한되거나, 추가인증을 하게 되는 등의 불편함이 생기게 되므로, AI 모형이 도입 전에 충분한 모니터링을 통해, 정상 고객이 불편함을 겪지 않게 되는 수준의 모형의 임계치를 검토한다.

체크리스트 가-2-1-B

AI 시스템이 인간의 의사결정을 전면적으로 대체 또는 중요 의사결정을 대체하는 경우, 감독·통제 절차가 마련되어 있는가?

체크리스트

※ AI시스템이 인간의 의사결정을 대체하는 경우가 없다면, 동 체크리스트는 "N/A" 처리할 수 있다.

1) 의사결정을 AI가 대체하는 경우 발생할 수 있는 잠재적 위험 가능성을 확인하고, 이에 대한 감독·통제 절차가 수립되어 있는가?

- AI를 활용하는 경우에도 기존 이상거래 탐지 업무에서 일반적으로 준수하고 있는 내부통제 절차가 수행되고 있는지 확인한다.
- AI 활용에 관한 내부 업무규정에 마련된 위험평가기준과 금융소비자 권리에 중대한 위험이 발생할 가능성을 기준으로 위험평가 결과의 적정성을 확인한다.
- 내부 규정 또는 업무 매뉴얼을 통해 운영 중인 AI 시스템의 위험평가 결과 보고 절차의 마련 여부와 결과 보고 및 승인 여부를 확인한다.
- AI가 기존 이상거래 탐지를 대체하는 경우 발생할 수 있는 잠재적 위험에 대해 사전·사후 대책이 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 업무별 담당자 지정 여부를 확인한다.
- AI 시스템 설계 과정에서 기능 오작동, 개인정보 이슈, 학습 및 의사결정 과정에서의 예상하지 못한 편향 등 발생 가능한 문제를 사전에 점검하고 이를 예방할 수 있는 절차가 마련되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 의사결정을 AI가 대체하는 경우, 인간의 개입이 필요한 경우에 대한 관련 절차가 정의되어 있는가?

- AI 기반의 이상거래 탐지 의사결정에 인간의 개입이 필요한 경우에 대한 업무 매뉴얼 또는 관련 내부 업무규정이 마련되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 의사결정을 AI가 대체하는 경우, 관리자, 사용자 및 기타 이해관계자가 해당 의사결정 과정에 대해 해석 및 추적이 가능하도록 설계되어 있는가?

- 이상거래 탐지 업무를 수행하는 AI 시스템의 결과에 대한 설명방안 설계 여부 등 자동 의사결정에 대해 추적 및 설명방안 마련이 이루어졌는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A카드는 AI 활용에 관한 규정을 마련하고 담당자를 지정하여 책임소재를 명확하게 하고 있으며, 발생할 수 있는 피해 방지를 위한 점검 프로세스를 통해 예상 가능한 피해를 최소화하고 있다.

- B은행은 EU 집행위의 위험단계 구분을 활용한 위험평가기준을 마련하여 위험평가를 수행하고 있으며, 결과 보고 후 승인된 경우에 한해 AI 시스템을 운영하고 있다.
  예시 위험평가 결과 : 용인불가 위험, 고위험, 제한된 위험, 최소한의 위험

- C은행은 기획 및 설계 시 AI 활용 목적, AI 서비스 특징 등을 종합적으로 고려하여 민감 정보 또는 이와 유사한 사생활 관련 정보 활용 필요성을 검토하여 정보유출, 편향적 탐지 등 잠재적 위험 가능성을 확인하고 보완대책을 검토 적용한다.

체크리스트 나-1-1-B

AI 시스템에 이용되는 학습데이터의 출처, 품질 등을 검증하고 개선필요시 조치를 취하였는가?

체크리스트

1) 학습데이터의 출처와 안정적인 데이터 수집 여부를 점검하였는가?

- 이상거래 탐지 모형 개발시, 거래시점, 거래장소 등 학습데이터 생성에 사용되는 원천데이터의 출처의 신뢰성과 데이터 수집의 안정성을 확인한다.
- 데이터 공급사와의 공급 방식에 대한 기록 및 공급내역 이력 보관 (메타기록, 내역 변경 등 변경 이력 보관)을 확인한다.
- 이상거래 탐지 모형 개발을 위한 데이터 획득 과정에서 관련 법/제도 준수 여부를 확인한다.
- AI 시스템 개발 시 고객이 일부데이터에 대해 사용 동의를 철회했을 경우 이를 반영할 수 있는 로직을 검토한다.
- 데이터 품질 관리를 위한 거버넌스 조직 또는 담당자가 금융기관 내에 있고, 양질의 데이터가 왜곡 없이 제공되도록 수시로 체크한다.

YES ☐ | NO ☐ | N/A ☐

2) 학습데이터의 품질 확보를 위해 데이터의 대표성 · 정합성을 체크하였는가?

- 학습데이터가 AI 시스템이 적용될 대상(모집단)을 대표하는지 확인한다.
  (개발 모집단 선정기준 및 개발대상 샘플 추출기준 점검)

- 전담 인력의 전문성 확보 기준 및 업무분장표 상 관련 업무연관 조직 구성 여부를 확인한다.
- 데이터 품질관리에 대한 내규 존재 여부를 확인한다.
- 학습데이터 생성에 대한 라벨링 작업 시 원천데이터와 라벨링 데이터의 동기화 여부를 확인한다.
- 사용 데이터의 정합성 검증 여부를 확인한다.
- 학습 데이터에 대한 조직의 품질 관리역량 강화를 위한 교육 및 지원 체계의 확보 여부를 확인한다.
- 사고 데이터의 유형과 출처에 따라 학습데이터 정제여부와 편향성 제거 조치여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 재학습을 수행할 경우, 학습데이터를 갱신하여 데이터의 최신성을 확보하였는가?

- 내부 규정 또는 업무매뉴얼 등을 통해 재학습 및 재배포 절차 마련 여부를 확인한다.
- 재학습 및 재배포 절차가 데이터 최신성 및 적정성 유지에 적합한지 확인한다.

YES ☐ | NO ☐ | N/A ☐

4) 학습데이터의 출처, 사전처리, 가공 등의 주요 과정을 문서화하였는가?

- 데이터의 수집과 처리 업무를 위한 절차로서 수집처리 방법 및 기준에 대한 내부 규정 또는 업무매뉴얼 반영 여부를 확인한다.
- 외부 데이터를 활용하는 경우 데이터 출처에 대해 명확한 기록 여부를 확인한다.
- 개별 데이터 처리에 대한 처리 로직기록 등 통한 데이터처리(변환/합성 등)에 대한 기록 보관 유지 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 데이터 수집과 처리 업무 절차를 규정 등에 반영하고 있으며, 마련되어 있는 절차에 따라 데이터를 목적에 부합하게 활용하고 있다.

- B카드는 모형의 개발목적에 적합하게 학습데이터의 대상과 기간을 설정한다. 또한, 학습데이터의 품질을 올리기 위해서 사고 데이터의 유형과 출처에 따라 학습데이터를 정밀하게 정제한다. 또한 데이터 수집과정에서 편향이 발생할 수 있기 때문에, 이를 예방하기 위해 다양한 샘플링 기법을 활용한다.

- C보험사는 AI 모형 학습 및 테스트 데이터 최신성 및 적정성 유지를 위해 내부 규정에 재학습 및 재배포 절차를 마련하여 주기적으로 재학습 및 재배포를 실시하여 데이터 품질을 유지하고 있다.

- D카드는 모형 개발 시, 학습데이터의 대상, 기간, 출처 등의 상세 정보를 이해관계자들이 이해하기 쉽도록 문서로 기록하여 관리하고 있다.

체크리스트 나-1-2-B

AI 시스템에 이용되는 학습데이터 또는 모형의 편향여부를 개발단계에서 테스트하고, 이를 완화하기 위한 적절한 조치를 취했는가?

체크리스트

1) AI 판단기준에 차별적인 요소가 들어가지 않도록 사전에 점검하였는가?

- 불합리한 차별을 방지하기 위하여 성별, 연령, 지역, 종교, 인종, 사회적 지위, 자산 등 불합리한 차별이 발생 가능한 평가요소를 선정하고 평가요소에 반영되었는지 점검한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 모형의 편향성 판단지표를 선정하여 개발 단계별 편향수준을 테스트하고, 편향을 완화하기 위한 적절한 조치를 취하였는가?

- 내부 정책에 따른 편향성 판단지표를 선정하였는지 확인한다.
- 개발표본 선정, 라벨링, 모델링 단계 등 세부 단계별로 편향이 존재하는지 점검한다.
- 업무 매뉴얼, 요구사항 문서 등을 통해 데이터 편향 완화 방안 마련 및 적용 여부를 확인한다.

※ 차별적 요소 포함 여부 및 편향 수준을 테스트하고, 편향을 완화하기 위한 조치를 검토하되, 편향성 판단기준을 엄격히 적용하면 AI를 통한 모형 개발이 어려운 점을 고려하여 금융회사가 자율적으로 편향성 기준 완화 여부를 결정할 수 있다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A카드는 데이터에 대한 명확한 수집 및 검수 기준을 수립하고, 다양하고 충분한 수의 데이터 수집을 확보하고, 작업자별 데이터 특성이 편향되지 않도록 편향 완화 방안을 마련함으로써 데이터 편향과 차별을 최소화하여 공정하게 적용하고 있다.

체크리스트 나-2-1-B

AI 시스템에 개인정보·민감정보를 활용하는 경우, 해당 정보의 필요성을 평가하고 안전조치를 수행하였는가?

체크리스트

1) AI 시스템에서 개인정보·민감정보의 활용 필요성을 점검하였는가?

- 이상거래 탐지 AI 시스템의 활용 목적, 적용대상을 고려하여 개인정보·민감정보 활용의 필요성을 점검하고, 반드시 필요한 경우에만 활용한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템에서 활용하는 개인정보·민감정보에 대해 안전조치를 취하였는가?

- AI 시스템의 개발을 위해 개인정보·민감정보를 수집·활용하는 경우 개인정보보호법 등 관련 법령을 준수하고, 활용 범위와 목적에 대한 명확한 기준을 마련했는지 확인한다.
- AI 시스템 개발 시 관련 정보의 유출, 악용 가능성이 없도록, 기술적·물리적 통제 방안을 마련해야 한다.

예시 개인정보 대체값 전환, 비식별화 조치 등

- 개인정보·민감정보 파기에 관해 관련 사항이 준수되고 있는지 확인한다.

예시 개인정보보호법과 개인정보처리방침에 따라 파기원칙, 파기절차, 파기방법에 따라 파기하였는지 확인

- 개인정보·민감정보의 수집·학습이 예상되는 경우, 초기 개발 단계에서 개인정보 수집과 생성에 대한 시스템을 점검할 수 있는 내부 절차를 구축했는지 확인한다.

예시 ① 활용동의 수집 여부, ② 수집방법, ③ 관련 정보보관 및 파기기간, ④ 관련 정보에 대한 활용 및 접근권한 범위 등 점검

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 FDS서비스 개발시 旣 동의하여 수집한 개인정보를 활용할 수 있는지, 민감정보를 보유한 경우 대체할 수 있는 정보는 없는지에 대해 사전 검토하고, 필요한 경우 대체값 전환 혹은 비식별화 조치하여 활용하고 있다.

- B카드사는 사고예방 목적으로 특정정보를 활용하기 위해 FDS 시스템 개발 시 준법감시 및 정보보호팀의 내부통제를 통해 ① 활용동의 수집 여부, ② 수집방법, ③ 관련 정보보관 및 파기기간, ④ 관련 정보에 대한 활용 및 접근권한 범위 등을 점검하고 있다.

체크리스트 나-3-1-B

AI 시스템의 설명가능성을 고려하고, 설명가능한 AI 기술 도입방안을 검토 또는 대안을 마련하였는가?

체크리스트

1) AI 시스템 개발 과정에서 설명가능성을 확보하기 위해 노력하였는가?
- AI 서비스 성격을 감안하여 설명 대상 및 설명하는 절차를 검토한다.
- 결과의 오해석 방지를 위해 설명 공유 대상과 범위를 설정한다.
- 설명가능한 AI 기술을 검토하고 설명방안을 마련하였는지를 확인한다.

YES □ | NO □ | N/A □

2) AI 시스템 개발 과정에서 설명가능한 AI 기술 적용이 어려운 경우, 대안을 마련하였는가?
- 설명가능한 AI 기술 개발 트렌드를 충분히 확인하여 적용이 어려운 타당한 이유와 이에 대한 대안이 충분히 마련되었는지 확인한다.
- 업무 매뉴얼, 요구사항 문서 등에 AI 시스템이 생성한 결과를 설명하기 위한 기법 검토 여부를 확인한다.

YES □ | NO □ | N/A □

3) AI 시스템 개발 과정에서 AI 시스템에 의한 결과를 고객에게 설명하기 위한 절차를 고려하였는가?
- 이상거래 탐지 서비스 관련 법령상 고객에 대한 설명의무가 없음

YES □ | NO □ | N/A □

준수사례 (예시)

- B카드사는 예측력이 우수한 AI모형(AI스코어)을 개발하고, 개발된 모형에 대해 고위험 대상으로 예측한 사유코드 등 별도의 방법론으로 개발하여 설명가능성을 확보하고 있으며, 또한 모니터링 화면 등에 고위험 스코어가 된 사유코드를 3가지를 실시간으로 볼 수 있도록 하여, 고객 문의시 응대할 수 있도록 하고 있다.

체크리스트 다-1-1-B

AI를 활용한 시스템별 성능 지표 선정 및 목표 수준을 설정하고 충족여부를 확인하였는가?

체크리스트

1) AI 기반 모형의 성능을 평가하기 위한 지표를 선정하고 있는가?

- AI 모형 성능에 대한 유지 목표 및 평가지표 선정 여부를 확인한다.
- 선정된 평가지표가 성능 유지 목표에 적합한지 검토한 결과를 확인한다.

예시 탐지율, 오탐율, PSI, SEI, KS, AUROC, AR 등

YES ☐ | NO ☐ | N/A ☐

2) 선정한 성능 평가 지표에 따라 목표 수준의 달성 여부를 점검하고 미달하였을 경우 조치하고 있는가?

- 성능 평가 결과가 목표 수준에 미달하였을 경우 원인을 분석하여 모형 재개발 여부를 검토하여 조치한다.
- 특정 학습데이터에 과적합이 되지 않았는지 다양한 테스트를 수행하고, 테스트 데이터에 따른 변동성을 모니터링하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 AI성능 목표 및 평가지표 선정하고, 학습하여 과적합 방지를 위해 학습하여 도출된 점수와 실제 사고와 매치 여부를 검토 후 가장 좋은 성능의 모형을 적용하여 FDS시스템의 탐지 성능을 안정적으로 유지하고 있다.
- B카드는 AI성능 목표 및 평가지표 선정하고, 재학습을 통한 복수의 모형을 개발하여 각 시뮬레이션 결과를 바탕으로 최적의 성능을 가진 모형을 적용하여 안정적으로 유지하고 있다.

체크리스트 다-2-1-B

불합리한 차별이 나타나지 않도록 공정성 판단기준을 설정하고 충족여부를 평가하고 개선하였는가?

체크리스트

1) 공정성 목표 수준 및 공정성 판단 지표를 선정, 관리하는가?

- 학습 전 데이터에 대한 편향성 판단 기준 및 절차 마련 여부를 확인한다.
- 학습 전 데이터에 대한 샘플링 데이터 추출 및 분포 검증 여부를 확인한다.

- AI 시스템이 학습을 통해 생성한 모형의 편향성 판단을 위한 합리적인 기준 또는 절차 마련 여부를 확인한다.
- 학습 후 AI 시스템이 생성한 모형이 분석한 결과에 대한 샘플링 데이터 추출 및 분포 검증 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 공정성 판단지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 기술적, 관리적 방안을 검토하여 조치하는가?

- 편향방지 방법론을 활용하는 등 편향을 완화하기 위한 방안을 검토하였거나 조치하였는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

※ FDS서비스는 특정 성별, 연령 등을 타깃으로 한 모형의 성과가 탁월(보이스피싱 등)하고, 특정 계층을 차별화하여야 해당 피해자를 더 구제할 수 있으므로, 편향 방지를 위한 데이터 추출과 합리적인 기준 또는 절차 마련은 필요하나, 편향성에 대하여 금융회사가 자율적으로 적용여부를 결정할 수 있다. (2)관련하여 N/A 적용 가능)

준수사례 (예시)

- A은행은 FDS서비스는 특정 성별, 연령 등을 타깃으로 한 모형의 성과가 탁월(보이스피싱 등)하고, 특정 계층을 차별화하여야 해당 피해자를 더 많이 구제할 수 있으므로, 편향성에 대하여 상황에 따라 자율적으로 적용여부를 결정하고 있다.
- B카드사는 FDS서비스의 경우 사고유형이나 승인방식 등에 따라 다양한 사고패턴을 보유하여 단일모형보다는 사고유형 및 승인방식(대면, 비대면 등)별 모형을 개발하여 예방고객이 편향되지 않게 노력하고 있으며, 산출된 모형 결과값 및 다양한 변수를 추가적으로 활용하는 규칙 기반(Rule Base)전략도 사용하여 모형에서 미처 탐지하지 못한 대상도 사고에서 보호 받을 수 있도록 하고 있다.

체크리스트 다-3-1-B

AI 시스템이 학습한 모형이 상황에 맞게 설명 가능한지 확인하고, 설명가능성을 적법한(또는 합리적인) 수준으로 개선하고자 노력하였는가?

체크리스트

1) 고객에 대한 설명의무, 금융서비스의 위험 수준 등을 고려하여 고객 및 이해관계자 등 설명 대상자들에 대해 AI 시스템 설명가능성 수준을 평가할 수 있는 기준 및 절차가 수립되어 있는가?

