금융에서의 생성형 인공지능 활용 현황과 법적 쟁점에 대한 연구*

이 석 준
(대법원 재판연구관)

【초 록】

GPT-3 출시 이래로 생성형 인공지능(이하 'AI')은 전 세계에서 여러 산업 분야에 걸쳐 중요한 기능을 하고 있고 미래의 시장 규모도 빠르게 성장할 것으로 예측된다. 생성형 AI는 금융 분야에서 고객 상담, 고객 투자 관리, 대출 등 위험 관리, 금융회사의 업무 개선에 많은 기여를 하고 있다.

생성형 AI에는 생성적 적대 신경망(GAN), 트랜스포머 모델, 다중모드 AI 등이 있다. 위 AI들은 각각의 장·단점을 가지고 있으면서 합성 데이터의 생성, 사기성 거래 등의 구별, 비대면뱅킹, 금융흐름의 예측, 감정 분석, 재무 분석 등에 있어서 탁월한 성능을 발휘한다.

그럼에도 금융 분야에서 생성형 AI를 활용함에 있어 몇 가지 법적 문제들이 있을 것으로 예상된다. 여기에는 조작 위험(딥페이크), 금융 안정성의 위험(플래시크래쉬, 금융시스템 리스크), 투자자보호 관련 위험, 데이터 유출 위험, 불공정의 위험(데이터 및 알고리즘 편향, 설명가능한 인공지능, 자동화된 의사결정) 등이 있어 이에 초점을 두고 그 쟁점 및 법적·기술적 해결 방안에 대하여 살펴보았다. 나아가 고객 상담시 투명성 위험, 반경쟁 위험에 관하여도 짚어보았다.

* 이 글은 필자의 개인적인 견해임을 밝힙니다.

주제어: 금융 / 생성형 인공지능 / 딥페이크 / 자동화된 의사결정 / 편향 / 설명가능한 인공지능 / 데이터 유출 / 챗봇 / 로보어드바이저 / 생성적 적대 신경망 / 변형 자동 인코더 / 트랜스포머 모델 / 다중모드 인공지능

논문투고일: 2024. 3. 19 / 논문심사일: 2024. 4. 17 / 게재확정일: 2024. 4. 24.

【차 례】

Ⅰ. 서 론

마켓리서치는 전 세계 금융서비스 산업에서 생성형 인공지능(Artificial Intelligence, 이하 'AI')의 시장규모가 2022년 약 8억 달러에서 2032년 약 94억 달러로 성장하여 연 평균 28.1%의 성장률을 보일 것으로 예상하였다.[1] 생성형 AI는 대량의 데이터 세트[2]를 빠르게 분석하고 인간이 놓칠 수 있는 복잡한 패턴을 식별하는 데 탁월하므로 금융 분야의 사기 탐지, 보안 강화, 위험 평가, 투자 예측 등 다양한 영역에 활용될 것으로 보인다.[3] 또한 생성형 AI에 대한 한 연구에 따르면 생성형 AI를 활용하는 근로자는 그렇지 않은 자들에 비하여 생산성을 평균 14% 정도 향상시킬 수 있다고 한다.[4] 실제로 챗봇과 같은 생성형 AI를 통하여 고객 추천용 상품정보를 생성하고 고객의 문의에 응답하며 투자설명서나 재무보고서를 신속하게 작성할 수 있는

1) MarketResearc report, July 2023, "Global Generative AI In Financial Services Market By Type (Solutions And Services), By Application (Credit Scoring, Fraud Detection, And Other), By Deployment Mode (Cloud And On-Premises), By Region And Companies - Industry Segment Outlook, Market Assessment, Competition Scenario, Trends, And Forecast 2023-2032"〈https://marketresearch.biz/report/generative-ai-in-financial-services-market/〉.

2) 여러 정보를 담고 있는 데이터를 추상적인 하나의 정보 단위로 보고 그 정보 단위인 데이터를 여러 개 모아놓은 집합을 말한다. 예를 들어, 금융기관 고객의 데이터 세트는 수많은 고객들의 주소, 연락처, 계좌번호, 입출금내역 등의 개별적인 정보들을 모아놓은 집합이라고 할 수 있다.

3) *Id.*

4) Erik Brynjolfsson, Danielle Li & Lindsey R. Raymond, "Generative AI at work", National bureau of economic research(2023), p.1.

등 생산성 향상 및 비용절감의 효과가 클 것으로 예상되고 있다.[5]

이에 따라 우리나라를 비롯한 전 세계의 금융업계에서는 생성형 AI의 도입을 위한 연구를 계속하고 있다. 반면 생성형 AI를 포함한 AI에 대한 법적 규율이 우리나라에서 전무한 상황이다. 따라서 금융에서 생성형 AI의 사용 현황을 일별하고 예상되는 법률적 쟁점에 대한 검토가 필요해 보인다. 참고로 생성형 AI는 기존의 예측형 AI와 결합하여 기능을 수행하거나 예측형 AI의 능력과 기능이 강화된 용도로 쓰이는 경우도 많아서 예측형 AI에 대한 일부 법적 쟁점이 생성형 AI에도 연결되므로 이를 포함하여 살펴보기로 한다.

II. 생성형 AI의 의의와 현황

1. 생성형 AI의 의의

생성형 AI는 현재 그 개념이 확고하게 정립된 것은 아니지만, 딥러닝(Deep Learning)[6] 모델의 일종으로 미리 학습된 데이터를 기반으로 문자, 음성, 이미지, 동영상 등 새로운 내용을 생성할 수 있는 AI 모델로 정의내릴 수 있다.[7] 생성형 AI는 1986년과 1997년에 각각 개발된 역전파(Back-propagating)[8]·순환신경망(Recurrent Neural Network: RNN)과 장·단기 기억(Long Short-Term Memory: LSTM) 신경망[9] 등을

5) American Banker August 25th, 2023 article, "How banks can adopt generative AI".

6) 인간처럼 사고하고 학습하도록 훈련된 프로그램을 말한다. '심층 신경망(Deep Neural Network)'이라고도 한다.

7) OECD, "Generative artificial intelligence in finance"(2023), p.8.

8) 역전파는 다층 신경망(입력층, 출력층, 은닉층과 같이 여러 계층으로 구성된 신경망을 말한다) 학습에서 출력 결과와 실제 결과 사이의 오차를 조정하기 위하여 가중치를 수정하는 과정에서 순전파와 달리 출력층부터 입력층으로 계산을 진행하는 것이다. 한편 순환신경망은 시간적 순서에 따라 특정 시점의 데이터가 입력되면 기억장치가 그 출력 결과를 스스로 저장하고 그 이후 그 결과를 다음 시점의 데이터로 순차적으로 입력하는 인공 신경망으로 작업이 반복될수록 정확성이 향상된다. David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, "Learning representations by back-propagating errors", *Nature*, Vol. 323(1986)에서 발표되었다.

