금융분야 AI 운영 가이드라인

1. 목적과 적용 범위

가. 가이드라인은 금융분야에서의 인공지능(이하 'AI'라 한다.) 시스템의 개발, 사업화 및 활용과 관련한 기획·설계, 평가·검증, 도입·운영 및 모니터링의 전 과정에서 신뢰성을 제고하여 AI 활성화를 제고하고 금융서비스에 대한 고객신뢰를 확보하는데 기여하는 것을 목적으로 한다.

나. 가이드라인은 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 제외한다.)하거나 활용하고자 하는 금융회사, 상품추천·신용평가 등 금융연관 서비스 제공을 위한 업무에 AI시스템을 직·간접적으로 활용하거나 활용하고자 하는 비금융회사(이하 '금융회사 등'이라 한다.) 등에 적용한다.

다. AI 시스템이란 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 시스템을 의미한다.

라. 금융회사 등은 AI 시스템이 활용된 서비스의 특성 및 고객 특성, AI 시스템이 활용된 서비스의 고객 수 등을 종합적으로 고려하여 AI 활성화 및 금융서비스에 대한 고객신뢰 확보라는 가이드라인의 취지를 훼손하지 않는 범위 내에서 가이드라인의 적용 범위 등을 조정할 수 있다.

2. 거버넌스의 구축

가. 금융회사 등은 조직이 추구하는 가치와 주된 AI 활용 맥락 등을 고려하여 AI 활용에 관한 윤리원칙과 기준을 수립한다.

나. 금융회사 등은 AI 시스템의 잠재적 위험을 평가하고 이를 관리하기 위하여 구성원의 역할·책임·권한 등을 AI 시스템의 전 과정에 걸쳐 구체적으로 정의한다. 금융회사 등은 AI 윤리원칙과 기준에 맞는 조직 관리를 위하여 AI 윤리위원회를 별도로 설치할 수 있다.

다. 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 마련한다. 위험관리정책은 소비자 권리보장을 위한 시스템 운영, AI 모델 및 학습데이터의 관리, AI 시스템 관련 문제 발생 시 감독당국과의 소통, 회사 내 AI 책임 문화 확산의 촉진 등의 업무처리에 관한 내용을 포함한다.

라. 금융회사 등이 개인에 대한 부당한 차별 등 개인의 권익와 안전, 자유에 대한 중대한 위험을 초래할 수 있는 서비스(이하 '고위험 서비스'라 한다.)에 대해 AI 시스템을 활용하는 경우, 적절한 내부통제 활동 및 승인절차를 마련하고, 승인 책임자를 지정한다, 승인책임자는 책임있는 업무 수행이 가능한 지위로 하되 최고위험관리책임자, 신용정보보호·관리인, 최고정보보호책임자 등 유사업무와 겸직할 수 있다.

3. AI 시스템의 기획 및 설계 단계

가. 금융회사 등은 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.

나. 금융회사 등은 AI 시스템의 목적 및 특성, 고객의 특성 등을 고려하여 탄력적으로 AI 시스템을 활용할 수 있다. 다만, AI 시스템이 인간의 의사결정을 전면적으로 대체하거나, 중요 의사결정을 대체하는 경우 금융회사 등은 AI 시스템을 효과적으로 감독·통제하고 책임성을 유지할 수 있도록 AI 시스템을 설계한다.

4. AI 시스템의 개발 단계

가. 금융회사 등은 올바른 AI 학습을 위하여 데이터의 출처, 품질, 편향성 등을 조사·검증하고 주기적인 데이터 갱신 등 데이터 품질 개선을 위한 방법을 검토한다.

나. 금융회사 등은 AI 시스템이 「개인정보 보호법」 제23조제1항 및 시행령 제18조에 따른 민감정보, 또는 이와 유사한 사생활 관련 정보 등을 활용하는 경우 사전 동의 획득 또는 비식별조치 등 안전한 정보 활용을 위한 충분한 조치를 거쳐야 하며, 해당 정보 활용의 필요성을 평가하고, 데이터 처리 과정에서 해당 정보의 재식별, 유출, 악용가능성이 없도록 한다.

다. 금융회사 등은 관련 법령 등에 따라 고객에 대한 설명의무가 있는 금융서비스 등에 AI 시스템을 활용하는 경우 또는 고위험 서비스에 AI 시스템을 활용하는 경우에는 개발 단계에서부터 설명 가능성을 고려하고, 가용한 설명 가능한 인공지능 기술 등을 확인하여 이를 도입하기 위한 노력을 기울인다.

5. AI 시스템의 평가 및 검증 단계

가. 금융회사 등은 AI 윤리원칙, AI 시스템의 목적, 오류 사례에 따른 고객 영향 및 잠재적 피해의 정도, AI 성능 측정 지표의 상충관계 등을 종합적으로 고려하여 AI 시스템의 적절한 성능 목표 수준 및 성능 측정 지표를 선정·관리한다.

나. 금융회사 등은 AI 윤리원칙, AI 시스템의 목적, 공정성 평가 지표별 고객 영향 및 잠재적 피해의 정도, AI 공정성 평가 지표의 상충관계 등을 종합적으로 고려하여 AI 시스템의 적절한 공정성 목표 수준 및 공정성 판단 지표를 선정·관리한다. 선정된 공정성 판단 지표에 따라 불균형이 발견된 경우, 공정성을 개선시킬 수 있는 기술적·관리적 노력을 기울인다.

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

6. AI시스템의 도입, 운영 및 모니터링 단계

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

나. 금융회사 등은 도입된 AI 시스템의 성능을 주기적으로 모니터링하고, 데이터 재학습 필요성 검토 등 성능 개선 가능성을 확인한다.

다. 금융회사 등은 AI 시스템에 고객 또는 제3자에 의한 데이터 오염 공격, 적대적 공격 등 오용·악용 가능성이 있는지 여부를 확인하고, 가용한 기술 범위 내에서 오용·악용 사례를 최소화할 수 있는 방안을 도입하기 위해 노력한다. 금융회사 등은 오픈소스 기반 AI 개발 프레임워크 등 AI 개발 환경의 보안 취약성에 관해 상시적으로 통지를 받을 수 있는 절차를 반영하고, 최선의 보안 시스템을 구축하기 위하여 노력한다.

7. AI 시스템 업무위탁에 관한 특례

가. 금융회사 등은 AI 시스템의 개발·운영 등을 외부기관에 위탁하고자 할 경우, 수탁기관이 동 가이드라인 및 가이드라인에 기초하여 마련된 AI 윤리원칙 및 위험관리정책을 준수하여 AI 시스템을 개발·운영할 수 있도록 하기 위한 위험관리지침을 마련하고 금융회사가 직접 AI 시스템을 개발·운영하는 경우에 비해 AI 시스템 운영에 따르는 위험이 확대되지 않도록 한다.

나. 금융회사 등은 외부기관에 의한 AI 시스템 개발·운영이 위험관리지침에 따라 이루어졌는지 주기적인 보고·점검 체계를 구축·운영하고, 고위험서비스에 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.

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