• 내부 규정, 업무매뉴얼 등에 설명가능성 수준에 대한 평가 기준 및 절차가 수립되어 있는지 확인한다.

예시 특정 고객 채널, 지역, 시간 등에 탐지 결과가 집중되는 이유에 대한 설명가능 수준에 대한 기준 마련 여부

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템의 설명가능성이 기준이 적법한(또는 합리적인) 수준에 도달하지 못할 경우 개선 방안(재학습이나 알고리즘 수정 등)을 검토하여 조치하고 있는가?

• 평가 기준에 도달하지 못했을 경우 수행했던 개선 방안 검토 및 조치 결과를 확인한다.

예시 과거 이상거래 탐지 기록과의 유사점과 차이점 비교를 추가적으로 수행하여 AI 시스템이 생성한 탐지 결과 설명 개선

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

• A카드는 다수의 딥러닝 알고리즘을 적용하여 이상거래 탐지에 있어 최적의 성능을 보인 방식을 적용하고 있으며 탐지 성능이 떨어지는 경우 설명가능성을 참조하여 성능 저하 원인을 파악하고 있다.

체크리스트 라-1-1-B

AI시스템 제공에 있어 AI 이용여부, 설명·이의제기권 등 고객의 권리 및 이의신청·민원제기 방법 등 소비자로서의 권리구제 방안을 마련하여 고지하였는가?

체크리스트

1) AI 시스템 적용 사실과 범위에 대해 고객에게 안내가 이뤄지고 있는가?

• 홈페이지, 모바일 앱 등의 채널을 통해 AI 시스템 활용 여부와 목적을 쉬운 용어를 사용하여 안내하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템으로 인한 고객의 불편 및 불이익에 대한 구제 방안이 마련되어 있는가?

• 내부 규정 또는 업무 매뉴얼에 신고 및 이의제기 절차가 구체적으로 마련되어 있는지 확인한다.

• 업무분장표 등을 통해 전담 조직 및 인력을 확인한다.

• 홈페이지, 모바일 앱 등의 채널을 통해 고객에게 적절한 권리구제 절차 및 기준 등 사전고지 여부를 확인한다.

※ FDS서비스는 Score로만 고객의 피해가 가는 거래정지*를 하지 않는 경우가 많아 고객의 불편 및 불이익 발생시키는 사례가 거의 없으며, FDS서비스 조치시 고객에게 FDS 서비스로 인한 조치사항을 고지하고 있다.
* Score와 거래패턴을 혼합하여 이상거래가 확실한 경우 거래정지 적용

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A사는 FDS서비스는 고객의 금융사기 피해방지를 위한 서비스로, 고위험 서비스가 아니지만, FDS를 위해 AI 활용하는 것을 홈페이지와 모바일앱을 통해 고객에게 사전 고지하고, 자율적으로 이용자 보호를 위한 절차 및 기준을 정하여 적용하고 있다.

체크리스트 라-2-1-B

AI 시스템의 성능을 주기적으로 모니터링하며 성능 개선이 필요한지의 여부를 확인하고 있는가?

체크리스트

1) AI 시스템 성능이 안정적으로 유지되는지 확인할 수 있는 모니터링 절차를 마련하였는가?

- 이상거래 탐지 AI의 성능 모니터링 주기, 범위 및 보고 등의 절차가 마련되어 있는지 확인한다.

예시 정기적 성능 점검 수행, 데이터 재학습 및 시스템 갱신 절차 등

- 성능 목표값에 대한 설정을 사전에 정의하고, 목표값 이하로 하락하는 경우, 파악할 수 있는 체계를 마련한다.

예시 보조지표(PSI, SEI, KS, AUROC, AR)를 활용하여 성능 목표값 정의

- 모니터링 결과 및 조치 이력의 문서화 여부 및 기록 내용을 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 유의미한 성능 하락 및 모집단 특성 변화에 따라 모형을 변경 또는 개선 필요 시 의사결정 단계나 절차를 사전에 정의하고 있는가?

- 주기적인 성능평가와 재학습 필요성을 검토하고, 필요 시 모형을 교체하거나 개선을 진행해야 하며 모형 변경 또는 개선을 위한 의사결정 단계와 절차가 마련되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) AI 재학습 및 모형 개선 절차를 수립·이행하고 있는가?

- AI 재학습 및 모형 개선 절차 수립과 이행 여부를 확인한다.

예시 이상거래 탐지 AI 개선을 위해 재학습 데이터의 생성, 재학습 절차, 재학습 결과에 따른 모델 사용 기준 등 포함

- 개선 방안 수립 시 데이터의 최신성, 정확성, 정합성, 비편향성을 고려하였는지 확인한다.
- 학습데이터 변경 이력 기록 여부 및 관리 문서를 확인한다.
- 활용된 데이터의 보관기간 준수 여부(5년, 10년 등)를 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A은행은 AI시스템에 대해 정기적 성능 점검을 수행하고 있으며 연 1회 이상 성능 모니터링을 통해 데이터 재학습 및 시스템 갱신 필요성을 검토하고 있다.

- B카드사는 AI시스템에 대해 기간을 명시한 정기적 성능 점검을 하고 측정된 지표의 추이를 확인하여 모형 성능을 관리하고 있으며, 성능이 기준치에 미달할 경우 학습 데이터 변경 및 재학습을 통해 이를 개선하고 있다.

- C카드사는 재학습 데이터의 생성, 재학습 절차, 재학습 결과에 따른 모델 사용 기준 등을 포함한 재학습 및 모형 개선 절차를 수립하여 이행하고 있으며 기준에 따라 재학습 결과를 검토하여 AI 시스템에 적용하고 있다.

- D카드사는 아래 보조지표를 추가로 활용하여 성능 모니터링하고 관리한다.

지표	내 용
PSI	스코어 구간별 alert건 구성비 분포변화를 측정하는 모형의 안정성지표
SEI	스코어구간별 불량(사고)건 구성비 분포 변화를 측정하는 모형의 안정성지표
KS	정상/불량 건 누적분포 차이의 최대값을 측정하는 모형 변별력 지표
AUROC	모형 전체적인 변별력 판단지표
AR	평가대상을 상대적으로 서열화 한경우 불량건의 상위 스코어 집중도 산출지표

체크리스트 라-3-1-B

AI 시스템 및 AI 시스템에서 사용하는 데이터의 오용·악용 가능성을 최소화하였는가?

체크리스트

1) 보안 시스템 구축 등을 통해 AI 이상거래 탐지 시스템에 대한 보안대책을 수립하여 적대적 공격 등 오용·악용 가능성을 최소화하고 있는가?

- AI 이상거래 탐지 시스템 구축 시 시스템, 인프라, 데이터, 네크워크, 이용자 보호 등과 관련하여 다양한 보안 위협 및 대응조치를 포함한 적합한 보안대책을 수립하고 검토하고 있는지 확인한다.

예시 보안대책 : 보안패치 의무화, 이중화 시스템 구축, 침입탐지 시스템 도입 등

- 이상거래 탐지 오동작 등 비정상 동작이나 예기치 못한 오류에 대한 대책을 수립하고, AI 이상거래 탐지 시스템에 대한 접근 권한을 관리하고 있는지 확인한다.

예시 비정상 동작, 예기치 못한 오류 처리를 위한 예외 처리 정책 마련 등

- AI 이상거래 탐지 분류 성능을 주기적으로 점검하여 성능 저하를 유발하는 공격 여부를 확인한다.
- 개인정보 암호화 이행 및 권한별 접근 통제 조치 여부를 확인한다.
- AI 이상거래 탐지 시스템을 운용하는 중요 서버에 백신 프로그램을 설치하고 주기적 업데이트 및 악성코드 점검, 실시간 검사 설정을 하고 있는지 확인한다.
- 보안 취약점에 대한 사전·정기 점검 수행 및 취약점 조치 여부를 확인한다.
- AI 이상거래 탐지 시스템 관련 침해사고 분석 시 필요한 로그에 대하여 보존 및 검토에 대한 정책을 수립하였는지 확인한다.
- AI 이상거래 탐지 시스템 관련 로그별 보존기간 및 검토 주기를 지정하였는지 확인한다.
- AI 이상거래 탐지 시스템 테스트 시 개인신용정보가 아닌 임의의 데이터를 생성하여 테스트를 수행하는지 확인한다. 단, 개인신용정보 이용이 불가피한 경우, 책임자의 승인 절차에 따라 가공하여 사용하고 즉시 폐기하는 등 관리대책이 수립·이행되었는지 확인한다.
- 암호화 대상, 사용 암호 알고리즘, 암호키 관리 방안을 포함한 개인정보 암호화 정책을 수립하였는지 확인한다.
- 정보자산 등 운영체제, 소프트웨어 패치 관리정책 및 절차를 수립·이행하고 인터넷 직접 접속을 통한 패치를 제한하고 있는지 확인한다.
- AI 이상거래 탐지 시스템 개발 과정에서 적대적 공격 등 오용·악용 가능성을 최소화하기 위해 보안성 검증을 실시하고 있는지 확인한다.

YES □ | NO □ | N/A □

2) 보안 시스템 구축 등을 통해 고객 또는 제3자에 의한 데이터 오염 공격 등을 통한 데이터의 오용·악용 가능성을 최소화하고 있는가?

- AI 이상거래 탐지 시스템에 사용되는 데이터 오용 · 악용을 감지 지표 설정 및 주기적 확인 여부 등 절차와 대처방안 수립 · 운영을 통해 관리하고 있는지 확인한다.
- 잘못된 개인신용정보 주입·조작 등 학습데이터 변조 여부를 주기적으로 확인한다.
- AI 이상거래 탐지 시스템에 사용되는 데이터의 무결성, 기밀성을 유지하기 위한 관리 규정을 준수 여부를 확인한다.

YES □ | NO □ | N/A □

3) 오픈소스(프레임워크, 라이브러리 등) 보안 취약성 관리를 위한 체계를 수립하여 AI 시스템의 보안성을 강화하고 있는가?

- AI 이상거래 탐지 시스템 개발에 필요한 오픈소스 사용 시 라이브러리 보안 취약점 확인 및 통지체계 수립 여부를 확인한다.

예시 보안취약성 상시통지 시스템 마련여부 등 확인

- 오픈소스 취약점을 주기적으로 확인하여 업데이트 및 패치를 적용하고 있는지 확인한다.
- 이상거래 탐지 관련 AI 알고리즘의 취약점 점검 여부를 확인한다.

YES □ | NO □ | N/A □

4) 침해사고 및 재해 등을 예방하기 위한 체계 및 침해사고 또는 재해가 발생했을 때 피해 확산 · 재발 방지와 신속한 복구를 위한 체계를 갖추고 있는가?

- 해킹, 악성코드, 논리폭탄, 메일폭탄, 서비스거부 또는 고출력 전자기파 등의 방법으로 사고 발생 시에 대비한 예방·복구 체계 마련 여부를 확인한다.
- 정상적 보호 · 인증절차를 우회하여 정보통신망에 접근할 수 있도록 하는 프로그램이나 기술적 장치 등에 대한 차단 체계 마련 여부를 확인한다.
- 침해사고 발생 시 기록 및 보고, 신고 및 통지, 비상 연락체계 등 내용이 포함된 침해사고 대응 절차를 수립하였는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A자산운용은 AI 시스템 이용자에 의한 오용·악용 가능성 방지를 위해 시스템 및 학습 데이터 보호를 위한 보안 조치를 검토하여 적용하고 있다. 보안 조치로서 2개 이상의 학습용 데이터 공급사를 확보하여 가용성을 확보하고 있으며, 공급사와의 데이터 송수신 시 무결성 및 기밀성이 확보되는 암호 통신 방식(HTTPS)을 적용하고 있다.

C. 챗봇

체크리스트 가-1-1-C

AI를 활용하는 목적이 명확하게 정의되고, 윤리원칙 부합 여부와 AI 활용에 따른 영향도와 잠재적 피해 가능성을 점검하였는가?

체크리스트

1) AI 도입 시, 기존 업무에 대한 영향도 분석 등을 통해 도입 타당성 검토가 이루어졌는가?

- 기존 민원상담 업무와의 비교 및 도입 시 기대효과 분석 등을 통해 업무 영향도 분석 및 타당성 검토가 이루어졌는지 확인한다.

예시 AI 도입 전·후 업무 절차, 담당자 역할 변화, 세부 업무별 AI 적용방식 적용 여부 등

※ 기존 업무가 존재하지 않는 경우, 생략 가능하다.

YES ☐ | NO ☐ | N/A ☐

2) AI 활용 목적과 업무범위, 역할 등이 명확하게 정의되어 있는가?

- 민원상담 업무 중 AI 시스템을 활용한 챗봇 서비스의 업무 범위, 활용 목적, AI의 역할 등이 명확하게 정의되어 있는지 확인한다.

예시 단순 질의 관련 민원 대응시 AI 기반의 챗봇 서비스를 적용

YES ☐ | NO ☐ | N/A ☐

3) AI 활용 목적이 해당 기관의 AI 윤리원칙*에 부합하는가?

* 인권보장, 프라이버시 보호, 다양성 존중, 공공성, 책임성, 안전성, 투명성 등

- 도입할 챗봇 서비스가 내부 AI 윤리원칙 등에 부합하는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

4) AI 시스템의 기획 및 설계 단계에서 고객에게 미치는 영향, 위험수준, 잠재적 피해 가능성 등이 고려되었는가?

- 챗봇 서비스를 위한 AI 시스템이 내부 AI 윤리원칙에 근거한 각 금융회사의 절차에 따라 AI 모형의 위험수준 평가기준이 마련되어 있는지 확인한다.
- 챗봇 서비스 기획 · 설계 과정에서 편견 혹은 차별 등의 윤리적 문제 발생 가능성을 점검하여, 사람의 업무를 챗봇이 대체하면서 발생 가능한 부작용과 불편사항을 사전에 점검하고 예상되는 사항에 대한 대안을 마련해야 한다.
- 챗봇 서비스는 사회적 약자나 취약계층 등 디지털 역량에 따른 금융소외 발생 가능성이 존재하므로, 이를 점검하고, 이들의 접근성을 충분히 보장하도록 설계함으로써

디지털 금융소외(Digital Exclusion)에 따른 부작용을 최소화할 수 있는 방안을 마련해야 한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A카드사는 고객상담 시, 고객이 자주 묻는 질문(FAQ)에 한정하여 챗봇 서비스를 운영하고, 사전에 설정된 답변을 채팅 형태로 보여주는 방식으로 업무 효율성과 고객 편의성을 높였다.
- B은행은 은행업무와 관련된 상담대화만을 하도록 서비스를 설계하였다.(차별적 발언이 발생할 수 있는 분야에 대한 적용배제)
- C카드사는 채팅을 통한 의사소통이 익숙하지 않은 디지털 취약계층 보호를 위해 챗봇 서비스 진행 과정에서 언제든 상담사 직접 연결이 가능한 버튼을 눈에 띄게 배치하였다.

체크리스트 가-2-1-C

AI 시스템이 인간의 의사결정을 전면적으로 대체 또는 중요 의사결정을 대체하는 경우, 감독 · 통제 절차가 마련되어 있는가?

체크리스트

※ AI시스템이 인간의 의사결정을 대체하는 경우가 없다면, 동 체크리스트는 "N/A" 처리할 수 있다.

1) 의사결정을 AI가 대체하는 경우 발생할 수 있는 잠재적 위험 가능성을 확인하고, 이에 대한 감독 · 통제 절차가 수립되어 있는가?
   - AI를 활용하는 경우에도 기존 민원상담 업무에서 일반적으로 준수하고 있는 내부 통제 절차가 수행되고 있는지 확인한다.
   - AI 활용에 관한 내부 업무규정에 마련된 위험평가기준과 금융소비자 권리에 중대한 위험이 발생할 가능성을 기준으로 위험평가 결과의 적정성을 확인한다.
   - 내부 규정 또는 업무 매뉴얼을 통해 운영 중인 AI 시스템의 위험평가 결과 보고 절차의 마련 여부와 결과 보고 및 승인 여부를 확인한다.
   - AI가 기존 민원상담 업무를 대체하는 경우 발생할 수 있는 잠재적 위험에 대해 예측하여 사전 · 사후 대책이 마련되어 있는지 확인한다.

- 업무분장표 등을 통해 업무별 담당자 지정 여부를 확인한다.
- AI 시스템 설계 과정에서 기능 오작동, 개인정보 이슈, 학습 및 의사결정 과정에서의 예상하지 못한 편향 등 발생 가능한 문제를 사전에 점검하고 이를 예방할 수 있는 절차가 마련되어 있는지 확인한다.

예시 인종 · 성별 · 연령 등에 따른 차별 발언 등

YES □ | NO □ | N/A □

2) 의사결정을 AI가 대체하는 경우, 인간의 개입이 필요한 경우에 대한 관련 절차가 정의되어 있는가?

- AI 기반의 챗봇 서비스에 인간의 개입이 필요한 경우에 대한 업무 매뉴얼 또는 관련 내부절차가 마련되어 있는지 확인한다.

YES □ | NO □ | N/A □

3) 의사결정을 AI가 대체하는 경우, 관리자, 사용자 및 기타 이해관계자가 해당 의사결정 과정에 대해 해석 및 추적이 가능하도록 설계되어 있는가?

- 챗봇 서비스를 수행하는 AI 시스템의 결과에 대한 설명방안 설계 여부 등 자동 의사결정에 대해 추적 및 설명방안 마련이 이루어졌는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A보험사는 챗봇 운영에 대한 내부지침을 보유하여 운영하고 있다.
- B은행은 업무 외적인 편견 · 차별적 상담데이터를 서비스 설계단계부터 배제하여 학습을 시키지 않았고, 편견/차별적 질문을 받을 시 “죄송합니다. 말씀을 이해하지 못했습니다.” 등의 문구를 활용하여 관련 대화를 원천부터 차단하였다.