9) 순환신경망의 일종으로 시간이 오래 지나간 경우 순환신경망의 데이터 처리 과정에서 기울기 소실(Vanishing Gradient) 문제가 발생하여 초기 학습된 데이터가 학습되지 않는 문제가 발생하자 이를 해결하기 위하여 만들어졌다. 통상의 순환신경망과의 차이는 데이터셀에서 정보를 추가하거나 제거할 수 있어 선택적 기억 및 망각이 가능하다는 점에 있다. Sepp Hochreiter, Jürgen Schmidhuber,

토대로 발전한 것으로,[10] 지도학습 방식(Supervised Learning)[11]과 비지도학습 방식(Unsupervised Learning)[12] 등을 이용하여 데이터 세트를 학습한다. 생성형 AI의 기능은 학습하는 데이터 세트의 내용에 따라 달라진다. 데이터의 종류는 문자·음성·이미지·동영상·코드 등 다양한데, 한 종류의 데이터 입력만 받는다면 단일모드(Unimodal) AI라고 하고, 두 종류 이상의 데이터 입력을 받는다면 다중모드(Multimodal) AI라고 한다.

생성형 AI는 기존의 예측형 AI와 구별하여 설명하는 경우가 많다. 예측형 AI는 입력된 데이터를 바탕으로 패턴을 찾아내고 특정 결과나 행위를 예측하는 AI 모델이다. 예측형 AI는 예를 들어 금융 분야에서 국제 시장의 동향, 국가기관의 환율 조정, 경제 상황 등을 종합하여 증권시장의 흐름을 예측하는 데 사용할 수 있다. 반면 생성형 AI는 학습된 데이터를 기반으로 새로운 콘텐츠를 생성하거나 문제를 해결하는 것으로 예측형 AI보다 한 단계 발전된 형태로 볼 수 있다. 생성형 AI는 다양한 데이터를 새로운 방식으로 분석할 수 있고 비정형화된 새로운 패턴을 도출할 수 있다. 이에 따라 증권시장의 흐름을 예측하는 것을 넘어서서 적절한 투자방법과 투자시점, 대상 등을 알리는 투자보고서를 작성하거나 인간이 인식할 수 없는 숨겨진 패턴과 추세를 발견하여 새로운 투자기법을 개발할 수 있다. 다만, 생성형 AI가 예측형 AI와 대립되는 개념으로 볼 수 없는 것이 예측형 AI의 데이터를 이용하여 새로운 콘텐츠를 생성하거나 생성형 AI의 시스템을 이용하여 기존 예측형 AI 시스템의 효율성과 정확성을 향상시킬 수 있기 때문이다. 결국 위 AI들은 상호보완적 관계라고 할 수 있다.

그리고 일반적인 AI는 금융, 법률, 의료, 문화예술, 보안 등 특정 용도로 활용되기 위하여 마련된 경우가 많으나, 생성형 AI는 특정 용도보다는 다양한 용도로 사용될 수 있고 특정 용도의 AI 시스템을 개발할 수 있는 기초가 되는 기반 AI가 되기도 한다. 마지막으로 생성형 AI가 자연어 처리(Natural Language Processing: NLP) AI와 결합한다면 인간과 유사한 문자를 생성하는 대규모 언어 모델(Large Language

---

"Long Short-Term Memory", *Neural Computation*, Vol. 9(1997)에서 발표되었다.

10) 그 후에 발전한 것으로 합성곱신경망(Convolutional Neural Network)과 뒤에서 나오는 자기 주의 기법(Self attention)도 중요한 역할을 한다.

11) 입력값과 짝을 이루는 출력값을 동시에 제공하는 것으로, 정답지를 알려주는 것에 비유하여 지도학습 방식이라고 한다.

12) 출력값 없이 입력값만 제공되는 것으로, 문제(입력)에 대한 정답(출력)을 알려주지 않은 것에 비유하여 비지도학습 방식이라고 한다.

Model)이 된다.

2. 금융 분야에서의 사용 현황

(1) 고객 편의 극대화

최초의 생성형 AI는 1966년 고객 상담을 위한 챗봇으로 만들어졌다. 그만큼 AI 상담원은 오래전부터 활용되어 왔고 개선을 거듭하였다. 현재 생성형 AI의 챗봇 수준은 고객들이 자신의 대화 상대가 인간인지, 기계인지 구별하기 어려운 상태에 도달해 있다. 이와 같이 AI를 기반으로 하는 챗봇은 고객 문의에 빠르게 응답하여 고객의 대기시간을 줄이고 고객 경험을 전반적으로 향상시켜 로보어드바이저를 구현할 수 있다.[13] 그리고 기존의 예측형 AI는 알고리즘에 따라 고객이 요구한 질문에 정해진 답변만 하였다면, 생성형 AI는 이를 넘어서 고객에게 맞춤화된 답변을 새로 생성하고 지속적으로 기능을 향상시킬 수 있는 장점이 있다. 생성형 AI를 기반으로 한 상담 시스템으로 챗봇 외에도 음성 봇, 가상 비서 등이 있다.[14][15]

생성형 AI는 최근에 성장하는 임베디드 금융(Embedded Finance)에 도움이 될 수 있다. 임베디드 금융은 은행이나 증권회사가 아닌 비금융회사가 자신의 플랫폼 내에 금융서비스를 판매할 수 있도록 하는 핀테크 시스템으로 별도의 은행 어플리케이션 없이 입출금 계좌서비스를 이용하거나 △△페이 등의 결제 서비스, 대출, 보험 등을 제공할 수 있다. 생성형 AI를 활용하면 대용량의 고객 데이터를 실시간으로 분석하고 고객의 선호도, 행동 패턴 등을 추출하여 임베디드 금융에 구현함으로써 고객 편의에 이바지할 수 있다.

(2) 고객 투자 등 관리

금융투자업자는 전문적인 금융지식과 경험을 토대로 고객의 투자금을 적정하

---

13) Joerg Osterrieder, "A Primer on Artificial Intelligence and Machine Learning for the Financial Services Industry"(2023), p.4.

14) 예를 들어, JPMorgan은 개인 투자자의 포트폴리오를 분석하고 투자 대상을 선택하는 AI 자문인 "IndexGPT"를 개발할 것을 공개적으로 발표하였다.

15) 국내 시중은행들도 대부분 AI챗봇을 운용하고 있고 이에 연계하여 모바일 어플리케이션 내에 'AI 뱅커'를 운용하고 있거나 도입할 예정인데, AI 뱅커는 말하는 입 모양이나 표정, 말투 등이 실제 은행원이 설명해 주는 듯한 느낌을 준다.