체크리스트 나-1-1-C

AI 시스템에 이용되는 학습데이터의 출처, 품질 등을 검증하고 개선필요시 조치를 취하였는가?

체크리스트

1) 학습데이터의 출처와 안정적인 데이터 수집 여부를 점검하였는가?

- 학습데이터 생성에 사용되는 원천데이터의 출처의 신뢰성과 데이터 수집의 안정성을 확인한다.

- 챗봇 대화 학습을 위한 데이터 공급 방식에 대한 기록 및 보관 (메타기록, 내역변경 등 변경 이력 보관)을 확인한다.
- 개인정보보호법을 준수하여 데이터 수집 시, 제공자에게 수집 사실을 고지하고 동의를 얻어야 한다.
- 챗봇 시스템 개발 시 활용한 고객의 데이터 중 일부에 대한 사용 동의를 철회했을 경우 이를 반영할 수 있는 로직을 검토한다.
- 데이터 품질 관리를 위한 거버넌스 조직 또는 담당자가 금융기관 내에 있고, 양질의 데이터가 왜곡 없이 제공되도록 수시로 체크한다.

예시 챗봇 대화기록 및 답변 오류 등에 대한 점검 등

YES ☐ | NO ☐ | N/A ☐

2) 학습데이터의 품질 확보를 위해 데이터의 대표성 · 정합성을 체크하였는가?

- 챗봇 학습데이터에 의한 대화 및 답변 시나리오가 정확하고, 답변 대상자에게 올바르게 적용될 수 있는지를 확인한다.
  (개발 모집단 선정기준 및 개발대상 샘플 추출기준 점검)
- 전담 인력의 전문성 확보 기준 및 업무분장표 상 관련 업무연관 조직 구성 여부를 확인한다.
- 챗봇 응답 데이터 품질관리에 대한 내규 존재 여부를 확인한다.
- 학습데이터 생성에 대한 라벨링 작업 시 원천데이터와 라벨링 데이터의 동기화 여부를 확인한다.
- 사용 데이터의 정합성 검증 여부를 확인한다.
- 학습 데이터에 대한 조직의 품질 관리역량 강화를 위한 교육 및 지원 체계의 확보 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 재학습을 수행할 경우, 학습데이터를 갱신하여 데이터의 최신성을 확보하였는가?

- 내부 규정 또는 업무매뉴얼 등을 통해 재학습 및 재배포 절차 마련 여부를 확인한다.
- 재학습 및 재배포 절차가 데이터 최신성 및 적정성 유지에 적합한지 확인한다.

예시 전담 부서를 통해 주기적으로 답변 적절성과 시의성 등을 점검

YES ☐ | NO ☐ | N/A ☐

4) 학습데이터의 출처, 사전처리, 가공 등의 주요 과정을 문서화하였는가?

- 데이터의 수집과 처리 업무를 위한 절차로서 수집처리 방법 및 기준에 대한 내부 규정 또는 업무매뉴얼 반영 여부를 확인한다.
- 외부 데이터를 활용하는 경우 데이터 출처에 대해 명확한 기록 여부를 확인한다.
- 개별 데이터 처리에 대한 처리 로직기록 등 통한 데이터처리(변환/합성 등)에 대한 기록 보관 유지 여부를 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A보험사는 각 질의어에 대한 TA 분석(HOT 키워드, 급상승 키워드, 연관어 등)을 통해 트렌드 문장(키워드)를 도출하고 학습에 활용하고 있다.
- B은행은 질의 관리 전담 조직을 통해 미답변 문의에 대한 데이터 점검 및 최신화를 진행하고 있다.
- C카드사는 챗봇을 위한 학습데이터의 출처, 사전처리, 가공 등의 과정에 대한 내부 지침을 제정하여 데이터 수집 및 가공 업무시 이를 준수하는지 확인하고 있다.
- D카드사는 고객이 챗봇에 시도하는 대화 및 질의에 대한 답변을 학습시키며, 이 과정에서 문의에 대한 정확한 답변이 이뤄지고 있는지를 사람이 점검하고 정책 및 제도 변경에 따른 답변을 업데이트 하고 있다.

체크리스트 나-1-2-C

AI 시스템에 이용되는 학습데이터 또는 모형의 편향여부를 개발단계에서 테스트하고, 이를 완화하기 위한 적절한 조치를 취했는가?

체크리스트

1) AI 판단기준에 차별적인 요소가 들어가지 않도록 사전에 점검하였는가?

- 불합리한 차별을 방지하기 위하여 성별, 연령, 지역, 종교, 인종, 사회적 지위, 자산 등 불합리한 차별이 발생 가능한 평가요소를 선정하고 평가요소에 반영되었는지 점검한다.

YES □ | NO □ | N/A □

2) AI 모형의 편향성 판단지표를 선정하여 개발 단계별 편향수준을 테스트하고, 편향을 완화하기 위한 적절한 조치를 취하였는가?

- 내부 정책에 따른 편향성 판단지표를 선정하였는지 확인한다.
- 개발표본 선정, 라벨링, 모델링 단계 등 세부 단계별로 편향이 존재하는지 점검한다.
- 업무 매뉴얼, 요구사항 문서 등을 통해 데이터 편향 완화 방안 마련 및 적용 여부를 확인한다.
- 챗봇 질의 응답 및 대화 시나리오 점검을 통해 응답 및 대화의 편향성을 제거해야 한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 매일 들어오는 운영데이터를 수시로 학습시키며 인종/성차별 등의 데이터는 학습에서 제외하고 있다. 또한 향후 외부데이터의 학습이 필요할 시 원천 소스를 기록 및 관리하여 데이터의 신뢰성을 제고할 계획이다.

체크리스트 나-2-1-C

AI 시스템에 개인정보·민감정보를 활용하는 경우, 해당 정보의 필요성을 평가하고 안전조치를 수행하였는가?

체크리스트

1) AI 시스템에서 개인정보·민감정보의 활용 필요성을 점검하였는가?

- 챗봇 서비스의 제공 목적, 적용대상을 고려하여 개인정보·민감정보 활용의 필요성을 점검하고, 반드시 필요한 경우에만 활용한다.

예시 챗봇을 통한 단순 서비스 안내 등을 진행할 경우 개인정보를 수집하지 않아야 한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템에서 활용하는 개인정보·민감정보에 대해 안전조치를 취하였는가?

- AI 시스템의 개발을 위해 개인정보·민감정보를 수집·활용하는 경우 개인정보보호법 등 관련 법령을 준수하고, 활용 범위와 목적에 대한 명확한 기준을 마련했는지 확인한다.
- AI 시스템 개발 시 관련 정보의 유출, 악용 가능성이 없도록, 기술적·물리적 통제 방안을 마련해야 한다.
- 개인정보·민감정보 파기에 관해 관련 사항이 준수되고 있는지 확인한다.

- 개인정보·민감정보의 수집·학습이 예상되는 경우, 초기 개발 단계에서 개인정보 수집과 생성에 대한 시스템을 점검할 수 있는 내부 절차를 구축했는지 확인한다.

예시 개발 단계에서 개인정보 생성 및 수집 여부 점검 절차를 마련한다

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A사는 챗봇 서비스 도입 전 기 동의하였던 개인정보를 챗봇에 활용할 수 있는지, 아니면 별도로 개인정보 동의가 필요한지 여부 등에 대하여 개인정보 관련 유관부서 협의 및 법률검토를 사전에 진행하고 있다.
- B카드사는 정보보호부를 통해 개인정보 활용에 대한 사전 검토를 수행하고 챗봇 개인정보 수집 이용동의서를 제정하였다.
- C카드사는 챗봇 대화 데이터 적재 단계부터 개인정보를 제외 처리한 후 학습 시 활용하고 있다.
- D보험사는 정보보호 운영 정책에 따라 각 영역별 (개인정보 유입 가능성, 개인정보 보호 대책, 비식별화 여부, 수집 여부, 수집 기간, 고객 동의 절차 등) 기준을 정립하였다.

체크리스트 나-3-1-C

AI 시스템의 설명가능성을 고려하고, 설명가능한 AI 기술 도입방안을 검토 또는 대안을 마련하였는가?

체크리스트

1) AI 시스템 개발 과정에서 설명가능성을 확보하기 위해 노력하였는가?
   - AI 서비스 성격을 감안하여 설명 대상 및 설명하는 절차를 검토한다.
   - 결과의 오해석 방지를 위해 설명 공유 대상과 범위를 설정한다.
   - 설명가능한 AI 기술을 검토하고 설명방안을 마련하였는지를 확인한다.

   YES ☐ | NO ☐ | N/A ☐

2) AI 시스템 개발 과정에서 설명가능한 AI 기술 적용이 어려운 경우, 대안을 마련하였는가?
   - 설명가능한 AI 기술 개발 트렌드를 충분히 확인하여 적용이 어려운 타당한 이유와 이에 대한 대안이 충분히 마련되었는지 확인한다.

• 업무 매뉴얼, 요구사항 문서 등에 AI 시스템이 생성한 결과를 설명하기 위한 기법 검토 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) AI 시스템 개발 과정에서 AI 시스템에 의한 결과를 고객에게 설명하기 위한 절차를 고려하였는가?

• 챗봇 서비스 관련 법령상 고객에 대한 설명의무가 없음

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

• A카드사는 챗봇 이용 중에 고객이 원할 경우 현재 챗봇이 어떤 종류의 AI 알고리즘과 고객 데이터에 기반하고 있는지에 대한 정보 페이지를 제공하는 버튼을 삽입하였다.

체크리스트 다-1-1-C

AI를 활용한 시스템별 성능 지표 선정 및 목표 수준을 설정하고 충족여부를 확인하였는가?

체크리스트

1) AI 기반 모형의 성능을 평가하기 위한 지표를 선정하고 있는가?

• AI 모형 성능에 대한 유지 목표 및 평가지표 선정 여부를 확인한다.

• 정밀도, 재현율 등 챗봇 출력 결과의 임계치 도출을 위해 모델 구현 과정에서 발생 가능한 문제점을 파악하고, 문제 발생 여부를 결정 짓는 중요 변수를 파악하고 있는지 확인한다.

• 임계치 기준 충족 여부에 대한 검증은 SVM, CNN, RNN, LSTM 등 다양한 인공지능 모델을 통해 가능하다.

YES ☐ | NO ☐ | N/A ☐

2) 선정한 성능 평가 지표에 따라 목표 수준의 달성 여부를 점검하고 미달하였을 경우 조치하고 있는가?

• AI 시스템 성능에 대한 검증을 통해 모형의 성능이 임계치 이하로 떨어지는 성능 저하 발생 기준을 마련하였는지 확인한다.

• 성능 저하 발생 기준에 따른 지속적인 모니터링이 수행되고 있는지 확인한다.

- 성능이 기준치 이하로 떨어지는 경우 사용자에 대한 안내 및 경고, 시스템 보완 프로세스 작동 등 보완 대책 마련을 통해 신뢰도 높은 서비스 제공 등이 이루어지고 있는지 확인한다.
- 특정 학습데이터에 과적합이 되지 않았는지 다양한 테스트를 수행하고, 테스트 데이터에 따른 변동성을 모니터링하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 챗봇 응답률 관리 및 임계치를 관리하며, 성능 저하 시 데이터 재학습 및 모델을 개선한다.
- B카드사는 챗봇 주요 NLP 알고리즘의 성능 저하 현상을 비정상답변 최소화를 위한 답변 검증 방식으로 모니터링하고 있다.
- C카드사는 정상답변율 지표는 답변 검증일 단위로 모니터링하며, 지표 저하 현상 확인 시 학습을 통한 알고리즘 최적화 작업 진행한다.
- D보험사는 추가 지식 학습을 통해 모델 신규 개발 시점마다 Accuracy 평가, F-Score 평가, K-fold 테스트 등을 통해 NLP 모델의 성능을 평가하고 기준을 넘을 경우에만 운영 이행을 하고 있다.

체크리스트 다-2-1-C

불합리한 차별이 나타나지 않도록 공정성 판단기준을 설정하고 충족여부를 평가하고 개선하였는가?

체크리스트

1) 공정성 목표 수준 및 공정성 판단 지표를 선정, 관리하는가?

- 인공지능 의사결정 추적을 위한 로그 수집을 구현하고 있는지 확인한다.
- 챗봇 프로그램 운영 과정에서 데이터 수집 • 학습이 반복적으로 이뤄질 경우, 추가되는 데이터를 기반으로 한 주기적인 평가 지표에 대한 관리 및 평가가 이루어지고 있는지 확인한다.
- 시스템 결정에 대한 세부적 기준을 내부적으로 확립하고, 시스템 운용 과정에서 이를 추적할 수 있는 방안으로 로그 수집을 수행하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 공정성 판단지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 기술적, 관리적 방안을 검토하여 조치하는가?

- 편향방지 방법론을 활용하는 등 편향을 완화하기 위한 방안을 검토하였거나 조치하였는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 지식 전담 관리 조직을 두어 AI 성능 지표 및 수준을 상시적으로 관리하고 있다.
- B보험사는 추가적으로 데이터를 학습할 때마다 대화 모형에 대한 평가 및 시스템 개선 등을 통해 공정성을 확보하고 있다.
- C은행은 챗본 이용자 로그 기록을 수집, 주 단위로 검토하여 특정 연령, 성별 등에서 챗봇 이용률 및 만족도에 문제가 있는지 파악하고 문제가 있는 경우 원인을 검토하여 조치한다.

체크리스트 다-3-1-C

AI 시스템이 학습한 모형이 상황에 맞게 설명 가능한지 확인하고, 설명가능성을 적법한(또는 합리적인) 수준으로 개선하고자 노력하였는가?

체크리스트

1) 고객에 대한 설명의무, 금융서비스의 위험 수준 등을 고려하여 고객 및 이해관계자 등 설명 대상자들에 대해 AI 시스템 설명가능성 수준을 평가할 수 있는 기준 및 절차가 수립되어 있는가?

- 내부 규정, 업무매뉴얼 등에 설명가능성 수준에 대한 평가 기준 및 절차가 수립되어 있는지 확인한다.

예시 AI 시스템의 이용자 질의 의도파악 방식, 답변 생성 방식, 오답 발생 원인 등에 대해 유의미하게 설명되는지 여부

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템의 설명가능성이 기준이 적법한(또는 합리적인) 수준에 도달하지 못할 경우 개선 방안(재학습이나 알고리즘 수정 등)을 검토하여 조치하고 있는가?

- 평가 기준에 도달하지 못했을 경우 수행했던 개선 방안 검토 및 조치 결과를 확인한다.

예시 이용자 질의 의도파악, 답변 생성 등에 대해 설명이 부족하고 이용도 만족도가 낮은 경우 지식관리체계 등 외부 데이터를 통해 학습 및 답변 지원

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A금융사는 이용자 질의에 대한 의도를 정확하게 파악할 수 있도록 측정지표를 마련하여 주기적으로 측정하고 수준을 벗어날 경우 학습을 통해 보완한다. 또한 금융서비스별로 표준질의 및 유사질의를 관리하고 학습하여 일정수준을 유지할 수 있도록 관리체계를 마련하고 있다.
- B은행은 AI가 활용할 수 있는 지식관리 시스템(KMS)을 갖추고 금융서비스에 대한 신뢰성 있는 답변을 할 수 있도록 준비하고 있다.
- C카드사는 챗봇이 한 답변을 로그를 이용하여 수집하고 신뢰도를 주기적으로 측정하고 설명의 의무를 위반한 답변이 없는지 분석한다.
- D보험사는 챗봇 이용자의 불만제기를 챗봇 또는 앱(MTS)를 통해 수집하고 내부 소비자보호 민원처리 지침에 따라 대응한다.

체크리스트 라-1-1-C

AI시스템 제공에 있어 AI 이용여부, 설명·이의제기권 등 고객의 권리 및 이의신청·민원제기 방법 등 소비자로서의 권리구제 방안을 마련하여 고지하였는가?

체크리스트

1) AI 시스템 적용 사실과 범위에 대해 고객에게 안내가 이뤄지고 있는가?

- 홈페이지, 모바일 앱 등의 채널을 통해 AI 시스템 활용 여부와 목적을 쉬운 용어를 사용하여 안내하고 있는지 확인한다.
- 고객의 권리에 대한 사전고지 여부 및 변경 이력을 확인한다.
- AI 시스템을 통한 챗봇 서비스 제공 범위와 한계를 설명함으로써 고객의 불필요한 오해 여지를 없애고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템으로 인한 고객의 불편 및 불이익에 대한 구제 방안이 마련되어 있는가?

- 내부 규정 또는 업무 매뉴얼에 신고 및 이의제기 절차가 구체적으로 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 전담 조직 및 인력을 확인한다.
- 홈페이지, 모바일 앱 등의 채널을 통해 고객에게 적절한 권리구제 절차 및 기준 등 사전고지 여부를 확인한다.
- 챗봇 서비스를 불편하게 느끼는 대상에 대한 상담원 등 사람 연결을 통한 해결 노력, AI 시스템 오류로 인한 불이익에 대한 보완책 마련 여부를 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A은행은 챗봇 메뉴 진입 시 인사말에 사람을 통한 상담이 아님을 고지한 후 대화를 시작하고, 챗봇이 가지는 한계점(다양한 요인들에 의한 영향 등)도 안내하였다.
- B카드사는 챗봇 인사말에 AI 챗봇 안내임을 노출하고, 챗봇을 통한 최초 업무처리 시 챗봇 이용약관에 동의하고 업무를 진행하도록 하고 있다.
- C보험사는 챗봇 상담에 대한 불편을 보완하기 위해 채팅상담사 연결 기능과 민원(오류)신고 기능을 제공하고 있다.
- D사는 기존에 직원의 오상담으로 인한 구제 절차를 챗봇의 오상담까지 확대 적용하였다.

체크리스트 라-2-1-C

AI 시스템의 성능을 주기적으로 모니터링하며 성능 개선이 필요한지의 여부를 확인하고 있는가?

체크리스트

1) AI 시스템 성능이 안정적으로 유지되는지 확인할 수 있는 모니터링 절차를 마련하였는가?

- 챗봇의 대화 적절성 등 성능 모니터링 주기, 범위 및 보고 등의 절차가 마련되어 있는지 확인한다.

예시 매달 응답율(챗봇에 들어온 질문 중 답을 할 수 없어 대답하지 못하고 '잘 모르겠다'고 답변한 비율)을 체크

- 답변 오류율 등 성능 목표값에 대한 설정을 사전에 정의하고, 목표값 이하로 하락하는 경우, 파악할 수 있는 체계를 마련한다.
- 모니터링 결과 및 조치 이력의 문서화 여부 및 기록 내용을 확인한다.
- 챗봇 서비스에 대한 평가 및 개선을 위해 유용성, 편의성, 효율성 등 다양한 영역에 대한 사용자 경험 평가를 통해 사용자가 체감하는 만족 정도를 평가하고 개선을 위한 방향을 설정했는지 확인한다.

예시 챗봇 이용자 대상 만족도 조사를 통해, 불편 사항 및 오류 응답을 점검하고 개선한다.

YES ☐ | NO ☐ | N/A ☐

2) 유의미한 성능 하락 및 모집단 특성 변화에 따라 모형을 변경 또는 개선 필요시 의사결정 단계나 절차를 사전에 정의하고 있는가?

- 주기적인 성능평가와 재학습 필요성을 검토하고, 필요 시 모형을 교체하거나 개선을 진행해야 하며 모형 변경 또는 개선을 위한 의사결정 단계와 절차가 마련되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) AI 재학습 및 모형 개선 절차를 수립·이행하고 있는가?

- AI 재학습 및 모형 개선 절차 수립과 이행 여부를 확인한다.
- 개선 방안 수립 시 데이터의 최신성, 정확성, 정합성, 비편향성을 고려하였는지 확인한다.
- 학습데이터 변경 이력 기록 여부 및 답변 관련 관리 문서를 확인한다.
- 활용된 데이터의 보관기간 준수 여부(5년, 10년 등)를 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 모든 챗봇 답변에 대해 고객의 긍정/부정 피드백을 수렴받고, 이를 향후 답변 모형에 반영하고 있다.
- B카드사는 챗봇 이용 시 비정기적으로 사용자 대상 건의사항을 수집받고 있다.
- C은행은 관제 기능을 통해 지식관리 시스템 인프라를 상시 모니터링하고, 고객으로부터 답변 결과에 대한 긍정/부정 의견을 받고, 지식관리 전담 조직을 통해 신규 지식 반영 및 기존 지식을 업데이트하고 있다.

체크리스트 라-3-1-C

AI 시스템 및 AI 시스템에서 사용하는 데이터의 오용·악용 가능성을 최소화하였는가?

체크리스트

1) 보안 시스템 구축 등을 통해 AI 시스템에 대한 모니터링을 수행하여 적대적 공격 등 오용·악용 가능성을 최소화하고 있는가?

- AI 챗봇시스템 구축 시 시스템, 인프라, 데이터, 네크워크, 이용자 보호 등과 관련하여 다양한 보안 위협 및 대응조치를 포함한 적합한 보안대책을 수립하고 검토하고 있는지 확인한다.
- 비정상적인 답변이나 예기치 못한 대화 오류 등에 대한 대책을 수립하고, 챗봇 시스템 접근 권한을 관리하고 있는지 확인한다.
- AI 챗봇 시스템의 성능을 주기적으로 점검하여 성능 저하를 유발하는 공격 여부를 확인한다.
- 개인정보 암호화 및 권한별 접근 통제 조치 여부를 확인한다.
- AI 챗봇시스템을 운용하는 중요 서버에 백신 프로그램을 설치하고 주기적 업데이트 및 악성코드 점검, 실시간 검사 설정을 하고 있는지 확인한다.
- 보안 취약점에 대한 사전·정기 점검 수행 및 취약점 조치 여부를 확인한다.
- AI 챗봇시스템 관련 침해사고 분석 시 필요한 로그에 대하여 보존 및 검토에 대한 정책을 수립하였는지 확인한다.
- AI 챗봇시스템 관련 로그별 보존기간 및 검토 주기를 지정하였는지 확인한다.
- AI 챗봇시스템 테스트 시 개인신용정보가 아닌 임의의 데이터를 생성하여 테스트를 수행하는지 확인한다. 단, 개인신용정보 이용이 불가피한 경우, 책임자의 승인 절차에 따라 가공하여 사용하고 즉시 폐기하는 등 관리대책이 수립·이행되었는지 확인한다.
- 암호화 대상, 사용 암호 알고리즘, 암호키 관리 방안을 포함한 개인정보 암호화 정책을 수립하였는지 확인한다.
- 정보자산 등 운영체제, 소프트웨어 패치 관리정책 및 절차를 수립·이행하고 인터넷 직접 접속을 통한 패치를 제한하고 있는지 확인한다.
- AI 챗봇시스템 개발 과정에서 적대적 공격 등 오용·악용 가능성을 최소화하기 위해 보안성 검증을 실시하고 있는지 확인한다.

YES □ | NO □ | N/A □

2) 보안 시스템 구축 등을 통해 고객 또는 제3자에 의한 데이터 오염 공격 등을 통한 데이터의 오용·악용 가능성을 최소화하고 있는가?

- 비정상적 챗봇 대화를 통한 학습 데이터 오용 · 악용 여부 감지 지표 설정 및 주기적 확인 여부 등 절차와 대처방안 수립 · 운영 여부를 확인한다.

예시 회사 보안관제 시스템에 챗봇에 대한 모니터링 기능을 추가한다.

- 잘못된 개인신용정보 주입·조작 등 학습데이터 변조 여부를 주기적으로 확인한다.
- AI 챗봇시스템에 사용되는 데이터의 무결성, 기밀성을 유지하기 위한 관리 규정을 준수 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 오픈소스(프레임워크, 라이브러리 등) 보안 취약성 관리를 위한 체계를 수립하여 AI 시스템의 보안성을 강화하고 있는가?

- 사전학습 언어모델 등 오픈소스 사용 시 라이브러리 보안 취약점 확인 및 통지체계 수립 여부를 확인한다.
- 오픈소스 취약점을 주기적으로 확인하여 업데이트 및 패치를 적용하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

4) 침해사고 및 재해 등을 예방하기 위한 체계 및 침해사고 또는 재해가 발생했을 때 피해 확산 · 재발 방지와 신속한 복구를 위한 체계를 갖추고 있는가?

- 해킹, 악성코드, 논리폭탄, 메일폭탄, 서비스거부 또는 고출력 전자기파 등의 방법으로 사고 발생 시에 대비한 예방·복구 체계 마련 여부를 확인한다.
- 정상적 보호 · 인증절차를 우회하여 정보통신망에 접근할 수 있도록 하는 프로그램이나 기술적 장치 등에 대한 차단 체계 마련 여부를 확인한다.
- 침해사고 발생 시 기록 및 보고, 신고 및 통지, 비상 연락체계 등 내용이 포함된 침해사고 대응 절차를 수립하였는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A카드사는 챗봇 오픈 전 보안성심의를 통과토록 하고 있다.
- B카드사는 전사 보안관제 시스템 내 챗봇 시스템에 대한 관제/모니터링 기능을 포함하여 실시하고 있다.
- C보험사는 기획/설계 단계, 이행 단계 시점에 시스템 운영 및 데이터 보안 관리 방안을 정보보호 보안성 심의를 통과하도록 하고 있다.
- D사는 기존 시스템 대상으로 적용되던 보안관리 절차(보안성 심의, 사후 모니터링, 자체점검 등)를 챗봇 시스템에도 동일한 수준으로 적용하고 있다.

D. 맞춤형 상품 추천

체크리스트 가-1-1-D

AI를 활용하는 목적이 명확하게 정의되고, 윤리원칙 부합 여부와 AI 활용에 따른 영향도와 잠재적 피해 가능성을 점검하였는가?

체크리스트

1) AI 도입 시, 기존 업무에 대한 영향도 분석 등을 통해 도입 타당성 검토가 이루어졌는가?

- 기존에 사람이 수행하던 상품추천 업무와의 비교 및 도입 시 기대효과 분석 등을 통해 업무 영향도 분석 및 타당성 검토가 이루어졌는지 확인한다.

예시 AI 도입 전·후 업무 절차, 담당자 역할 변화, 세부 업무별 AI 적용방식 적용 여부 등

※ 기존 업무가 존재하지 않는 경우, 생략 가능하다.

YES ☐ | NO ☐ | N/A ☐

2) AI 활용 목적과 업무범위, 역할 등이 명확하게 정의되어 있는가?

- 상품추천 업무 중 AI 시스템을 활용한 맞춤형 상품추천 서비스의 업무 범위, 활용 목적, AI의 역할 등이 명확하게 정의되어 있는지 확인한다.

예시 출퇴근 30대 직장인의 소비패턴에 맞는 맞춤형 신용카드 추천 시 AI를 활용하여 대중교통 할인, 커피 할인에 특화된 신용카드 제시

YES ☐ | NO ☐ | N/A ☐

3) AI 활용 목적이 해당 기관의 AI 윤리원칙*에 부합하는가?

* 인권보장, 프라이버시 보호, 다양성 존중, 공공성, 책임성, 안전성, 투명성 등

- 도입할 맞춤형 상품추천 서비스가 내부 AI 윤리원칙 등에 부합하는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

4) AI 시스템의 기획 및 설계 단계에서 고객에게 미치는 영향, 위험수준, 잠재적 피해 가능성 등이 고려되었는가?

- 맞춤형 상품추천을 위한 AI 시스템이 내부 AI 윤리원칙에 근거한 각 금융회사의 절차에 따라 AI 모형의 위험수준 평가기준이 마련되어 있는지 확인한다.
- 일부 보험상품은 고객의 사고 발생 시 제공되는 특약의 묶음을 제한적으로 제공하는 경우가 있으므로, 고객의 생명과 안전을 침해하지 않으며 금전적 손해를 끼치지 않도록 구성되는지 확인해야 한다.

예시 자동차 보험의 상품구성에 긴급호출 서비스가 없어, 실제 위기상황 발생에도 자동차 보험 서비스를 사용할 수 없는 경우

- 추천된 상품 자체가 사회 전반의 공공의 이익에 부합하지 않거나, 금융소비자 편익에 역행하는 상품이 일시적으로 출시될 수 있으므로 이러한 경우, 상품추천 시스템 적용을 최소화하거나, 제거해야 할 수 있는 방안을 마련해야 한다.

예시 여행을 많이 다닐수록(GPS위치등록) 금리를 우대하는 예금의 경우 코로나, 메르스 같은 사회적 질병 발생 시 상품 구성이 공공 이익에 일시적으로 역행할 수 있음

YES □ | NO □ | N/A □

준수사례 (예시)

- A금융사는 상품추천시 컴플라이언스 체크리스트를 마련하여 사전점검이 가능한 시스템을 도입하였다.

- B금융사는 상품추천 알고리즘 개발 시 상품 보유현황 및 라이프스타일별 선호상품/서비스를 입력 자료로 구성하고, 이 외에 국적이나 사회적 계층 등 차별적 요소가 있는 자료는 입력자료로 사용하지 않고 있다.

- C금융사는 고객 대표자 회의를 분기마다 운영하여 고객 VOC에서 관련 의견을 파악하고 AI 서비스로 추천된 상품 및 서비스가 사회경제적으로 부정적인 요소가 있는지 사전에 파악하고 있다.

- D금융사는 상품 추천에 대해 불만족스러운 사항 발생 시 소비자가 이에 대해 개선 의견을 개진할 수 있는 메뉴를 마련하여 운영하고 있다.

- E그룹의 경우, AI 기술 기업으로서 윤리헌장을 마련하여, AI 알고리즘의 기본 원칙으로, 알고리즘과 관련된 모든 노력을 우리 사회 윤리안에서 다하며, 이를 통해 인류의 편익과 행복을 추구한다는 방향을 설정하고, 이를 AI 기술 전반에 적용하고 있다.

체크리스트 가-2-1-D

AI 시스템이 인간의 의사결정을 전면적으로 대체 또는 중요 의사결정을 대체하는 경우, 감독 · 통제 절차가 마련되어 있는가?

체크리스트

※ AI시스템이 인간의 의사결정을 대체하는 경우가 없다면, 동 체크리스트는 “N/A” 처리할 수 있다.

1) 의사결정을 AI가 대체하는 경우 발생할 수 있는 잠재적 위험 가능성을 확인하고, 이에 대한 감독·통제 절차가 수립되어 있는가?

- AI를 활용하는 경우에도 기존 상품추천 업무에서 일반적으로 준수하고 있는 내부 통제 절차가 수행되고 있는지 확인한다.

예시 고객에게 불리한 상품 추천 등 고객의 발생 가능한 위험 및 피해를 예방하기 위한 적절한 감독 및 통제 인력 프로세스 등

- AI 활용에 관한 내부 업무규정에 마련된 위험평가기준과 금융소비자 권리에 중대한 위험이 발생할 가능성을 기준으로 위험평가 결과의 적정성을 확인한다.
- 내부 규정 또는 업무 매뉴얼을 통해 운영 중인 AI 시스템의 위험평가 결과 보고 절차의 마련 여부와 결과 보고 및 승인 여부를 확인한다.
- AI가 기존 상품추천 업무를 대체하는 경우 발생할 수 있는 잠재적 위험에 대해 예측하여 사전·사후 대책이 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 업무별 담당자 지정 여부를 확인한다.
- AI 시스템 설계 과정에서 성별, 연령, 장해, 소득 등 다양한 요소에 대한 데이터 학습, 활용, 의사결정 등 AI 시스템 전반에서 편향이 발생할 가능성이 있는지 점검하고 편향을 예방할 수 있는 모니터링 등 확인 절차, 편향 발생시 이를 조치할 수 있는 통제 절차를 마련하였는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 의사결정을 AI가 대체하는 경우, 인간의 개입이 필요한 경우에 대한 관련 절차가 정의되어 있는가?

- AI 기반의 맞춤형 상품추천 의사결정에 인간의 개입이 필요한 경우에 대한 업무 매뉴얼 또는 관련 내부절차가 마련되어 있는지 확인한다.
- 금융회사는 AI 시스템에 의한 상품추천 외에 필요에 따라 신상품 또는 정부 요청 등에 의해 한시적인 상품 추천 로직 개발·운영이 있을 수 있으며, 감독기관의 별도 통제에 따라 적용되도록 운영할 수 있다.

YES ☐ | NO ☐ | N/A ☐

3) 의사결정을 AI가 대체하는 경우, 관리자, 사용자 및 기타 이해관계자가 해당 의사결정 과정에 대해 해석 및 추적이 가능하도록 설계되어 있는가?

- 맞춤형 상품추천을 수행하는 AI 시스템의 결과에 대한 설명방안 설계 여부 등 자동 의사결정에 대해 추적 및 설명방안 마련이 이루어졌는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

〈 관련 법령에 따른 별도 체크리스트 〉

4) (금융회사 이해상충 방지 기준에 따른 별도 체크리스트) 맞춤형 상품 추천 AI 시스템이 금융소비자가 본인에게 유리한 조건으로 검색·상품배열을 할 수 있도록 설계되어 있는가?

- 맞춤형 상품 추천 AI 시스템이 금융회사와 금융소비자간 이해상충 방지를 위해, 금융소비자보호법 감독규정 제6조 제7항(알고리즘 요건)이 반영되어 있는지 확인한다.
- 신상품 출시 또는 기존상품 우대 조건이 발생할 경우, 고객 우대혜택을 재계산해 최선의 상품을 추천하는 알고리즘을 구성할 수 있다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A보험사는 맞춤형 상품 추천 서비스에서 추천한 보험 상품이 고객의 마음에 들지 않을 경우 손쉽게 다른 상품을 선택해 설계할 수 있도록 서비스 메뉴를 구성하고 있다.
- B금융사는 고객중심경영최고책임자(Chief Customer Officer; CCO)를 두어 신규 서비스, 상품 도입 시 사전 검토 및 최종의사결정을 거친다.
- C보험사는 고객에게 맞춤형 상품 추천 시 각 추천상품마다 주로 고객의 어떤 특성에 의거하여 이 상품이 추천되었는지를 명시하고 있다.
- D보험사는 맞춤형 상품 추천 후 어떤 보장의 강화를 원하는지 고객이 직접 입력하여 추천상품을 수정할 수 있도록 하는 기능을 갖추고 있다.

체크리스트 나-1-1-D

AI 시스템에 이용되는 학습데이터의 출처, 품질 등을 검증하고 개선필요시 조치를 취하였는가?

체크리스트

1) 학습데이터의 출처와 안정적인 데이터 수집 여부를 점검하였는가?

- 학습데이터 생성에 사용되는 원천데이터의 출처의 신뢰성과 데이터 수집의 안정성을 확인한다.
- 데이터 공급사와의 공급 방식에 대한 기록 및 공급내역 이력 보관 (메타기록, 내역 변경 등 변경 이력 보관)을 확인한다.
- 데이터 획득 과정에서 관련 법/제도 준수 여부를 확인한다.

예시 고객에게 수집데이터 활용범위를 명확하게 안내하였는지, 데이터 활용에 있어 고객 동의 유무를 확인하였는지, 고객 동의의 유효기간이 언제인지 등

- 시스템 개발 시 고객이 일부데이터에 대해 사용 동의를 철회했을 경우 이를 반영할 수 있는 로직을 검토한다.