게 관리하고 투자한다.[16] 여기에는 고객 포트폴리오 제시가 포함되고 있고, 포트폴리오는 고객의 개별적인 투자 목표, 시장 상황 및 위험 성향에 따라 전략적으로 구성된다. 생성형 AI의 발전으로 금융투자업자는 고객에게 맞는 포트폴리오를 생성하여 고객에게 투자 및 자산관리에 대한 아이디어를 제시할 수 있다.[17] 또한 고객을 직접 만나기 어려운 상황에서도 비대면으로도 고객과 투자전략을 논의하거나 고객에 대한 금융업무를 수행하기 용이해졌다.

그리고 생성형 AI는 고객 데이터에서 복잡한 패턴을 식별하여 고객의 행동을 분석하는 데 사용할 수 있다. 예를 들어 어떤 고객이 장내파생상품이나 ETF 상품을 매수할 여지가 있는지, ETF를 매수한다면 어떠한 내용으로 투자처를 구성할 것인지, 가상자산 선물과 같은 위험자산 투자를 선호하는지 등의 고객 선호도를 예측 및 분석할 수 있다. 또한 투자중개업자나 투자자문업자 등이 생성형 AI의 보조를 받아 빠르고 효율적으로 고객에게 맞춤식의 투자추천을 할 수 있다.

(3) 위험 관리

생성형 AI는 대량의 데이터를 분석하여 사기 또는 조작, 자금세탁 등의 이상거래를 포착할 수 있다.[18] 착오주문과 착오결제 방지에 있어서도 중요한 역할을 할 수 있다. 나아가 시스템적 위험 또한 인지할 가능성도 있다. 이러한 역할은 생성형 AI가 인간이 놓칠 수 있는 패턴을 식별하기 때문에 더욱 가능한 일이다. 특히 군집(Clustering) 알고리즘을 병용하는 경우 일정한 범주 내에 있는 정상 거래를 집단화하고 그 범주를 넘어선 나머지 거래를 이상거래로 간주할 수 있다. 참고로 기존 AI로도 사기, 조작, 자금세탁, 착오주문 등을 탐지할 수는 있으나 생성형 AI는 기존 AI에 비하여 방대한 양의 데이터를 빠른 시간 내에 분석하므로 매우 복잡한 패턴으로 이루어진 이상거래 등을 더 신속하고 용이하게 식별할 수 있다.

생성형 AI는 신용 평가 및 위험 감소에도 긍정적인 영향을 미칠 수 있다. 은행의 대출 분야에서 특히 쓰임새가 많은데 대출 분야에서 중요한 것은 대출 심사와

16) AI 플랫폼인 K Score는 생성형 AI와 자연어 처리기술을 이용하여 대용량의 금융뉴스 등의 데이터를 처리하여 투자자들의 투자판단에 도움을 준다.

17) 모건스탠리의 Next Best Action.

18) 가상자산 결제서비스 업체인 Sardine은 법정화폐 또는 가상자산 거래에서의 사기 방지 목적으로 미국의 온라인 대출 서비스 업체인 Cross River 은행의 AI 결제 포털을 사용할 예정이다. American Express와 JPMorgan Chase의 경우에도 사기 탐지를 위하여 생성형 AI를 활용한다.

신용점수 평가 및 적절한 신용한도의 설정일 것이다. 대출 심사에 있어서는 특정 고객의 대출 승인 여부, 승인시 금리의 결정 등의 의사결정을 자동화 또는 반자동화시킬 수 있어 활용도가 높다.

특히 생성형 AI는 고객에 대한 신용 보고서, 계좌 잔액, 고용 이력, 교육 수준, 대출 이력 등 상당한 양의 데이터를 분석하여 고객의 신용도를 판단하여 고객이 대출채무를 불이행할 가능성이 있는지 체크하고 대출한도액과 금리를 정할 수 있다. 나아가 생성형 AI는 다양한 분야에서 비정형적인 데이터를 수집하고 분석할 수 있어 대출신청인의 자력과 신용위험에 대한 통찰력 있는 정보를 제공해 줄 수 있다. 예를 들어 생성형 AI는 개인의 소셜 미디어 활동을 분석하여 소비 패턴, 생활 습관 등 신용도에 영향을 미칠 수 있는 잠재적인 개인 생활도 식별할 수 있다.[19]

이와 같은 양상은 신용카드 발급 및 갱신, 보험계약의 체결에 있어서도 마찬가지로 신용카드 채무불이행 가능성이나 보험사기 발생 가능성을 탐지하고 분석하는 데 생성형 AI를 활용할 수 있을 것으로 보인다.

(4) 업무 개선

생성형 AI는 다음과 같은 이유들로 금융기관들의 업무 개선에 도움을 줄 수 있다. 먼저 생성형 AI의 보조를 받아 재무분석 등에 있어 새로운 통찰력을 제공할 수 있다.[20] AI는 특정 금융상품의 가격과 거래량, 뉴스 기사, 투자자들의 동향 등 방대한 양의 데이터 세트를 학습하여 시장을 분석하고 투자 기회를 인지하며 특히 Arima[21]와 같은 시계열 알고리즘을 병용하면 미래 시장 동향을 예측할 수 있으므로 금융투자업자가 투자결정을 내리는 데 도움이 될 수 있다.[22]

둘째, 생성형 AI는 사람이 볼 때 몇 시간이 걸리는 기초서류를 수초만에 내용을 정확하게 읽고 적절하고 형식에 맞는 보고서나 재무제표를 작성하게 해준다. 또한 한눈에 알아보기 쉽게 엑셀파일을 만들거나 통계 등의 자료를 생성할 수 있다.

---

19) Siti Aishah Binti Mohd Yusof, Fatin Aqilah Binti Mohamad Roslan, "The Impact of Generative AI in Enhancing Credit Risk Modeling and Decision-Making in Banking Institutions", *Emerging Trends in Machine Intelligence and Big Data*, Vol. 15(2023), pp.43-44.

20) 도이치방크의 사례.

21) Autoregressive Integrated Moving Average, 자기회귀(AR, 과거 시점 값이 현재 시점의 값에 영향을 미치는 것으로 과거 시점 값들의 가중치 합)와 이동평균(MA, 과거 시점의 오차가 현재 시점이 값에 영향을 미치는 것)을 결합한 시계열 예측 모델이다.

22) Joerg Osterrieder, *supra* note 13, p.4.

해외의 유력 금융 관련 자료도 어법에 맞게 번역을 해준다. 생성형 AI는 집합투자업자나 투자매매·중개업자의 투자설명서나 보조자료, 보험회사의 보험상품 설명서를 적은 비용으로 신속·정확하게 생성할 수도 있다. 이와 같이 생성형 AI는 사람에게 어려운 작업을 쉽게 보조해 주거나 어렵지 않더라도 매우 지루하고 실수할 수 있는 작업을 효율적으로 처리하도록 하여 시간을 절약해 준다.

셋째, 생성형 AI는 금융공학 데이터를 학습하는 경우 새로운 파생상품, 구조화된 ETF 등의 금융상품을 개발할 수 있다. 또한 생성형 AI가 기존의 금융상품의 구조에 문제가 없는지 검토하여 그 기능을 최적화하거나 개선된 형태의 금융상품을 출시할 수도 있을 것이다.

그 밖에 생성형 AI를 활용하여 금융 소프트웨어 개발자에게 코드 작성을 지원할 수 있고,[23] 계약서를 분석할 수 있다.[24] 그리고 고객 이탈 요인이나 패턴을 분석할 수 있고, 금융경쟁사의 전략과 시장 동향 등에 관한 정보를 획득하여 전략적 이점을 얻을 수도 있을 것이다.

3. 생성형 AI 등에 대한 법적 규율 현황

1) 현재 우리나라에서는 생성형 AI를 포함하여 AI를 규율하고 있는 법률은 전무한 상황이다. 다만, AI를 총론적으로 직접 규율하기 위해 국회에 계류 중인 법안으로 2022. 12. 7. 발의된 인공지능 산업 육성 및 신뢰 기반 조성 등에 관한 법률안[25]과 2023. 2. 28. 발의된 인공지능책임법안[26]이 있다. 그 밖에 AI와 관련한 것으로 금융 영역에 적용될 여지가 있는 개정안으로 2023. 2. 20. 발의된 개인정보 보호법 일부개정법률안,[27] 2023. 5. 22. 발의된 콘텐츠산업 진흥법 일부개정법률

23) 골드만삭스의 경우 개발자의 코드 작성을 지원하는 자체적인 생성형 AI를, JPMorgan의 경우 소프트웨어 개발과 관련하여 빠르고 정확한 코드를 추천하는 Senatus AI 프로그램을 운영하고 있다.

24) 예를 들어 ZBrain이 있다.

25) 윤두현 의원 대표발의, 의안번호 2118726, AI 산업을 진흥하고 AI 사회의 신뢰기반 조성에 필요한 기본적인 사항을 규정함을 목적으로 한 것으로 고위험 AI 등이 아니고서는 AI 서비스를 자유롭게 허용함을 원칙으로 한다. 금융서비스업은 위 법안 제2조 제3호 가.목 내지 바.목에서 규율하는 고위험 영역에 명시되어 있지는 않다.

26) 황희 의원 대표발의, 의안번호 2120353, 인공지능사업자와 고위험 AI 등을 규율하면서 인공지능사업자의 책무를 규정하고 있다. 금융서비스업은 제2조 제4호의 고위험 영역으로 명시되지는 않지만 로그인, 결제 목적으로 생체인식을 사용할 경우 같은 호 나.목에 따라 고위험 영역에 해당할 여지는 있다.

27) 김영배 의원 대표발의, 의안번호 2120130, AI 회사의 알고리즘으로 인하여 개인정보가 유출된 경우 개인정보보호위원회가 해당 알고리즘 제출을 요구할 수 있도록 하였다.

안[28] 등이 있다.

금융위원회가 2021년 마련한 '금융분야 AI 운영 가이드라인'(이하 '가이드라인')은 생성형 AI를 포함한 전체 AI를 규율하는 지침으로, 금융업계에 중요한 이정표 역할을 할 것으로 기대된다. 위 가이드라인은 제1항부터 제7항까지에 걸쳐 이루어져 있는데, 그 요지로는 AI의 잠재적 위험에 대처하기 위한 금융기관 등의 위험관리정책 마련(제2항), AI 시스템 기획·설계 단계에서 윤리원칙에 부합하도록 할 것(제3항), AI 시스템 개발 단계에서 데이터 품질 개선 방안을 검토하고 민감정보 등에 대한 비식별조치를 할 것(제4항), AI 시스템 평가 및 검증 단계에서 공정성 판단지표 개발 및 설명가능 인공지능 기술을 적용할 것(제5항), AI 시스템 도입·운영 단계에서 오용 및 악용 사례를 최소화하도록 할 것(제6항), AI 시스템 개발 및 운영 업무를 외부기관에 위탁시 엄격하게 점검할 것(제7항) 등이다.[29]

2) 해외 동향을 살펴보면, 유럽연합(이하 'EU')에서는 최근 인공지능법(Artificial Intelligence Act, 이하, 'AI법')이 통과되어 그 시행을 눈앞에 두고 있다. 위 법의 가장 큰 특징으로 용도별 위험성의 정도에 따른 규제의 세분화로 꼽힌다. 즉, 위 법은 AI의 사용용도에 따라 위험성의 정도를 달리 규율하는데, 허용불가 AI, 고위험 AI, 제한적 위험 AI, 최소한 위험 AI 4단계로 구성된 AI 체계를 규정하면서 특히 고위험 AI에 규제력을 집중하고 있다(제6조). 고위험 AI에 해당하는 분야를 세부적으로 규정한 부속서 II, III에 금융영역이 명시적으로 들어가 있지는 않으나 생체인식, 신용도 평가, 생명 및 건강보험과 관련한 위험 평가 등 일부 금융과 연결될 수 있는 분야를 고위험 영역으로 규율하고 있다. 고위험 AI는 위험관리시스템의 구축, 일정한 품질이 검증된 데이터 세트의 사용, 적정한 기술문서의 사용, 인간의 감독, 일정한 사이버 보안조치의 마련 등의 규제를 받는다(제8조 내지 제29a조). 이를 제외한 대부분의 금융 용도 AI는 제한적 또는 최소 위험 AI로 취급되는데 최소 위험 AI는 어떠

28) 이상헌 의원 대표발의, 의안번호 2122180, AI 기술을 이용하여 콘텐츠를 제작한 경우에는 해당 콘텐츠가 AI 기술을 이용하여 제작된 콘텐츠라는 사실을 표시하도록 하였다.

29) 한편 금융위원회가 2022. 8. 발표한 금융 분야 인공지능 활용 활성화 및 신뢰확보 방안의 후속조치로 'AI 기반 신용평가모형 검증체계'와 '금융분야 AI 보안 가이드라인'을 마련하기도 하였다. AI 기반 신용평가모형 검증체계는 AI 특성을 고려하여 신용정보회사가 데이터를 적절히 관리하는지, 신용평가모형에 사용되는 알고리즘과 변수를 합리적으로 선정하였는지 점검하고 금융소비자에게 신용평가모형과 신용평가 결과에 대해 충분히 설명할 수 있는지 검증한다. 한편 금융분야 AI 보안 가이드라인은 AI 모델을 개발할 때 고려해야 할 보안사항을 개발단계별로 제시하고, AI 챗봇 서비스에 대한 보안성 체크리스트를 추가로 제공한다(금융위원회, "금융분야 인공지능의 신뢰를 높인다", 보도자료, 2023. 4. 17).