예시 고객 동의의 유효기간 관리를 하고 있는지, 고객 동의 철회에 따라 즉시 데이터 활용을 중단하고 있는지 등

- AI 시스템의 효율성 강화를 위해 데이터를 통한 학습이 지속적으로 실시될 수 있도록 안정적인 데이터 수집의 가능 여부를 확인한다.
- 데이터 품질 관리를 위한 거버넌스 조직 또는 담당자가 금융기관 내에 있고, 양질의 데이터가 왜곡 없이 제공되도록 수시로 체크한다.
- 마이데이터 등 금융기관의 일부데이터는 외부 데이터 거래기관 또는 관계사 등으로부터 물리적으로 같은 공간에 놓여지는 경우가 있는데, 이러한 경우 상품추천 시스템에 적용할 수 없다.

예시 출처에 따라 데이터를 구분·관리하고 있는지, 출처가 다른 데이터를 오용 여부에 대해 상시 모니터링을 진행하고 있는지

YES □ | NO □ | N/A □

2) 학습데이터의 품질 확보를 위해 데이터의 대표성 · 정합성을 체크하였는가?

- 학습데이터가 AI 시스템이 적용될 대상(모집단)을 대표하는지 확인한다.
(개발 모집단 선정기준 및 개발대상 샘플 추출기준 점검)
- 데이터 품질관리에 대한 내규 존재 여부를 확인한다.
- 학습데이터 생성에 대한 라벨링 작업 시 원천데이터와 라벨링 데이터의 동기화 여부를 확인한다.
- 사용 데이터의 정합성 검증 여부를 확인한다.

예시 모델 개발 시 성별/나이/직업군 등에 따라 최소 샘플 수 및 비율을 일정하게 확보(밸런싱)하여 학습하고, 나이 등 연속적인 값의 경우 다른 변수값을 고정한 채 나이를 변경하여 일정한 방향으로 결과가 나오는지 체크하여 특이값 지점 체크

- 학습 데이터에 대한 조직의 품질 관리역량 강화를 위한 교육 및 지원 체계의 확보 여부를 확인한다.

YES □ | NO □ | N/A □

3) 재학습을 수행할 경우, 학습데이터를 갱신하여 데이터의 최신성을 확보하였는가?

- 내부 규정 또는 업무매뉴얼 등을 통해 재학습 및 재배포 절차 마련 여부를 확인한다.
- 재학습 및 재배포 절차가 데이터 최신성 및 적정성 유지에 적합한지 확인한다.
- 금융기관의 정보는 데이터의 형태에 따라 배치(batch) 주기가 정해지고, 상품추천 시스템은 다양한 배치 주기의 데이터를 조합해 완성된다. 따라서 신상품 출시 등 데이터가 갱신 또는 업데이트 될 때 최단 시간에 반영할 수 있는 로직을 개발 적용한다.

YES ☐ | NO ☐ | N/A ☐

4) 학습데이터의 출처, 사전처리, 가공 등의 주요 과정을 문서화하였는가?

- 데이터의 수집과 처리 업무를 위한 절차로서 수집처리 방법 및 기준에 대한 내부 규정 또는 업무매뉴얼 반영 여부를 확인한다.
- 외부 데이터를 활용하는 경우 데이터 출처에 대해 명확한 기록 여부를 확인한다.
- 개별 데이터 처리에 대한 처리 로직기록 등 통한 데이터처리(변환/합성 등)에 대한 기록 보관·유지 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A금융사는 맞춤형 상품추천 시스템 학습 시 입력하는 데이터는 고객기본정보 및 거래정보 등의 적합한 내부정도와 고객의 동의하에 수집된 정보를 활용하고 있다.
- B금융사는 맞춤형 상품 추천을 위한 데이터 수집 지침을 제정하고 정보보호위원회의 심의를 거쳐, 이 지침을 준수한 적법한 학습데이터를 활용하도록 하고 있다.
- C금융사는 학습데이터를 사전에 정의한 업데이트 기준에 맞추어 고객의 객관적 상황에 적합한 상품추천 결과를 제공하고, 주기적으로 업데이트하고 있다.

체크리스트 나-1-2-D

AI 시스템에 이용되는 학습데이터 또는 모형의 편향여부를 개발단계에서 테스트하고, 이를 완화하기 위한 적절한 조치를 취했는가?

체크리스트

1) AI 판단기준에 차별적인 요소가 들어가지 않도록 사전에 점검하였는가?

- 불합리한 차별을 방지하기 위하여 성별, 연령, 지역, 종교, 인종, 사회적 지위, 자산

등 불합리한 차별이 발생 가능한 평가요소를 선정하고 평가요소에 반영되었는지 점검한다.

예시 맞춤형 상품추천에 관련한 합리적·불합리한 차별요소를 구분하고, 상품내용에 따른 적정한 평가요소를 관리·변경·반영하고 있는지 등

YES ☐ | NO ☐ | N/A ☐

2) AI 모형의 편향성 판단지표를 선정하여 개발 단계별 편향수준을 테스트하고, 편향을 완화하기 위한 적절한 조치를 취하였는가?

- 내부 정책에 따른 편향성 판단지표를 선정하였는지 확인한다.
- 개발표본 선정, 라벨링, 모델링 단계 등 세부 단계별로 편향이 존재하는지 점검한다.

예시 상이한 평가요소(성별, 연령, 지역, 자산 등)를 지닌 모델을 각 세부 단계에 대입하여 단계별 편향여부를 점검하고, 이상이 있는 경우 수정 필요

- 업무 매뉴얼, 요구사항 문서 등을 통해 데이터 편향 완화 방안 마련 및 적용 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

〈 관련 법령에 따른 별도 체크리스트 〉

3) 개발 단계별 편향수준 테스트 및 완화조치에도 불구하고, 시스템 활용 결과시 발생 가능한 편향에 대한 통제를 위해 적절한 검증 방안을 마련하였는가?

- 개발 단계에서 인식하지 못한 AI 시스템이 내린 결과의 편향여부의 기준을 세우고, 편향성 여부를 검증하는 "AI 공정성 진단 검증도구" 등을 도입, 활용하여 점검한다.

예시 AI 편향성을 측정하고, AI 모델 문제점을 진단·교정하는 알고리즘을 활용하는 AI 공정성 진단 검증도구 운용

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A보험사는 맞춤형 상품 추천 AI 시스템을 위한 학습 데이터 갱신 시마다 편향성이 존재하는지 검토하고 평가해야 하는 자체 규정을 가지고 있다.
- B금융사는 고객차별적인 상품/서비스 추천을 모니터링할 소비자패널 회의를 정기적으로 운영한다.

체크리스트 나-2-1-D

AI 시스템에 개인정보·민감정보를 활용하는 경우, 해당 정보의 필요성을 평가하고 안전조치를 수행하였는가?

체크리스트

1) AI 시스템에서 개인정보·민감정보의 활용 필요성을 점검하였는가?

- AI 시스템의 활용 목적, 적용대상을 고려하여 개인정보·민감정보 활용의 필요성을 점검하고, 반드시 필요한 경우에만 활용한다.

예시 보험상품 추천시 필요한 개인정보(나이, 소득, 보험가입내역 등)과 불필요한 개인정보를 정기적으로 구분하고, 정보수집 동의여부에도 이를 반영하여 적합한 개인정보를 수집할 수 있도록 양식 등을 수정하였는지 확인

YES □ | NO □ | N/A □

2) AI 시스템에서 활용하는 개인정보·민감정보에 대해 안전조치를 취하였는가?

- AI 시스템의 개발을 위해 개인정보·민감정보를 수집·활용하는 경우 개인정보보호법 등 관련 법령을 준수하고, 활용 범위와 목적에 대한 명확한 기준을 마련했는지 확인한다.
- AI 시스템 개발시 관련 정보의 유출, 악용 가능성이 없도록, 기술적·관리적·물리적 통제 방안을 마련해야 한다.

예시 출처별 개인정보 데이터(마이데이터, 마케팅 등)를 명확히 구분·관리하고, 데이터 유출 억제 위한 망분리, 별도공간 확보 등 통제방안 검토·실시

- 개인정보·민감정보 파기에 관해 관련 사항이 준수되고 있는지 확인한다.
- 개인정보·민감정보의 수집·학습이 예상되는 경우, 초기 개발 단계에서 개인정보 수집과 생성에 대한 시스템을 점검할 수 있는 내부 절차를 구축했는지 확인한다.

예시 개인식별정보(전화번호, 주민등록번호, 전화번호 등) 탐색 프로그램을 단계별로 구동하여 마스킹처리, 암호화여부 등 필요한 보안조치의 구현 여부를 확인

YES □ | NO □ | N/A □

준수사례 (예시)

- A보험사는 상품/서비스 기획단계시 사전체크리스트를 도입하여 민감정보가 필요한지 검토한 후 AI 시스템에 반영한다.
- B은행은 맞춤형 상품 추천 AI 시스템을 위한 학습 데이터 전체를 비식별화 조치 완료해 운영하고 있다.
- C보험사는 고객정보, 민감정보가 포함된 데이터베이스는 중요 정보시스템으로 선정하여 접속 및 활용을 통제한다.

체크리스트 나-3-1-D

AI 시스템의 설명가능성을 고려하고, 설명가능한 AI 기술 도입방안을 검토 또는 대안을 마련하였는가?

체크리스트

1) AI 시스템 개발 과정에서 설명가능성을 확보하기 위해 노력하였는가?
- AI 서비스 성격을 감안하여 설명 대상 및 설명하는 절차를 검토한다.
- 결과의 오해석 방지를 위해 설명 공유 대상과 범위를 설정한다.
- 설명가능한 AI 기술을 검토하고 설명방안을 마련하였는지를 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템 개발 과정에서 설명가능한 AI 기술 적용이 어려운 경우, 대안을 마련하였는가?
- 설명가능한 AI 기술 개발 트렌드를 충분히 확인하여 적용이 어려운 타당한 이유와 이에 대한 대안이 충분히 마련되었는지 확인한다.
- 업무 매뉴얼, 요구사항 문서 등에 AI 시스템이 생성한 결과를 설명하기 위한 기법 검토 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) AI 시스템 개발 과정에서 AI 시스템에 의한 결과를 고객에게 설명하기 위한 절차를 고려하였는가?
- 맞춤형 상품추천 관련 금융소비자보호법 제19조(설명의무)의 준수 여부를 확인한다.
- 고객에게 추천된 금융상품의 주요 내용을 설명할 수 있는 절차가 마련되어 있으며 고객에게 적절한 안내가 이루어지고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

〈금융소비자보호법 제19조(설명의무) 관련 준수사항〉

- 금융회사는 맞춤형 상품추천 관련 AI 시스템을 활용하여 일반금융소비자에게 계약 체결을 권유하거나 설명을 요청받을시 아래 사항을 이해할 수 있도록 설명하여야 한다.
- 금융상품별 주요내용(보장성·투자성·예금성·대출성 상품별 주요사항 상이), 연계·제휴 서비스 관련사항(내용·이행책임·제공기간 등), 청약철회 관련사항(기한· 행사방법·효과), 민원처리 및 분쟁조정 절차, 예금자보호법 등 타법에 따른 보호여부, 일반금융소비자의 의사결정 지원 및 권익보호를 위해 금융위 고시사항 관련

준수사례 (예시)

- A은행은 맞춤형 상품 추천 AI 시스템 학습 시마다 금융소비자보호 총괄부서와 협의하여 금융상품의 위험도와 복잡성이 반영되었는지 확인 후 운영하고 있다.
- B보험사는 고객관련 시스템 개발 시 내부통제시스템 사전 검토를 통하여 금융소비자보험법 및 금융 관련 법규 준수를 사전에 검토한다.
- C은행은 맞춤형 상품 추천 서비스에 각 추천 상품이 어떤 사유로 추천되었는지, 그리고 어떤 장단점을 가지고 있는지 함께 확인할 수 있도록 서비스를 제공하고 있다.
- A카드사는 카드 대출 상품의 금리 및 최대 이용가능금액을 AI 서비스를 이용하여 제공한다면, 대출 상품 신청 전, 사용자가 통상적으로 인지할 수 있는 영역에 '라. 대출성 상품' 의 안내를 제공한다.

체크리스트 다-1-1-D

AI를 활용한 시스템별 성능 지표 선정 및 목표 수준을 설정하고 충족여부를 확인하였는가?

체크리스트

1) AI 기반 모형의 성능을 평가하기 위한 지표를 선정하고 있는가?

- AI 모델의 유형에 따라 추천 모형의 성능을 측정할 수 있는 지표를 산출하여 활용하고 있는지 확인한다.

예시 Accuracy, Hit Ratio, MRR, NDCG 등의 지표

- 2개 이상의 개별 샘플 리뷰를 통해 전문가 집단의 경험적·이론적 예상과 부합 여부를 검토하고 있는지 확인한다.
- 성능 및 안정성 확보를 위한 모니터링 기준, 사용자 특성에 따른 설명 평가 기준 등을 수립하여 작성한 업무 매뉴얼에 따라 기록을 보관하고, 내역 변경 등 변경 이력을 보관하고 있는지 확인한다.
- 사전 정의된 데이터 명세를 벗어나는 값의 집계, 과거 입력변수와 최신 입력변수의 차이가 허용 임계값(현업 적용 시 허용 가능한 수준에서 정의) 이상으로 발생하는지에 대한 여부 등의 기준을 마련하고, 해당 모형이 적정한 성능으로 운영되고 있음을 판단할 수 있는 최소 기준이 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 선정한 성능 평가 지표에 따라 목표 수준의 달성 여부를 점검하고 미달하였을 경우 조치하고 있는가?

- 성능 평가 결과가 목표 수준에 미달하였을 경우 원인을 분석하여 모형 재개발 여부를 검토하여 조치한다.
- 특정 학습데이터에 과적합이 되지 않았는지 다양한 테스트를 수행하고, 테스트 데이터에 따른 변동성을 모니터링하고 있는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A은행은 맞춤형 상품추천 AI시스템 학습 후 테스트 시 전문가 집단의 상품 선택과의 유사성을 측정하여 내부기준을 만족하였는지 평가하고 있다.
- B보험사는 맞춤형 상품추천 AI시스템의 추천 결과가 지점/사람이 하는 추천시스템과 유사한 결과가 나오는지 확인하기 위하여, 영업지점과 상품개발부서에 확인하는 절차를 거친다.
- C금융사는 대 고객 관련 시스템의 시스템오류, 추천 엔진의 성능을 모니터링 하는 시스템을 도입하였다
- D금융사는 모델 및 데이터의 형상관리를 통하여 상품추천시스템이 어떠한 고객/금융정보를 활용하여 AI/ML 모델을 만들었는지 관리한다.
- E카드사는 데이터 업데이트 주기, 모델개발 주기 및 운영시스템 반영을 정기적으로 진행하고 있다. 또한 신상품 및 제도 변경 시 이를 반영하기 위하여 비정기 반영도 진행한다.

체크리스트 다-2-1-D

불합리한 차별이 나타나지 않도록 공정성 판단기준을 설정하고 충족여부를 평가하고 개선하였는가?

체크리스트

1) 공정성 목표 수준 및 공정성 판단 지표를 선정, 관리하는가?

- 학력, 성별, 연령, 종교 등 불합리한 차별 요소를 선정하여 차별이 발생하고 있지 않은지 점검하고 있는지 확인한다.

예시 맞춤형 상품추천에 관련한 합리적·불합리한 차별요소를 구분하고, 상품내용에 따른 적정한 평가요소를 관리·변경·반영하고 있는지 등

- 공정성 목표 수준 및 공정성 판단 지표를 선정 및 관리하고 있는지 확인한다.

예시 데이터 샘플링을 통해 데이터 분포 검증을 수행하는지 확인한다.
관련 업무 매뉴얼에 따라 공정성 평가 지표별로 고객의 잠재적 영향 정도를 종합적으로 고려할 수 있는 관리 방안을 마련하였는지 확인한다.
공정성 판단 지표를 지속적·주기적으로 모니터링할 수 있는 프로세스를 구축하고, 관리·평가 체계를 마련하였는지 확인한다.

YES □ | NO □ | N/A □

2) 공정성 판단지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 기술적, 관리적 방안을 검토하여 조치하는가?

- 편향방지 방법론을 활용하는 등 편향을 완화하기 위한 방안을 검토하였거나 조치하였는지 확인한다.
- 고객의 공정한 선택을 돕기 위해 상품설명서, 금융회사 홈페이지 내 상품 안내페이지, 모바일 앱 등의 채널을 통해 AI시스템에서 사용한 정보(데이터 고려항목, 추천배열 방법 등)를 쉬운 용어를 사용하여 안내하고 있는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A금융사는 상품추천 결과를 사내부서 및 고객패널 등을 검토를 통하여 평가하는 프로세스 후 서비스를 제공한다.
- B은행은 상품추천시스템을 분기에 1회 업데이트하며, 신상품 출시 제도 변경시 추가 업데이트 한다.
- C보험사는 맞춤형 상품 추천 서비스에서 고객에게 추천 상품과 함께 주로 어떤 측면을 고려했는지에 대한 안내가 함께 제공된다.

체크리스트 다-3-1-D

AI 시스템이 학습한 모형이 상황에 맞게 설명 가능한지 확인하고, 설명가능성을 적법한(또는 합리적인) 수준으로 개선하고자 노력하였는가?

체크리스트

1) 고객에 대한 설명의무, 금융서비스의 위험 수준 등을 고려하여 고객 및 이해관계자 등 설명 대상자들에 대해 AI 시스템 설명가능성 수준을 평가할 수 있는 기준 및 절차가 수립되어 있는가?

- 내부 규정, 업무매뉴얼 등에 설명가능성 수준에 대한 평가 기준 및 절차가 수립되어 있는지 확인한다.