한 규제도 받지 않고, 제한적 위험 AI로 판단되는 경우 투명성 조치만 이행하면 그 사용이 전면적으로 허용된다(제52조).

위 법의 또 다른 특징으로 생성형 AI 등 범용 AI를 규율하는데 있다[제3조(44b)]. 원래의 초안에서는 이를 명시하지 않았지만 GPT-3의 출시를 계기로 생성형 AI를 별도의 규제 대상으로 명시하였다. 특별한 용도가 없는 생성형 AI는 그 자체로는 위험성이 높다고 취급되지는 않지만 금지된 AI 및 고위험 AI를 포함하여 각 분야의 AI를 구축시킬 수 있는 기반이 될 수 있기 때문이다. 생성형 AI를 포함한 범용 AI는 원칙적으로 제한적 위험이 있는 AI로 취급되는데 이에 따르면 생체 인식이나 감정 인식 AI, 딥페이크 AI를 활용하는 사용자는 이에 노출된 사람들에게 AI를 활용하여 생성 또는 조작되었음을 명확히 공개하여야 한다. 특히 소셜네트워크를 통하여 해당 정보에 노출된 모든 사람들에게는 그것이 AI에 의하여 생성된 사실을 인지할 수 있도록 조치하여야 한다. 다만 고위험 AI 시스템 또는 그 구성요소로 사용될 수 있을 때에는 고위험 AI로 취급되어 해당 규제를 받게 된다(제52a조 내지 제52d조).

3) 일본의 경우 우리나라와 마찬가지로 AI에 대한 직접적이고 구속력 있는 법률 및 규정을 두고 있지 않고 별다른 법률 제정의 움직임도 보이고 있지 않다. 다만 'AI 및 데이터 활용에 관한 계약 지침'(AI・データの利用に関する 契約ガイドライン) 및 'AI 원칙 구현을 위한 거버넌스 지침'(AI原則実践のためのガバナンス・ガイドライン) 등을 마련하여 금융영역을 포함한 관련 업계에 일정한 내용을 안내하고 있다.

4) 미국은 2019년 행정부가 행정명령 제13859호로 발표한 '미국 인공지능 구상' 이후, '생성적 적대 신경망 출력물 확인법'(Identifying Outputs of Generative Adversarial Networks Act)과 '국가 인공지능 구상법 2020'(National Artificial Intelligence Initiative Act of 2020)을 제정하여 AI 연구・개발과 훈련 지원을 중심으로 한 명문 규정을 마련하였다. 그 밖에 정부인공지능법 2020(AI in Government Act of 2020), 인공지능진흥법(Advancing American AI Act) 등의 법령을 두고 있으나 EU의 AI법과 같은 규제적 의미는 크지 않다.[30] [31]

---

30) 한편 미국에서는 '2022 알고리즘 책임법안'이 발의되어 있는데, 특히 오류나 편향, 자동화된 의사결정에 대하여 규율하면서 이에 대한 투명성 조항과 감독 조항을 명시하고 있다.

31) 세계법제정보센터, "세계 각국의 AI 규제 관련 입법동향"(2023. 8)〈https://world.moleg.go.kr/web/dta/lgslTrendReadPage.do?A=A&searchType=all&searchPageRowCnt=10&CTS_SEQ=50807&AST_SEQ=3891&ETC=1〉.

Ⅲ. 생성형 AI의 개별적인 기술구조와 금융에서의 각각의 활용

생성형 AI는 배포 초기 단계로 아직 기술적으로 미성숙한 상태에 있으나, 금융공학과 AI의 기술이 빠르게 발전하고 있기 때문에 그 기술적 활용이 어떻게 진행될지 섣불리 예측하기 곤란하다. 다만, 현재까지 나온 생성형 AI 모델들의 현황들을 살펴볼 때 위 각 모델들을 활용하여 이미 구현하였거나 구현할 수 있을 것으로 예상되는 금융 분야를 살펴보면 다음과 같다.

1. 생성적 적대 신경망(Generative Adversarial Networks: GAN)[32]

생성적 적대 신경망은 서로 적대 또는 대립하는 관계인 2가지 AI 모델, 즉 생성기(Generator)와 판별기(Discriminator)를 동시에 사용하는 비지도학습방법이다. 생성기는 문자, 음성, 이미지 등을 생성하는 데 최대한 사실에 가깝게 위조한다. 판별기는 위 위조 내용과 실제 내용을 모두 수신하고, 두 정보를 최대한 정확하게 구별하는 것을 목표로 한다.[33] 판별기에서 나타내는 출력값이 1이면 위 위조 내용이 실제에 부합하는 것이고, 0이면 가짜로 판명되는 것이다. 따라서 생성기가 생성한 내용은 1에 가까울수록 좋다고 볼 수 있다.[34]

예를 들어, 생성기는 인공적으로 생성된 금융거래 데이터를 판별기에 제공하고 판별기는 생성기가 만든 거래 데이터와 실제 금융거래 데이터를 모두 학습하여 그 차이(손실)를 최대한 줄이고 그 내용을 생성기에 전달하여 성능을 개선하게 된다.[35] 이와 같이 두 모델이 경쟁하면서 발전하여 나가는데,[36] 훈련 수(Epoch)가 증가할 때마다 생성기가 생성하는 거래 데이터가 더 정교해진다.

금융 분야에서 생성적 적대 신경망을 활용하는 방법은 다음과 같다. ① (사기성

32) 현재는 잘 사용되고 있지는 않지만 딥페이크와 관련하여 대표적인 생성형 AI이므로 이를 소개한다.

33) Antonia Creswell *et al.*, "Generative Adversarial Networks: An Overview", *IEEE Signal Process Magazine* 35(1)(2018), pp.53-54

34) 이때 생성된 데이터가 실제 데이터와 일치할 정도가 된다면 판별기는 생성기가 제공하는 모든 입력 데이터에 대하여 0.5 이상을 나타내어 혼란스러운 반응을 보이게 된다(Antonia Creswell *et al.*, *id.*).