예시 추천 알고리즘을 통해 나온 결과를 일정 기간 트래킹하여 실제값과 비교 → 모델의 정확도 계산

- 맞춤형 상품 추천 AI 시스템의 목적 및 모형 알고리즘의 특성에 따라 신뢰도 필요성 여부 및 신뢰도 산출 가능한지 판단 기준을 확보하였는지 확인한다.
- 정확도, 정밀도, 재현율 등 모형 성능지표 이외에 신뢰도 지표 선정 방안이 마련되어 있는지 확인한다.
- 신뢰도 지표 관리를 통해 신뢰도 수준별 의미가 정의되어 있는지 확인한다.

YES □ | NO □ | N/A □

2) AI 시스템의 설명가능성이 기준이 적법한(또는 합리적인) 수준에 도달하지 못할 경우 개선 방안(재학습이나 알고리즘 수정 등)을 검토하여 조치하고 있는가?

- 설명가능성 관련한 법·제도, 가이드라인 등이 있는 경우 해당 원칙에 대한 준수 여부를 확인한다.
- 추천 결과에 부정적 요소가 발견되거나, 이용자가 불만을 제기하는 경우, 추천 원인 주요 사유를 제공하고, 만족도를 조사하여 그 결과를 향후 모델링에 반영하는 체계가 마련되어 있는지 확인한다.
- AI 시스템의 맞춤형 상품 추천에 대해 소비자의 주요 불만 제기시 금융소비자보호법 및 금융소비자보호에 관한 감독규정 내부통제기준에 따라 보고 및 개선을 위한 체계가 마련되어 있는지 확인한다.

예시 금융소비자보호감독규정 별표2의 금융소비자보호 내부통제위원회 운영에 관한 사항

YES □ | NO □ | N/A □

준수사례 (예시)

- A금융사는 상품추천시스템이 정확한 추천을 하는지에 대한 측정지표를 갖추고 있으며, 영업부서 및 상품개발 부서에 정기적인 확인절차를 거친다.
- B은행은 상품추천시스템을 매월 추가학습, 반기 모형재학습 및 비정기적으로 업데이트를 진행하여 신뢰도 저하를 사전에 예방한다.
- C보험사는 고객이 원할 시 추천도 상위 5개를 추천하도록 하고 각각에 대한 만족도를 조사하여 학습 데이터에 반영할 수 있도록 하고 있다.
- D금융사는 상품추천결과에 대한 부정적인 VOC(Voice Of Customer) 접수 시 CCO (Chief Customer Officer) 산하 정례회에 보고되어 개선하는 방향을 도출 한 후 개선하는 절차를 시행한다.

체크리스트 라-1-1-D

AI시스템 제공에 있어 AI 이용여부, 설명·이의제기권 등 고객의 권리 및 이의신청·민원제기 방법 등 소비자로서의 권리구제 방안을 마련하여 고지하였는가?

체크리스트

1) AI 시스템 적용 사실과 범위에 대해 고객에게 안내가 이뤄지고 있는가?

- 상품설명서, 홈페이지, 모바일 앱 등의 채널을 통해 AI 시스템 활용 여부와 목적을 쉬운 용어를 사용하여 안내하고 있는지 확인한다.

예시 사용자에게 컨텐츠를 추천시 사용자가 기존에 소비한 컨텐츠에 대한 카테고리, 텍스트, 이미지 데이터 등을 사용하여 비슷하거나 특별한 관계가 있는 다른 콘텐츠를 추천하는 “컨텐츠 기반 필터링”을 사용하고 있음을 사전에 안내

- 고객의 권리에 대한 사전고지 여부 및 변경 이력을 확인한다.
- 맞춤형 상품 추천 서비스를 설명하는 별도의 안내 페이지를 구성하고, 적용된 맞춤형 상품 추천 서비스의 제공 방식 및 알고리즘 등에 대하여 설명을 제공하는지 여부를 확인한다.
- 맞춤형 서비스를 안내하는 영역 내에, 적용된 맞춤형 상품 추천 서비스에 대하여 인풋 데이터에 따른 서비스 결과 예시가 일반고객이 쉽게 이해할 수 있는 수준으로 최소 2가지 이상 있는지를 확인한다.

YES □ | NO □ | N/A □

2) AI 시스템으로 인한 고객의 불편 및 불이익에 대한 구제 방안이 마련되어 있는가?

- 내부 규정 또는 업무 매뉴얼에 신고 및 이의제기 절차가 구체적으로 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 전담 조직 및 인력을 확인한다.
- 홈페이지, 모바일 앱 등의 채널을 통해 고객에게 적절한 권리구제 절차 및 기준 등 사전고지 여부를 확인한다.
- 맞춤형 서비스 결과에 대하여 설명하고 이의 제기에 관하여 답변하는 운영 정책(예를 들면, 담당부서/회신기일 등)의 여부를 확인한다.
- 맞춤형 서비스를 안내하는 영역 내에, 맞춤형 서비스 관련 설명을 요구할 수 있다는 내용과 이의를 제기할 수 있다는 내용이 안내되어 있는지를 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A금융사는 VOC프로세스 고객패널, 디지털패널 등을 통하여 상품추천 시스템에 대한 고객의견을 접수하고 있으며, 수집된 의견을 상품추천시스템 업데이트에 활용하고 있다.
- B카드사는 고객센터의 소비자포털을 통하여 법령에서 정의하고 있는 고객의 권리 안내 및 민원 창구를 소개하고 있으며, AI 서비스에 대해서도 특정 채널을 통해 고객의 불편 및 구제 방안에 대해 안내하고 있다.

체크리스트 라-2-1-D

AI 시스템의 성능을 주기적으로 모니터링하며 성능 개선이 필요한지의 여부를 확인하고 있는가?

체크리스트

1) AI 시스템 성능이 안정적으로 유지되는지 확인할 수 있는 모니터링 절차를 마련하였는가?

- 성능 모니터링 주기, 범위 및 보고 등의 절차가 마련되어 있는지 확인한다.
- 성능 목표값에 대한 설정을 사전에 정의하고, 목표값 이하로 하락하는 경우, 파악할 수 있는 체계를 마련한다.
- 모니터링 결과 및 조치 이력의 문서화 여부 및 기록 내용을 확인한다.
- 사용 데이터의 최신화 또는 유효기간에 관한 정책 수립 여부를 확인한다.

예시 고객의 맞춤형 상품추천 관련 데이터 최신화 위한 주기를 설정하고, 이에 따른 업데이트가 구현되고 있는지 정기 모니터링 시기를 적절히 설정 필요

- 맞춤형 상품 추천 시스템에 적용된 모형의 온/오프라인 지표의 모니터링 시스템 존재 여부 또는 주기적인 모니터링의 방식, 범위, 결과확인에 관한 프로세스 수립 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 유의미한 성능 하락 및 모집단 특성 변화에 따라 모형을 변경 또는 개선 필요시 의사결정 단계나 절차를 사전에 정의하고 있는가?

- 주기적인 성능평가와 재학습 필요성을 검토하고, 필요 시 모형을 교체하거나 개선을 진행해야 하며 모형 변경 또는 개선을 위한 의사결정 단계와 절차가 마련되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) AI 재학습 및 모형 개선 절차를 수립·이행하고 있는가?

- AI 재학습 및 모형 개선 절차 수립과 이행 여부를 확인한다.
- 개선 방안 수립 시 데이터의 최신성, 정확성, 정합성, 비편향성을 고려하였는지 확인한다.
- 학습데이터 변경 이력 기록 여부 및 관리 문서를 확인한다.
- 활용된 데이터의 보관기간 준수 여부(5년, 10년 등)를 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A금융사는 정보보호위원회, 사전체크리스트, 데이터 활용기간/폐기주기 등 적용하여 데이터입수/활용/폐기 등을 적법하게 시행하고 있다.
- B은행은 데이터특성에 맞추어 실시간 외부데이터 확보, 일/주/월 데이터 업데이트를 진행하고 있다.
- C보험사는 최소 1개월에 한번 최신 경향이 반영된 데이터를 통해 맞춤형 상품 추천 AI 시스템의 재학습이 이루어지도록 하고 있다
- D금융사는 AI상품추천시스템에 나온 결과를 오프라인에서 활용 시에 해당결과를 피드백/모니터링 하는 프로세스를 갖추고 있다.
- E금융사는 AI상품추천시스템에 나온 결과를 온라인에서 활용 시에 해당 결과를 피드백/모니터링 하는 프로세스를 갖추고 있다.

체크리스트 라-3-1-D

AI 시스템 및 AI 시스템에서 사용하는 데이터의 오용·악용 가능성을 최소화하였는가?

체크리스트

1) 보안 시스템 구축 등을 통해 AI 상품 추천시스템에 대한 보안대책을 수립하여 적대적 공격 등 오용·악용 가능성을 최소화하고 있는가?

- AI 상품 추천시스템 구축 시 시스템, 인프라, 데이터, 네트워크, 이용자 보호 등과 관련하여 다양한 보안 위협 및 대응조치를 포함한 적합한 보안대책을 수립하고 검토하고 있는지 확인한다.
- 사용자 상품거래 내역 유출 등 비정상 동작이나 예기치 못한 오류에 대한 대책을 수립하고, AI 상품 추천시스템에 대한 접근 권한을 관리하고 있는지 확인한다.
- 사용자 상품거래 내역 등에 대한 개인정보 암호화 및 권한별 접근 통제 조치 여부를 확인한다.

예시 고객의 맞춤형추천 관련 이력, 내용 등이 타인에게 노출되지 않도록 통제

- AI 상품 추천시스템을 운용하는 중요 서버에 백신 프로그램을 설치하고 주기적 업데이트 및 악성코드 점검, 실시간 검사 설정을 하고 있는지 확인한다.
- 보안 취약점에 대한 사전·정기 점검 수행 및 취약점 조치 여부를 확인한다.
- AI 상품 추천시스템 관련 침해사고 분석 시 필요한 로그에 대하여 보존 및 검토에 대한 정책을 수립하였는지 확인한다.
- AI 상품 추천시스템의 성능을 주기적으로 점검하여 성능 저하를 유발하는 공격 여부를 확인하고, 특히 학습데이터를 교란하는 등 데이터 오염 공격의 경우 온라인을 통해 이뤄지는 경우가 많으므로 AI의 트래픽, 로그와 이벤트의 상관관계 분석, 패턴 기반 탐지 등을 통한 보안조치를 지속 강화한다.
- AI 상품 추천시스템의 성능을 주기적으로 점검하여 성능 저하를 유발하는 공격 여부를 확인한다.
- AI 상품 추천시스템 관련 로그별 보존기간 및 검토 주기를 지정하였는지 확인한다.
- AI 상품 추천시스템 테스트 시 개인신용정보가 아닌 임의의 데이터를 생성하여 테스트를 수행하는지 확인한다. 단, 개인신용정보 이용이 불가피한 경우, 책임자의 승인 절차에 따라 가공하여 사용하고 즉시 폐기하는 등 관리대책이 수립·이행되었는지 확인한다.
- 암호화 대상, 사용 암호 알고리즘, 암호키 관리 방안을 포함한 개인정보 암호화 정책을 수립하였는지 확인한다.

- 정보자산 등 운영체제, 소프트웨어 패치 관리정책 및 절차를 수립·이행하고 인터넷 직접 접속을 통한 패치를 제한하고 있는지 확인한다.
- AI 상품 추천시스템 개발 과정에서 적대적 공격 등 오용·악용 가능성을 최소화하기 위해 보안성 검증을 실시하고 있는지 확인한다.

YES □ | NO □ | N/A □

2) 보안 시스템 구축 등을 통해 고객 또는 제3자에 의한 데이터 오염 공격 등을 통한 데이터의 오용·악용 가능성을 최소화하고 있는가?

- AI 상품 추천시스템에 사용되는 데이터 오용 · 악용 여부 감지 지표 설정 및 주기적 확인 여부 등 절차와 대처방안 수립 · 운영 여부를 확인한다.
- 잘못된 개인신용정보 주입·조작 등 학습데이터 변조 여부를 주기적으로 확인한다.
- AI 추천시스템에 사용되는 데이터의 무결성, 기밀성을 유지하기 위한 관리 규정을 준수 여부를 확인한다.

YES □ | NO □ | N/A □

3) 오픈소스(프레임워크, 라이브러리 등) 보안 취약성 관리를 위한 체계를 수립하여 AI 시스템의 보안성을 강화하고 있는가?

- AI 상품 추천시스템에 사용되는 오픈소스 사용 시 라이브러리 보안 취약점 확인 및 통지체계 수립 여부를 확인한다.
- 맞춤형 상품 추천 관련 AI 알고리즘의 취약점 점검 여부를 확인한다.
- 오픈소스 취약점을 주기적으로 확인하여 업데이트 및 패치를 적용하고 있는지 확인한다.

YES □ | NO □ | N/A □

4) 침해사고 및 재해 등을 예방하기 위한 체계 및 침해사고 또는 재해가 발생했을 때 피해 확산 · 재발 방지와 신속한 복구를 위한 체계를 갖추고 있는가?

- 해킹, 악성코드, 논리폭탄, 메일폭탄, 서비스거부 또는 고출력 전자기파 등의 방법으로 사고 발생 시에 대비한 예방·복구 체계 마련 여부를 확인한다.

  예시 맞춤형추천 관련 고객의 개인정보, 신용정보, 이용이력 등을 복구할 수 있는 체계를 갖춰, 사고 이후에도 이전과 동일한 상품추천 알고리즘을 구현할 수 있도록 지원

- 정상적 보호 · 인증절차를 우회하여 정보통신망에 접근할 수 있도록 하는 프로그램이나 기술적 장치 등에 대한 차단 체계 마련 여부를 확인한다.

• 침해사고 발생 시 기록 및 보고, 신고 및 통지, 비상 연락체계 등 내용이 포함된 침해사고 대응 절차를 수립하였는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

• A금융사는 추천시스템을 시스템로그 및 추천이력을 내부통제 규정에 의해 생성/보관/폐기 운영한다.

• B금융사는 악성코드 공격에 대응하기 위하여 보안패치 업데이트, 비밀번호 정기변경, 이중화 시스템 구축 등을 시행한다.

• C보험사는 오픈소스 활용에 따른 보안취약점 보완하기 위하여 주기적으로 보안패치 현황을 모니터링하고 업데이트 및 패치 등을 진행한다.

• D보험사는 데이터를 관리하는 서버와 AI시스템을 운용하는 서버에 침입탐지 시스템을 도입하고 이에 대한 모니터링을 수행하여 학습데이터와 모형을 보호하고 있다.

E. 로보어드바이저

체크리스트 가-1-1-E

AI를 활용하는 목적이 명확하게 정의되고, 윤리원칙 부합 여부와 AI 활용에 따른 영향도와 잠재적 피해 가능성을 점검하였는가?

체크리스트

1) AI 도입시, 기존 업무에 대한 영향도 분석 등을 통해 도입 타당성 검토가 이루어졌는가?

• 기존 투자자문 업무와의 비교 및 도입 시 기대효과 분석 등을 통해 업무 영향도 분석 및 타당성 검토가 이루어졌는지 확인한다.

예시 AI 도입 전·후 업무 절차, 담당자 역할 변화, 세부 업무별 AI 적용방식 적용 여부 등

※ 기존 업무가 존재하지 않는 경우, 생략 가능하다.

YES ☐ | NO ☐ | N/A ☐

2) AI 활용 목적과 업무범위, 역할 등이 명확하게 정의되어 있는가?

- 투자자문 · 일임, 집합투자 재산 운용 등 업무 중 AI 시스템을 활용한 로보어드바이저 서비스의 업무 범위, 활용 목적, AI의 역할 등이 명확하게 정의되어 있는지 확인한다.

YES □ | NO □ | N/A □

3) AI 활용 목적이 해당 기관의 AI 윤리원칙*에 부합하는가?

* 인권보장, 프라이버시 보호, 다양성 존중, 공공성, 책임성, 안전성, 투명성 등

- 도입할 로보어드바이저 서비스가 내부 AI 윤리원칙 등에 부합하는지 확인한다.

YES □ | NO □ | N/A □

4) AI 시스템의 기획 및 설계 단계에서 고객에게 미치는 영향, 위험수준, 잠재적 피해 가능성 등이 고려되었는가?

- 로보어드바이저를 위한 AI 시스템이 내부 AI 윤리원칙에 근거한 각 금융회사의 절차에 따라 AI 모형의 위험수준 평가기준이 마련되어 있는지 확인한다.
- 집합투자나 투자일임업 수행을 위해 로보어드바이저 서비스가 사용되는 경우, 오류나 오작동으로 인한 고객 피해 가능성을 검토하고 방지방안을 마련하였는가?
- 로보어드바이저 서비스를 운영하는 금융회사의 이익을 고객의 이익보다 우선하도록 설계하여 이해상충 문제가 존재하지는 않는지 확인하고, 이에 대한 검토 및 방지방안이 마련되어 있는지 확인한다.

YES □ | NO □ | N/A □

준수사례 (예시)

- A증권사는 로보어드바이저 기획 · 설계 시 로보어드바이저 시스템이 인간의 기본적 권리를 침해하지 않고 고객의 자율적 행동 및 결정을 방해하지 않는지 여부 등에 대한 내부 검토를 통해 윤리원칙에 부합토록 개발 · 운영하고 있다.

- B증권사는 로보어드바이저 시스템이 특정 개인 또는 집단의 이익만을 대변하거나, 성별 · 인종 등에 편향되지 않도록 기획 · 설계 단계부터 내부 검토를 통해 윤리 원칙에 부합토록 개발 · 운영하고 있다.

체크리스트 가-2-1-E

AI 시스템이 인간의 의사결정을 전면적으로 대체 또는 중요 의사결정을 대체하는 경우, 감독 · 통제 절차가 마련되어 있는가?

체크리스트

※ AI시스템이 인간의 의사결정을 대체하는 경우가 없다면, 동 체크리스트는 "N/A" 처리할 수 있다.

1) 의사결정을 AI가 대체하는 경우 발생할 수 있는 잠재적 위험 가능성을 확인하고, 이에 대한 감독 · 통제 절차가 수립되어 있는가?

- AI를 활용하는 경우에도 기존 투자자문, 투자일임 등의 업무에서 일반적으로 준수하고 있는 내부통제 절차가 수행되고 있는지 확인한다.
- AI 활용에 관한 내부 업무규정에 마련된 위험평가 기준과 금융소비자 권리에 중대한 위험이 발생할 가능성을 기준으로 위험평가 결과의 적정성을 확인한다.
- 내부 규정 또는 업무 매뉴얼을 통해 운영 중인 AI 시스템의 위험평가 결과 보고 절차의 마련 여부와 결과 보고 및 승인 여부를 확인한다.
- AI가 기존 투자자문 업무를 대체하는 경우 발생할 수 있는 잠재적 위험에 대해 예측하여 사전 · 사후 대책이 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 업무별 담당자 지정 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) 의사결정을 AI가 대체하는 경우, 인간의 개입이 필요한 경우에 대한 관련 절차가 정의되어 있는가?

- AI 기반의 로보어드바이저 시스템 의사결정에 인간의 개입이 필요한 경우에 대한 업무 매뉴얼 또는 관련 내부 절차가 마련되어 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 의사결정을 AI가 대체하는 경우, 관리자, 사용자 및 기타 이해관계자가 해당 의사결정 과정에 대해 해석 및 추적이 가능하도록 설계되어 있는가?

- 로보어드바이저 서비스를 수행하는 AI 시스템의 결과에 대한 설명방안 설계 여부 등 자동 의사결정에 대해 추적 및 설명방안 마련이 이루어졌는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A증권사는 기획 · 설계 단계부터 고객 금융자산의 불합리한 피해 발생 가능성을 검토하여 감독 · 통제 절차를 수립하고 있으며, 서비스시에는 해당사항을 고객에게 사전 고지하는 한편, 로보어드바이저 시스템 상에서 이상거래 탐지 등으로 인해 거래 중지 등의 피해가 발생할 경우 이를 신고하고 의견을 제시할 수 있는 절차를 마련해 홈페이지 등을 통해 고객에게 사전에 고지하고 있다.

체크리스트 나-1-1-E

AI 시스템에 이용되는 학습데이터의 출처, 품질 등을 검증하고 개선필요시 조치를 취하였는가?

체크리스트

1) 학습데이터의 출처와 안정적인 데이터 수집 여부를 점검하였는가?

- 학습데이터 생성에 사용되는 원천데이터의 출처의 신뢰성과 데이터 수집의 안정성을 확인한다.
- 데이터 공급사와의 공급 방식에 대한 기록 및 공급내역 이력 보관 (메타기록, 내역 변경 등 변경 이력 보관)을 확인한다.
- 데이터 획득 과정에서 관련 법/제도 준수 여부를 확인한다.
- 시스템 개발 시 고객이 일부데이터에 대해 사용 동의를 철회했을 경우 이를 반영할 수 있는 로직을 검토한다.
- 데이터 품질 관리를 위한 거버넌스 조직 또는 담당자가 금융기관 내에 있고, 양질의 데이터가 왜곡 없이 제공되도록 수시로 체크한다.

YES ☐ | NO ☐ | N/A ☐

2) 학습데이터의 품질 확보를 위해 데이터의 대표성 · 정합성을 체크하였는가?

- 학습데이터가 AI 시스템이 적용될 대상(모집단)을 대표하는지 확인한다.
  (개발 모집단 선정기준 및 개발대상 샘플 추출기준 점검)
- 전담 인력의 전문성 확보 기준 및 업무분장표 상 관련 업무연관 조직 구성 여부를 확인한다.
- 데이터 품질관리에 대한 내규 존재 여부를 확인한다.
  ※ 데이터 공급에 문제가 생길 경우를 대비한 Contingency Plan 등
- 학습데이터 생성에 대한 라벨링 작업 시 원천데이터와 라벨링 데이터의 동기화 여부를 확인한다.
- 사용 데이터의 정합성 검증 여부를 확인한다.

※ 물리적 검증(건수 검증, 합계 검증, Pysical R-I 검증) 및 논리적 검증(Cross 검증, Logical R-I 검증) 방식 등 활용

- 학습 데이터에 대한 조직의 품질 관리역량 강화를 위한 교육 및 지원 체계의 확보 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) 재학습을 수행할 경우, 학습데이터를 갱신하여 데이터의 최신성을 확보하였는가?

- 내부 규정 또는 업무매뉴얼 등을 통해 재학습 및 재배포 절차 마련 여부를 확인한다.
- 재학습 및 재배포 절차가 데이터 최신성 및 적정성 유지에 적합한지 확인한다.

YES ☐ | NO ☐ | N/A ☐

4) 학습데이터의 출처, 사전처리, 가공 등의 주요 과정을 문서화하였는가?

- 데이터의 수집과 처리 업무를 위한 절차로서 수집처리 방법 및 기준에 대한 내부 규정 또는 업무매뉴얼 반영 여부를 확인한다.
- 외부 데이터를 활용하는 경우 데이터 출처에 대해 명확한 기록 여부를 확인한다.
- 개별 데이터 처리에 대한 처리 로직기록 등 통한 데이터처리(변환/합성 등)에 대한 기록 보관 유지 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

※ 로보어드바이저 서비스는 사람의 투자판단을 AI가 대체하면서 고객 및 금융투자회사에게 새로운 투자 기회를 제공할 수 있다는 장점이 있는 반면, 투자에 따른 손실 위험은 항상 상존하고 있다. 고객이 직접 경험하는 AI 서비스인 만큼 이로 인한 고객의 투자 손실 위험에 대한 데이터의 충분한 검토가 필요하다.

준수사례 (예시)

- A자산운용사는 데이터 품질 확보를 위해 입수한 데이터에 대해 물리적 검증(건수 검증, 합계 검증, Pysical R-I 검증) 및 논리적 검증(Cross 검증, Logical R-I 검증) 방식 중 2가지 이상의 방식을 혼합한 검증 체계를 마련하여 데이터 정합성을 확보하고 있으며, 데이터 공급에 문제가 생길 경우를 대비해 Contingency Plan을 마련하여 데이터 대체 공급 방안을 준비해두고 있다.
- B사는 데이터 출처 및 처리 과정의 투명성을 확보하기 위해 데이터 공급사, 공급 방식, 공급 내역, 개인정보 수집 동의 등의 내용을 보관하고 있으며, 업무 매뉴얼을 통해 표준 데이터 처리 과정을 규정화하여 관리하고 있다.
- C증권사는 데이터 품질 관리 인력에 대한 기준(데이터 관련 업무 유관경력 5년 이상)을 내부 업무 매뉴얼에 마련하고 있으며, 데이터 품질 관리를 위해 데이터품질관리팀을 전문 인력으로 구성하여 업무 매뉴얼에 따라 데이터 품질을 통제하고 있다.

체크리스트 나-1-2-E

AI 시스템에 이용되는 학습데이터 또는 모형의 편향여부를 개발단계에서 테스트하고, 이를 완화하기 위한 적절한 조치를 취했는가?

체크리스트

1) AI 판단기준에 차별적인 요소가 들어가지 않도록 사전에 점검하였는가?

- 불합리한 차별을 방지하기 위하여 성별, 연령, 지역, 종교, 인종, 사회적 지위, 자산 등 불합리한 차별이 발생 가능한 평가요소를 선정하고 평가요소에 반영되었는지 점검한다.

YES □ | NO □ | N/A □

2) AI 모형의 편향성 판단지표를 선정하여 개발 단계별 편향수준을 테스트하고, 편향을 완화하기 위한 적절한 조치를 취하였는가?

- 내부 정책에 따른 편향성 판단지표를 선정하였는지 확인한다.
- 개발표본 선정, 라벨링, 모델링 단계 등 세부 단계별로 편향이 존재하는지 점검한다.
- 업무 매뉴얼, 요구사항 문서 등을 통해 데이터 편향 완화 방안 마련 및 적용 여부를 확인한다.

YES □ | NO □ | N/A □

※ 로보어드바이저 서비스는 투자판단에 따른 손실위험이 항상 존재하고 있으므로, 동일한 조건에 있는 고객이 다른 이익을 취하거나 손해를 받지 않도록 검토하여야 한다.

준수사례 (예시)

- A사는 통계적 판단을 위한 정량적 지표를 업무 매뉴얼에 반영해 관리하고 있으며, 이를 통해 사용 데이터 및 모델의 품질 및 편향성을 관리하고 있다.
- B사는 로보어드바이저 알고리즘이 상대적으로 자산이 많은 부유한 층에 더 유리한 결과가 도출되는 것을 발견하고 데이터 편향을 제거하고 수정하였다.

체크리스트 나-2-1-E

AI 시스템에 개인정보·민감정보를 활용하는 경우, 해당 정보의 필요성을 평가하고 안전조치를 수행하였는가?

체크리스트

1) AI 시스템에서 개인정보·민감정보의 활용 필요성을 점검하였는가?

- AI 시스템의 활용 목적, 적용대상을 고려하여 개인정보·민감정보 활용의 필요성을 점검하고, 반드시 필요한 경우에만 활용한다

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템에서 활용하는 개인정보·민감정보에 대해 안전조치를 취하였는가?

- AI 시스템의 개발을 위해 개인정보·민감정보를 수집·활용하는 경우 개인정보보호법 등 관련 법령을 준수하고, 활용 범위와 목적에 대한 명확한 기준을 마련했는지 확인한다.
- AI 시스템 개발시 관련 정보의 유출, 악용 가능성이 없도록, 기술적·물리적 통제 방안을 마련해야 한다.
- 개인정보·민감정보 파기에 관해 관련 사항이 준수되고 있는지 확인한다.
- 개인정보·민감정보의 수집·학습이 예상되는 경우, 초기 개발 단계에서 개인정보 수집과 생성에 대한 시스템을 점검할 수 있는 내부 절차를 구축했는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

※ 로보어드바이저 서비스는 고객의 자산상황, 직업, 학력 등을 기초로 고객에게 맞는 최적의 투자자문을 제공하는 서비스로 개인의 중요한 민감정보를 활용하고 있다. 이러한 개인정보 및 민감정보를 보호할 수 있는 적절한 조치를 취하여야 한다.

준수사례 (예시)

- A증권사는 로보어드바이저와 관련하여 개인정보를 수집하는 경우 개인정보보호법과 내부 규정에 따라 정보의 사용 목적 및 범위 등을 고지하고 정보 주체의 동의를 받고 있다.
- B증권사는 로보어드바이저와 관련하여 개인정보를 이용하거나 제3자에게 제공하는 경우 개인정보보호법과 내부 규정에 따라 당초 수집 목적과의 관련성, 정보주체의 이익에 대한 부당한 침해 가능성, 가명처리 또는 암호화 등 안전성 확보 조치 여부 등을 검토하고 있다.
- C증권사는 로보어드바이저 업무와 관련하여 수집·이용한 개인정보를 보관하거나 파기할 경우 개인정보보호법과 개인정보처리방침에 따라 보관하고 파기원칙, 파기절차, 파기방법에 따라 파기하여 안전성을 확보하고 있다.

체크리스트 나-3-1-E

AI 시스템의 설명가능성을 고려하고, 설명가능한 AI 기술 도입방안을 검토 또는 대안을 마련하였는가?

체크리스트

1) AI 시스템 개발 과정에서 설명가능성을 확보하기 위해 노력하였는가?
- AI 서비스 성격을 감안하여 설명 대상 및 설명하는 절차를 검토한다.
- 결과의 오해석 방지를 위해 설명 공유 대상과 범위를 설정한다.
- 설명가능한 AI 기술을 검토하고 설명방안을 마련하였는지를 확인한다.

YES ☐ | NO ☐ | N/A ☐

2) AI 시스템 개발 과정에서 설명가능한 AI 기술 적용이 어려운 경우, 대안을 마련하였는가?
- 설명가능한 AI 기술의 개발 트렌드를 충분히 확인하여 적용이 어려운 타당한 이유가 존재하며, 이에 대한 대안이 충분히 마련되었는지 확인한다.
- 업무 매뉴얼, 요구사항 문서 등에 AI 시스템이 생성한 결과를 설명하기 위한 기법 검토 여부를 확인한다.
- 대고객 서비스 후 사후적으로 해당문제 발생 여부 체크 점검 체계 마련 여부를 확인한다.

YES ☐ | NO ☐ | N/A ☐

3) AI 시스템 개발 과정에서 AI 시스템에 의한 결과를 고객에게 설명하기 위한 절차를 고려하였는가?
- 로보어드바이저를 활용한 투자일임계약 체결시 온라인 등을 통해 설명의무를 이행할 수 있도록 관련 절차가 마련되어 있는지, 고객에게 적절한 안내가 이루어지고 있는지 확인한다.
- 투자일임재산 운용과정에서 로보어드바이저를 활용시 위험요소, 투자전략, 시장 상황분석 등을 상세히 기재하여 고객에게 제공하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A증권사는 내부 업무 매뉴얼에 로보어드바이저 기획·설계 시 AI 모델의 설명 가능성을 확보토록 하고 있으며, 윤리적 검증을 위한 데이터를 별도로 마련하여 AI 모델로 인한 사회적 이슈 발생 가능성을 경감하고 있다.

- B증권사는 내부 업무 매뉴얼에 따라 로보어드바이저 시스템 활용 과정에서 발생할 수 있는 위험, 민감정보 활용 등에 관해 사용자에게 사전에 고지하고 있다.

- C운용사는 로보어드바이저를 투자일임재산 운용에 활용하는 경우 일반적인 투자위험(원금손실 가능여부, 과거 주가추이 및 산업동향 등), 조사분석자료 등 객관적인 근거에 기인한 회사의 실적 및 향후 주가전망, 투자를 선택한 이유 및 보유기간에 대한 전략 등을 투자일임보고서에 기재하여 고객에게 제공하고 있다.

체크리스트 다-1-1-E

AI를 활용한 시스템별 성능 지표 선정 및 목표 수준을 설정하고 충족여부를 확인하였는가?

체크리스트

1) AI 기반 모형의 성능을 평가하기 위한 지표를 선정하고 있는가?

- AI 모형 성능에 대한 유지 목표 및 평가지표 선정 여부를 확인한다.
- 내부 규정 및 업무 매뉴얼에 테스트 절차와 기준, 벤치마크 지수 선정 방법과 기준 등이 마련되어 있는지 확인한다.
- 성능의 이상 여부 판단 기준이 마련되어 있는지 확인한다.
- 로보어드바이저 AI 시스템 학습을 통해 생성된 모형에 대해 상품화 평가 기준을 제공하는지 확인한다.

예시 투자자 성향별 포트폴리오 산출역량, 다계좌 운용역량 등

YES ☐ | NO ☐ | N/A ☐

2) 선정한 성능 평가 지표에 따라 목표 수준의 달성 여부를 점검하고 미달하였을 경우 조치하고 있는가?

- 성능 평가 결과가 목표 수준에 미달하였을 경우 원인을 분석하여 모형 재개발 여부를 검토하여 조치한다.
- 특정 학습데이터에 과적합이 되지 않았는지 다양한 테스트를 수행하고, 테스트 데이터에 따른 변동성을 모니터링하고 있는지 확인한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A은행은 로보어드바이저 코드 및 알고리즘에 대한 품질 검증 방법 및 성능 측정 기준 마련 방식을 업무 매뉴얼에 반영하고 있으며, 매뉴얼에 따라 벤치마크 지수를 선정하고,

벤치마크 지수를 비교하여 성능의 이상유무를 판단하여 상품화 여부 등 코드와 알고리즘에 대한 품질을 측정 및 관리하고 있다.

- B사는 알고리즘 검증을 위해 학습 과정과 결과를 기록하고 관리하며, 코드와 알고리즘의 검증을 위해 이중으로 분석 결과를 검증하고 있으며, 로보어드바이저 대상이 되는 종목들의 동일 비중 포트폴리오를 기본적인 벤치마크 지수로 활용하며 해당 알고리즘과 유사한 형태의 지수가 있는 경우는 해당 지수를 벤치마크로 활용하고 있다. 예를 들어 ETF 자산 배분의 경우는 많이 사용되는 자산 배분 모형의 경우 (주식 00: 채권 00: …)을 벤치마크로 선정하고 있다. 벤치마크 지수와의 0년 이상 장기간 성과분석 항목(샤프 비율/ MDD/ 연환산수익률/ 회전율/ 변동성 등)의 비교 및 검증 과정을 통해 일정 성능을 충족하는 알고리즘을 선정하고 있다.

체크리스트 다-2-1-E

불합리한 차별이 나타나지 않도록 공정성 판단기준을 설정하고 충족여부를 평가하고 개선하였는가?

체크리스트

1) 공정성 목표 수준 및 공정성 판단 지표를 선정, 관리하는가?
   - 투자자 정보(예: 연령, 재산, 금융투자 목적 및 경험, 위험에 대한 태도 등)가 동일한 투자자 정보가 동일 포트폴리오를 생성하는지 확인한다.
   - 투자자 정보가 동일하나 성별, 인종, 민족, 국가 등 정보에 따라 차별적인 포트폴리오가 생성되지 않는지 확인한다.
   - 동일 시점, 동일 종목 매매처리 등 동일한 운용 조건을 가지고 있는 계좌들 사이 운용성과 괴리율을 측정, 기록하고 있는지 확인한다.
   - 다른 조건이 동일한 상황에서 성별, 인종, 민족, 국가 등 불합리한 차별 요소로 인해 운용성과 및 괴리율이 일정 이상(예: XX%) 차이나지 않도록 관리하고 있는지 확인한다.