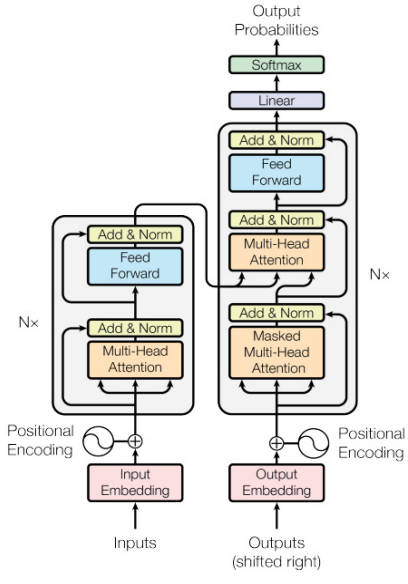
35) 생성기는 실제 데이터에 직접 접근할 수는 없고 이를 학습하는 유일한 방법은 판별기와의 상호작용을 통해서만 가능하다. 구체적으로 위 방법은 판별기의 잘못된 분류에 대하여 패널티를 부과하고 위 차이에 대하여 역전파를 통하여 입력 가중치를 새로 설정하는 방식으로 교정하게 된다.

36) Suman Kalia, "Potential impact of generative artificial intelligence(AI) on the financial industry", *IJCI Journal*, Vol. 12(2023), p.38.

거래, 시스템적 위험 탐지) 평상시와 비교하여 비정상적인 거래 패턴이나 수치를 식별함으로써 합법적인 거래와 사기성 거래를 구분하거나[37] 시스템적 위험을 미리 감지할 가능성이 있다. ② (악성코드 탐지) 금융보안을 침해하는 악성코드를 탐지하고 이를 분류할 수 있다.[38] ③ (데이터 편향 문제 해결) 많은 데이터를 최대한 학습할 필요 없이 정확하고 대표성 있는 합성 금융 데이터를 생성하여 데이터 편향(Bias) 문제를 해결할 수 있다.[39] 이에 따라 금융위험의 모델링과 리스크 관리, 포트폴리오의 최적화에 활용될 수 있다.

2. 트랜스포머 모델(Transformer model)

[트랜스포머 모델의 구동 메커니즘]



(출처: Ashish Vaswani *et al.*, "Attention Is All You Need", *Neural Information Processing Systems*(2017), p.3)

37) 특히 신용카드 사기거래 탐지에 있어서 상당한 효과를 발휘할 수 있다. 이에 대하여는 Emilija Strelcenia, Simant Prakoonwit, "A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection", *Mach. Learn. Knowl. Extr.* 5(1)(2023), pp.304-329 참조〈https://www.mdpi.com/2504-4990/5/1/19〉.

38) 이에 대하여는 Ziyue Wang *et al.*, "CNN-and GAN-based classification of malicious code families: A code visualization approach", *International Journal of Intelligent Systems*, Vol. 37(2022), pp.12472-12489 참조.

39) Emilija Strelcenia, Simant Prakoonwit, *supra* note 37, p.311.

트랜스포머 모델은 오늘날 우리가 볼 수 있는 대규모 언어 모델(LLM)과 생성형 AI의 기반을 마련한 것으로, 기존의 순환신경망이나 합성곱신경망과는 별개인 '자기 주의 기법'(Self-attention)을 활용한다. 기계번역을 예로 들면 기존의 순환신경망이나 장·단기 기억네트워크는 번역을 단어 하나씩 처리하고 문장과 같은 긴 시퀀스의 유기적인 해석을 용이하게 해 내지 못하는 단점이 있는 반면에, 위 모델은 단어들 간의 관계, 의미를 추론하여 문장 정보를 정확하게 파악하고 문맥의 전체적인 흐름을 이해하여 번역하는 데 도움을 준다.

자기 주의 기법의 핵심 개념은 쿼리(Query, 해석하려는 단어와 다른 단어들과의 관련성을 파악), 키(Key, 단어들과의 유사성을 측정), 밸류(Value, 해당 단어의 중요도 측정)인데, 개별적인 여러 단어들에 대하여 각각 쿼리, 키, 밸류로 이루어진 세 가지 벡터를 생성한다. 위 세 가지 벡터를 통하여 입력된 각 단어와 다른 단어들과의 관련성과 유사도를 측정하고 (문맥에 따라 측정되는) 중요성의 정도에 따라 가중치를 부여함으로써 입력된 단어들 간의 상호작용을 강화한다. 이에 따라 트랜스포머 모델은 기계 번역, 문장 생성 및 문자 기반 활동에 효과적인 기능을 발휘할 수 있다.[40]

금융 분야에서 트랜스포머 모델을 활용하는 방법은 다음과 같다. ① (감정 분석) 금융 관련 뉴스, 소셜미디어 게시물 등에 나와 있는 감정이나 표현을 문맥의 흐름에 따라 이해함으로써 시장 동향과 투자자들의 정서를 분석할 수 있다. 또는 고객들의 상담 평가에 나타난 감정을 포착하여 고객에게 개별화된 투자조언을 할 수 있다. ② (금융 예측) 트랜스포머 모델은 시계열 예측(Time series)[41]에 좋은 성능을 발휘한다.[42] 데이터에서 나타나는 변화 양상과 추세를 포착함으로써 주가, 에너지나 곡물 수요 예측 등 다양한 금융응용 분야에 대하여 정확하게 예측할 수 있고, 특히 금융 포트폴리오 최적화에 도움을 줄 수 있다.[43] ③ (재무 분석) 트랜스포머 모델은 재무적인 흐름을 관찰하여 재무보고서나 설명서를 자동으로 생성하는 데 도움을 줄 수 있다.[44]

---

40) 자세한 구동 메커니즘은 Ashish Vaswani *et al.*, "Attention Is All You Need", *Neural Information Processing Systems*(2017) 참조.

41) 시간의 흐름에 따라 기록된 것.

42) Caosen Xu *et al.*, "A Financial Time-Series Prediction Model Based on Multiplex Attention and Linear Transformer Structure", *Appl. Sci.* 13(8)(2023), pp.14-15.

43) 이에 대하여는 Edmond Lezmi, Jiali Xu, "Time Series Forecasting with Transformer Models and Application to Asset Management"(2023) 참조〈https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4375798〉.

44) Tuna Tuncer *et al.*, "Asset Price and Direction Prediction via Deep 2D Transformer and Convolutional Neural Networks", Proceedings of the Third ACM International Conference on AI

3. 다중모드(Multimodal) AI[45]

생성적 적대 신경망, 변형 자동 인코더, 확산 모델, 트랜스포머 모델의 발전과 범용성이 있는 대형 언어 모델의 출현 등을 기반으로 다중모드 AI가 개발되었다. 다중모드 AI는 이미지, 음성, 문자 등 여러 유형의 데이터를 결합하여 더 정확한 예측을 하거나 통찰력 있는 결정을 내리는 모델이다. 다중모드 AI와 기존의 단일모드 AI의 차이점은 학습하는 데이터 유형이다.[46] 예를 들어, 단일모드인 금융 AI는 거시경제 데이터와 산업 데이터 및 금융 데이터만을 활용하여 금융시장을 예측하거나 분석하는 단일한 업무를 한다. 반면 다중모드 AI는 금융시장의 흐름에 영향을 미치는 객관적인 지표 외에도 투자자들의 감정, 감각 등도 수집 및 처리하여 특정한 조건이나 환경에서 더 자세하고 정확한 결과를 낼 수 있다.