   YES ☐ | NO ☐ | N/A ☐

2) 공정성 판단지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 기술적, 관리적 방안을 검토하여 조치하는가?
   - 편향방지 방법론을 활용하는 등 편향을 완화하기 위한 방안을 검토하였거나 조치하였는지 확인한다.

   YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A증권은 로보어드바이저 업무 전담팀을 구성하고 있으며, 관련 업무 매뉴얼에 따라 편향 가능성 최소화, 윤리기준 준수를 위한 모니터링 지표를 마련하여 주기적으로 평가하여 수준에 미달하는 경우를 관리·개선하고 있다.

- D자산운용은 로보어드바이저를 활용한 투자일임서비스를 제공하며 투자일임계약 체결 시 예시의 투자자 정보를 확인하고, 해당 정보가 동일한 경우 동일한 모형 포트폴리오를 산출하는지 확인. 성별, 지역, 국적 등의 정보에 따라 차별적인 모형 포트폴리오를 산출하는지 확인하고 있다.
  예시 연령, 재산상황, 금융투자 목적 및 경험, 금융투자상품에 대한 이해도, 투자 위험에 대한 태도

체크리스트 다-3-1-E

AI 시스템이 학습한 모형이 상황에 맞게 설명 가능한지 확인하고, 설명가능성을 적법한(또는 합리적인) 수준으로 개선하고자 노력하였는가?

체크리스트

1) 고객에 대한 설명의무, 금융서비스의 위험 수준 등을 고려하여 고객 및 이해관계자 등 설명 대상자들에 대해 AI 시스템 설명가능성 수준을 평가할 수 있는 기준 및 절차가 수립되어 있는가?
   - 내부 규정, 업무매뉴얼 등에 설명가능성 수준에 대한 평가 기준 및 절차가 수립되어 있는지 확인한다.

   예시 금융투자상품 위험도 분류표 등을 참고하여 작성된 로보어드바이저 서비스의 위험도 분류에 따른 초고위험, 고위험 여부에 따른 설명가능성 기준, 설명의 맥락(위치, 타이밍, 노출 등) 적절성 평가 기준

   YES ☐ | NO ☐ | N/A ☐

2) AI 시스템의 설명가능성이 기준이 적법한(또는 합리적인) 수준에 도달하지 못할 경우 개선 방안(재학습이나 알고리즘 수정 등)을 검토하여 조치하고 있는가?
   - 평가 기준에 도달하지 못했거나, 설명이 맥락에 적합하지 못한 경우 수행했던 개선 방안 검토 및 조치 결과를 확인한다.

   YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- A증권은 로보어드바이저 시스템 평가·검증 시 내부 업무 매뉴얼 마련된 절차에 따라 고객에 대한 설명의무가 있는 금융서비스 여부와, 고위험 서비스 여부를 다시 한번 확인하고 있으며, 해당될 경우 설명 가능성 확보 여부를 검증한다.

- B사는 AI 알고리즘을 활용하여 로보어드바이저 시스템 개발 시 판단 결과를 설명하기 위해, 투입값을 변화시키면서 결과값을 확인하고 투입 요인의 비중을 분석하는 민감도 분석 방법과 다수 요인을 투입 값으로 넣었을 때의 결과값을 보고 AI 동작 원리를 분석하는 첨가요인 민감도(SHAP) 알고리즘을 활용한다.

- C자산운용은 로보어드바이저 관련 업무 매뉴얼에 설명 필요성에 대해 규정하고 있으며, 업무 매뉴얼에 따라 설명이 필요한 위치, 타이밍과 설명이 필요한 위치별 맥락에 대해 검증하고 맞지 않는 경우 개선을 수행하고 있다.

체크리스트 라-1-1-E

AI시스템 제공에 있어 AI 이용여부, 설명·이의제기권 등 고객의 권리 및 이의신청·민원제기 방법 등 소비자로서의 권리구제 방안을 마련하여 고지하였는가?

체크리스트

1) AI 시스템 적용 사실과 범위에 대해 고객에게 안내가 이뤄지고 있는가?

- 상품설명서, 홈페이지, 모바일 앱 등의 채널을 통해 AI 시스템 활용 여부와 목적을 쉬운 용어를 사용하여 안내하고 있는지 확인한다.
- 고객의 권리에 대한 사전고지 여부 및 변경 이력을 확인한다.

YES □ | NO □ | N/A □

2) AI 시스템으로 인한 고객의 불편 및 불이익에 대한 구제 방안이 마련되어 있는가?

- 내부 규정 또는 업무 매뉴얼에 신고 및 이의제기 절차가 구체적으로 마련되어 있는지 확인한다.
- 업무분장표 등을 통해 전담 조직 및 인력을 확인한다.
- 홈페이지, 모바일 앱 등의 채널을 통해 고객에게 적절한 권리구제 절차 및 기준 등 사전고지 여부를 확인한다.
- 대고객 채널(홈페이지, 모바일앱 등)을 통한 전담 조직 또는 인력에 대한 고지 여부를 확인한다.

YES □ | NO □ | N/A □

〈 관련 법령에 따른 별도 체크리스트 〉

3) 로보어드바이저를 활용하여 집합투자재산을 운용하는 경우 집합투자기구의 투자목적 · 투자방침과 투자전략에 맞게 운용하고 있는가?

- 로보어드바이저의 활용이 집합투자규약등에 명기된 투자목적 · 투자방침과 투자전략 등에 부합하는지 주기적으로 점검한다.

YES ☐ | NO ☐ | N/A ☐

4) 로보어드바이저를 활용하여 투자일임업을 수행하는 경우 투자일임재산을 실제로 운용하는 로보어드바이저를 투자자의 동의를 얻지 아니하고 교체하였는가?

- 로보어드바이저 시스템을 교체하고자 하는 경우 반드시 투자자의 동의를 얻어야 한다.
- 다만, 기존 로보어드바이저와 동일성이 유지되는 범위 내에서 단순 수정, 개선하는 등의 경우는 제외한다.

YES ☐ | NO ☐ | N/A ☐

준수사례 (예시)

- C자산운용는 로보어드바이저와 관련한 고객의 권리를 홈페이지 및 모바일앱을 통해 사전 고지하고 있으며, 업무 매뉴얼에 따라 사전고지내용을 주기적으로 검토 후 개정하고 있다.

- B자산운용은 투자일임계약 체결 권유 및 계약 체결 시 고객이 소비자의 권리 등 계약 내용을 이해할 수 있도록 설명·고지하고 고객에게 계약서 및 설명서를 파일로 제공하고 있으며, 모바일 및 홈페이지를 통해 이용방법을 확인하고 이의신청·민원제기 등을 할 수 있는 시스템을 마련해 두고 있다.

- C사는 로보어드바이저 이용 고객을 보호하기 위해 내부 업무 매뉴얼에 따라 이의 제기 절차 및 이의 신청 및 민원제기에 대응하기 위한 전담 조직을 마련해두고 있다.

체크리스트 라-2-1-E

AI 시스템의 성능을 주기적으로 모니터링하며 성능 개선이 필요한지의 여부를 확인하고 있는가?

체크리스트

1) AI 시스템 성능이 안정적으로 유지되는지 확인할 수 있는 모니터링 절차를 마련하였는가?

- 성능 모니터링 주기, 범위 및 보고 등의 절차가 마련되어 있는지 확인한다.

- 성능 목표값에 대한 설정을 사전에 정의하고, 목표값 이하로 하락하는 경우, 파악할 수 있는 체계를 마련한다.
- 모니터링 결과 및 조치 이력의 문서화 여부 및 기록 내용을 확인한다.

YES □ | NO □ | N/A □

2) 유의미한 성능 하락 및 모집단 특성 변화에 따라 모형을 변경 또는 개선 필요시 의사결정 단계나 절차를 사전에 정의하고 있는가?

- 주기적인 성능평가와 재학습 필요성을 검토하고, 필요 시 모형을 교체하거나 개선을 진행해야 하며 모형 변경 또는 개선을 위한 의사결정 단계와 절차가 마련되어 있는지 확인한다.

YES □ | NO □ | N/A □

3) AI 재학습 및 모형 개선 절차를 수립·이행하고 있는가?

- AI 재학습 및 모형 개선 절차 수립과 이행 여부를 확인한다.
- 개선 방안 수립 시 데이터의 최신성, 정확성, 정합성, 비편향성을 고려하였는지 확인한다.
- 학습데이터 변경 이력 기록 여부 및 관리 문서를 확인한다.
- 활용된 데이터의 보관기간 준수 여부(5년, 10년 등)를 확인한다.

YES □ | NO □ | N/A □

※ 로보어드바이저 서비스가 특정 고객, 또는 특정상황에서 다른 투자자문을 제공하지 않도록 결과에 대한 충분한 모니터링 및 검토가 필요하다.

〈 관련 법령에 따른 별도 체크리스트 〉

4) AI 시스템을 유지 · 보수하기 위하여 요건을 갖춘 전문인력을 1인 이상 배치하고 있는가?

- 전문인력은 다음 각목의 어느 하나에 해당하여야 한다.
  - 가. 정보보호 또는 정보기술(IT)분야의 전문학사학위를 취득한 후 3년 이상 「전자금융거래법」에 따른 정보보호 또는 정보기술(IT) 분야 업무를 수행한 경력이 있는자
  - 나. 정보보호 또는 정보기술(IT)분야의 학사학위를 취득한 후 1년 이상 정보보호 또는 정보기술(IT) 분야 업무를 수행한 경력이 있는자
  - 다. 정보보호 또는 정보기술(IT) 분야의 석사학위 이상 취득한자

라. 7년 이상 정보보호 또는 정보기술(IT) 분야 업무를 수행한 경력이 있는 자
마. 전문학사학위를 취득한 후 5년 이상 정보보호 또는 정보기술(IT) 분야 업무를 수행한 경력이 있는 자
바. 학사학위를 취득한 후 3년 이상 정보보호 또는 정보기술(IT) 분야 업무를 수행한 경력이 있는자
사. 석사학위를 취득한 후 1년 이상 정보보호 또는 정보기술(IT) 분야 업무를 수행한 경력이 있는 자
아. 「전자금융거래법 시행령」 별표 1 제1호나목 각호의 어느하나에 해당하는 전문자격을 취득한 후 1년 이상 정보보호 또는 정보기술(IT)분야 업무를 수행한 경력이 있는 자

YES □ | NO □ | N/A □

5) 로보어드바이저를 활용하여 투자자문/일임업을 수행하는 경우 투자자의 투자목적, 재산상황, 투자경험 등을 고려하여 투자자의 투자성향을 분석하고 있는가?

- 투자자문의 내용 또는 투자일임재산에 포함된 투자대상자산이 하나의 종류 · 종목에 집중되지 아니하도록 한다.
- 매 분기별로 1회 이상 다음의 사항을 평가하여 투자자문의 내용 또는 투자일임재산의 운용방법의 변경이 필요하다고 인정되는 경우 그 투자자문의 내용 또는 투자일임재산의 운용방법을 변경한다.
  가. 투자자문 내용 또는 투자일임재산의 안전성 및 수익성
  나. 투자자의 투자성향 분석을 고려하여 투자자문의 내용 또는 투자일임재산에 포함된 투자대상자산의 종목 · 수량 등이 적합한지 여부

YES □ | NO □ | N/A □

준수사례 (예시)

- A증권사는 로보어드바이저 시스템에 대한 성능 평가 기준(실제 수익률과의 일치성, 고객 만족도 등)을 업무 매뉴얼에 명시하고 있으며, 관련 업무 절차에 따라 성능 모니터링 및 평가 결과에 따른 재학습을 주기적(분기당 1회)으로 이행하고 있다.

- B자문사는 투자자문의 내용이 하나의 종목에 집중되지 않도록 로보어드바이저 시스템을 설계하였으며 매분기별로 자문한 내용을 점검하여 투자자의 성향에 적합한지를 파악하고 있다.

- C운용사는 AI 시스템을 유지, 보수하기 위하여 정보보호 업무에 7년이상 경력이 있는 자를 채용하여 배치하고 있다.

체크리스트 라-3-1-E

AI 시스템 및 AI 시스템에서 사용하는 데이터의 오용·악용 가능성을 최소화하였는가?

체크리스트

1) 보안 시스템 구축 등을 통해 AI 시스템에 대한 보안대책을 수립하여 적대적 공격 등 오용·악용 가능성을 최소화하고 있는가?

- AI 로보어드바이저 시스템 구축 시 시스템, 인프라, 데이터, 네크워크, 이용자 보호 등과 관련하여 다양한 보안 위협 및 대응조치를 포함한 적합한 보안대책을 수립하고 검토하고 있는지 확인한다.
- 부적절한 투자자문 등 비정상 동작이나 예기치 못한 오류에 대한 대책을 수립하고, AI 로보어드바이저 시스템에 대한 접근 권한을 관리하고 있는지 확인한다.
- 사용자 자산운용 이력 등에 대한 개인정보 암호화 및 권한별 접근 통제 조치 여부를 확인한다.
- AI 로보어드바이저 시스템을 운용하는 중요 서버에 백신 프로그램을 설치하고 주기적 업데이트 및 악성코드 점검, 실시간 검사 설정을 하고 있는지 확인한다.
- 보안 취약점에 대한 사전·정기 점검 수행 및 취약점 조치 여부를 확인한다.
- AI 로보어드바이저 시스템 관련 침해사고 분석 시 필요한 로그에 대하여 보존 및 검토에 대한 정책을 수립하였는지 확인한다.
- AI 로보어드바이저 시스템의 성능을 주기적으로 점검하여 성능 저하를 유발하는 공격 여부를 확인한다.
- AI 로보어드바이저 시스템 관련 로그별 보존기간 및 검토 주기를 지정하였는지 확인한다.
- AI 로보어드바이저 시스템 테스트 시 개인신용정보가 아닌 임의의 데이터를 생성하여 테스트를 수행하는지 확인한다. 단, 개인신용정보 이용이 불가피한 경우, 책임자의 승인 절차에 따라 가공하여 사용하고 즉시 폐기하는 등 관리대책이 수립·이행되었는지 확인한다.
- 암호화 대상, 사용 암호 알고리즘, 암호키 관리 방안을 포함한 개인정보 암호화 정책을 수립하였는지 확인한다.
- 정보자산 등 운영체제, 소프트웨어 패치 관리정책 및 절차를 수립·이행하고 인터넷 직접 접속을 통한 패치를 제한하고 있는지 확인한다.

- AI 로보어드바이저 시스템 개발 과정에서 적대적 공격 등 오용·악용 가능성을 최소화하기 위해 보안성 검증을 실시하고 있는지 확인한다.

YES □ | NO □ | N/A □

2) 보안 시스템 구축 등을 통해 고객 또는 제3자에 의한 데이터 오염 공격 등을 통한 데이터의 오용·악용 가능성을 최소화하고 있는가?

- AI 로보어드바이저 시스템에 사용되는 데이터 오용 · 악용 여부 감지 지표 설정 및 주기적 확인 여부 등 절차와 대처방안 수립 · 운영 여부를 확인한다.
- 잘못된 개인신용정보 주입·조작 등 학습데이터 변조 여부를 주기적으로 확인한다.
- AI 챗봇시스템에 사용되는 데이터의 무결성, 기밀성을 유지하기 위한 관리 규정을 준수 여부를 확인한다.

YES □ | NO □ | N/A □

3) 오픈소스(프레임워크, 라이브러리 등) 보안 취약성 관리를 위한 체계를 수립하여 AI 시스템의 보안성을 강화하고 있는가?

- AI 로보어드바이저 시스템 개발에 필요한 오픈소스 사용 시 라이브러리 보안 취약점 확인 및 통지체계 수립 여부를 확인한다.
- 오픈소스 취약점을 주기적으로 확인하여 업데이트 및 패치를 적용하고 있는지 확인한다.
- 로보어드바이저 관련 AI 알고리즘의 취약점 점검 여부를 확인한다.

YES □ | NO □ | N/A □

4) 침해사고 및 재해 등을 예방하기 위한 체계 및 침해사고 또는 재해가 발생했을 때 피해 확산 · 재발 방지와 신속한 복구를 위한 체계를 갖추고 있는가?

- 해킹, 악성코드, 논리폭탄, 메일폭탄, 서비스거부 또는 고출력 전자기파 등의 방법으로 사고 발생 시에 대비한 예방·복구 체계 마련 여부를 확인한다.
- 정상적 보호 · 인증절차를 우회하여 정보통신망에 접근할 수 있도록 하는 프로그램이나 기술적 장치 등에 대한 차단 체계 마련 여부를 확인한다.
- 침해사고 발생 시 기록 및 보고, 신고 및 통지, 비상 연락체계 등 내용이 포함된 침해사고 대응 절차를 수립하였는지 확인한다.

YES □ | NO □ | N/A □

※ 로보어드바이저 서비스는 최적의 투자자문을 제공하는 서비스로 오류가 생길 시 고객에게 실시간으로 손해를 발생하게 하거나 그 손해가 매우 커질 수 있다. 그러므로 주기적으로 오류를 점검하고 취약점을 보완하기 위한 조치를 취하여야 한다.

준수사례 (예시)

- A은행은 로보어드바이저 시스템과 관련하여 침입차단시스템, 침입탐지시스템, 악성코드방지프로그램 등의 안전 조치를 적용하고 있으며, 학습 데이터의 오염을 방지하기 위해 출처(데이터 공급사)가 검증된 데이터만 활용하고 있다.

- B자산운용은 해킹, 컴퓨터 바이러스 등의 침해에 대한 예방체계 대응 체계를 갖추고 있으며 시스템을 유지보수하기 위하여 전문인력을 상시 배치하고 있다.

금융분야 AI
개발·활용 안내서