다중모드 AI는 입력 모듈과 출력 모듈 외에 융합 모듈이 따로 있고 위 융합 모듈에서 이미지, 음성, 문자, 동영상 등의 각 데이터를 하나의 데이터 세트로 결합, 정렬 및 처리하는 기능을 담당한다. 이에 따라 음성 데이터만으로 일정한 이미지를 생성하거나 특정 이미지에 나타난 풍경, 감성 등을 문자로 설명하는 등 서로 다른 데이터 간 가로지르기(Crossmodal)가 가능해진다. 다중모드 AI에는 텍스트 분석기술,[47] 자연어 처리기술(Natural language processing),[48] 이미지 및 동영상 데이터에 대한 컴퓨터 시각기술, 여러 유형의 데이터에 대한 통합 시스템 기술 등이 필요하다.

금융 분야에서 다중모드 AI를 활용하는 방법은 다음과 같다. ① (정확한 투자 판단) 앞서 나온 예시와 같이 주식시장의 금융 분석을 함에 있어 객관적인 금융지표 외에도 상황 포착 및 인식에 탁월한 성능이 있어서 투자자들의 감정 및 정서 분석을 도입하여 금융기관의 정확한 투자 판단에 도움을 줄 수 있다.[49] ② (고객 상담에

---

in Finance(November 2022), pp.79-86.

45) 이 외에 확산 모델(Diffusion model)도 금융에서 중요한 역할을 하는 생성형 AI이나 여기서는 생략한다.

46) 다중모드 AI는 이미지 생성기술과 대형언어모델을 접목한 문자-이미지 생성 모델(주어진 문자로 이미지를 생성하는 AI)이 대표적이고, 이를 계기로 GPT-4와 같은 대화형 인공지능의 입력 프롬프트 다변화(음성, 이미지, 문자 등)가 시작되었다[채명식·조유리, "2023 인공지능", KISTEP 한국과학기술기획평가원(2023), 13쪽].

47) 이를 통하여 서면에 나타난 문서의 의도와 목적을 이해할 수 있게 해준다.

48) 자연어 처리기술은 음성 출력 및 인식 기능, 음성-문자 변환 기능 등을 제공한다.

49) 이에 대한 연구로 Yu-Fu Chen, Szu-Hao Huang, "Sentiment-influenced trading system based on

활용) 상담을 진행하는 금융소비자의 반응을 정확하게 판단할 수 있다. 예를 들어 "아주 잘하네요"라고 말하는 고객을 가정해 보자. 위 내용은 칭찬의 뜻을 내포할 수도 있는 반면에 비꼬는 표현이 될 수도 있는데, 기존 AI는 그 문구 자체만 인식을 하나, 다중모드 AI는 말투나 얼굴 표정과 같은 다른 유형의 데이터도 수집을 하여 정확한 반응을 구별할 수 있다. ③ (회사 및 개인의 신용 분석에 활용) 회사의 신용등급, 개인의 신용점수 등에 대하여 다중모드 AI를 활용하여 데이터를 융합하면 더 우수하고 정확한 결과를 도출할 수 있다.[50] ④ (금융업무의 간소화) 간단한 설명을 기초로 문서, 이미지, 동영상 등을 생성할 수 있으므로 내부 보고서, 엑셀파일, 광고 동영상 등 여러 업무를 쉽고 간편하게 처리할 수 있다.

Ⅳ. 금융법상 쟁점과 과제

앞서 본 바와 같이 생성형 AI는 금융 분야에서 다양하게 활용될 것으로 보인다. 그럼에도 불구하고 생성형 AI의 사용에 있어 ⓐ 조작 위험(딥페이크), ⓑ 금융안정성의 위험(플래시크래쉬, 금융시스템 리스크), ⓒ 투자자보호 관련 위험, ⓓ 데이터 유출 위험, ⓔ 불공정의 위험(데이터 편향, 자동화된 의사결정) 등과 같은 법적 문제들이 있을 것으로 예상되므로 아래에서 살펴보기로 한다.

1. 조작 위험: 딥페이크

1) '딥페이크'(Deepfake)란 AI 등 기계 학습 기술을 활용하여 생성하거나 수정된 실제 또는 가상의 동영상, 이미지, 음성 및 문자를 말한다. 딥페이크를 만드는 방법은 여러 가지가 있지만 최근에 활용되는 것은 얼굴 교환 기술을 사용하는 생성형 AI를 이용한 것이다. 특히 생성적 적대 신경망은 딥페이크의 결함을 감지하고 개선하여 딥페이크 탐지기가 그 출처를 추적하기 어렵게 만든다.[51] 아울러 다중모드 AI

multimodal deep reinforcement learning", *Applied Soft Computing* 112(4)(2021) 참조.

50) 관련 연구로 Mahsa Tavakolia *et al.*, "Multi-Modal Deep Learning for Credit Rating Prediction Using Text and Numerical Data Streams"(2023)이 있다〈https://arxiv.org/abs/2304.10740〉.

51) TC Helmus, "Artificial intelligence, deepfakes, and disinformation: A primer", *Rand*(2022), p.3.

의 발전으로 인하여 딥페이크는 매우 정교해지고 쉽게 접근할 수 있어[52] 잠재적으로 여러 금융상 피해가 발생할 수 있다.[53]

2) 금융 분야에는 생성형 AI를 이용한 딥페이크 사기 피해가 주로 세 가지 측면에서 이루어지고 있다. 첫째, 금융 신원 도용으로 인한 사기 피해이다. 예전에는 탈취한 데이터로 피해자의 명의로 신용카드를 발급받거나 대출을 받았다면, 문자·이미지 등을 쉽게 생성하는 생성형 AI의 발달에 따라 온라인으로 피해자의 얼굴과 음성을 흉내내어 금융기관이나 가상자산거래소의 신원 확인 절차를 무력화할 수 있다.[54]

둘째, 딥페이크를 이용한 허위 거래정보 유포를 통한 불공정거래이다. 주가 등락에 따른 금융상 이익을 얻기 위하여 테러가 발생하였다는 자막과 함께 딥페이크로 된 영상을 올릴 수 있다.[55] 또는 생성형 AI를 통하여 허위 재무제표나 투자보고서를 만들 수 있고, SNS에 투자자들에게 오해를 일으키는 시장 분석이나 조작된 투자추천을 올릴 수 있다.[56]

셋째, 딥페이크를 활용한 보이스피싱이나 메시지 피싱 등으로 인한 금융상의 피해를 들 수 있다.[57] 이 경우 피해자들 각각의 특성을 고려한 전화 음성, 이메일 또는 문자메시지를 짧은 시간 내에 맞춤식으로 대량 제작하여 살포할 수 있다.[58] 또한 실제로는 존재하지 않는 투자자문업체 직원과 줌(Zoom) 등 화상회의 프로그램으로 영상 대화를 하면서 신뢰관계를 쌓은 다음 자신의 금융정보를 공개하거나 투자 명목으로 거액을 송금하도록 할 수 있다. 그 외에도 생성형 AI를 활용한 악성코

---

52) 예를 들어 Midjourney는 자연어에서 이미지를 생성하는 생성형 AI 모델인데, 위 AI를 통하여 생성된 여러 개의 딥페이크 이미지(체포되는 도날드 트럼프 전 미국 대통령의 사진과 하얀색의 멋진 재킷을 입은 프란치스코 교황 사진 포함)를 남용하는 행위가 급증하여 무료평가판의 사용이 중단되었다.

53) The New York Times Aug. 30th, 2023 article "Voice Deepfakes Are Coming for Your Bank Balance"에 따르면 최근 1년 사이 딥페이크를 활용한 보이스피싱이 급증하였음을 기술하고 있다.

54) 이와 유사한 방법으로 신용 도용은 아니지만 계좌를 개설하여 가짜 유동성이나 자산의 존재를 만들어 낼 수도 있다.

55) CNN May 23th, 2023 article, "'Verified' Twitter accounts share fake image of 'explosion' near Pentagon, causing confusion". 이 기사는 미국 국방부 근처에서 폭발이 일어났다고 주장하는 가짜 이미지가 소셜네트워크인 X의 여러 계정에 유포되어 주가가 하락한 사건에 관한 것이다.

56) David S. Krause, "Mitigating Risks for Financial Firms Using Generative AI Tools"(2023), pp.6-7 〈https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4452600〉.

57) Techcircle May 2nd, 2023 article, "Nearly half of Indian internet users faced AI-driven voice scams this year".

58) The Wall Street Journal pro Cybersecurity August 30th, 2019 article, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case"에 따르면 영국의 한 회사가 딥페이크 음성을 이용하여 모회사의 CEO로 가장한 자로부터 긴급 송금 요청을 받고 243,000달러를 송금하여 사기 피해를 입기도 하였다.

드 생성 및 수정, 비밀번호 해독을 함으로써 금융기관의 보안을 침해할 수 있다.

3) 이에 대응하는 기술적 조치로 '딥페이크 탐지' 및 '원본 인증'이 거론되고 있다. 딥페이크 탐지는 주로 조작되었다는 점을 뒷받침할 만한 증거를 찾고 그 증거를 수치로 표시하거나 시각적 형태로 제시하여 조작 여부에 대한 추가적인 분석이 필요함을 알리는 방법으로 이루어진다.[59] 시중에서 개발된 탐지 프로그램은 대표적으로 GPTZero, FakeCatcher, FaceForensics Benchmark 등이 있다. 다만 이와 같은 탐지기는 학습하기 위하여 사용되는 학습데이터 세트가 다양하지 않으면 성능이 저하되는 문제가 있다.[60] 그뿐만 아니라 위 탐지기는 사람이 작성한 콘텐츠를 AI가 생성한 것으로 식별하거나 그 반대로 기능하는 이른바 '오탐지'(false positive)[61] 문제가 생길 수 있다.

원본 인증은 이미지나 동영상 등 원본에 고유정보를 입력하고 그것이 복제되거나 수정된 경우 고유정보가 변경됨으로써 복제 또는 수정되었다는 것을 알 수 있도록 설계할 수 있다. 그 방법으로 원본에 보이지 않는 디지털 워터마크를 하는 방안,[62] 원본에 고유의 해시정보를 삽입하여 딥페이크 수정본의 그것과 비교하는 방안, 장치 및 브라우저에 고유 지문(Device and Browser Fingerprinting)을 부착하는 방안[63] 등이 있다. 다만, 이 경우에는 인증 표준이 채택되어 보편화되어야 할 것인데, 전 세계적으로 이에 대한 확립은 이루어지지 않은 상황이다.

4) 한편 딥페이크에 대한 법적 규제와 관련하여 금융상의 조치에 한정하여 이루어진 것은 전무하다. 그 범위를 넓혀 보면 앞서 본 EU의 AI법에서는 생성형 AI로 딥페이크를 생성한 자에게 용도를 불문하고 투명성 조치를 의무화하였다. 이에 따라 생성자는 딥페이크에 노출된 사람들에게 AI를 활용하여 딥페이크를 생성하거나 원본을 조작하였음을 명확히 공개하여야 한다.[64] 그리고 소셜네트워크를 통하여 해

---

59) National Secuirty Agency *et al.*, "Contextualizing Deepfake Threats to Organizations"(2023), p.6.

60) Binh Le *et al.*, "Why Do Facial Deepfake Detectors Fail?", WDC '23: Proceedings of the 2nd Workshop on Security Implications of Deepfakes and Cheapfakes(2023), p.5.

61) 거짓 양성(False Positive) 및 거짓 음성(False Negative).

62) CNBC May 10th, 2023 article, "Google will label fake images created with its A.I."

63) 딥페이크 생성자의 컴퓨터와 운영체계, 브라우저, 플러그인 등에 대하여 관련 데이터를 수집하고 이를 지문이라고 부르는 고유 식별자와 결합하여 생성자의 장치를 추적하고 의심스러운 연결을 차단하는 방법을 말한다.

64) 미국에서는 연방 의회 차원에서 딥페이크 책임법안(Deepfake Accountability Act)이 발의되어 계류중인데, 딥페이크된 동영상, 이미지 등에 투명성 조치의무를 부과하고 특정 신분증과 관련한 사기범죄에 딥페이크도 포함하도록 규정하고 있다. 주 차원에서는 텍사스주, 캘리포니아주, 버지니아주에

당 정보에 노출된 모든 사람들에게 그것이 AI에 의하여 생성된 사실을 인지할 수 있도록 조치하여야 한다.

다만 투명성 조치를 교묘하게 회피하는 모습이 나타날 여지가 있다. 예를 들어, 투명성 조치에 따라 수정 또는 조작되었다는 것을 알리는 문구나 음성 등이 동영상에 게재되었다고 하더라도 그것이 동영상 말미에 나타내거나 이미지 구석에 시각적으로 보기 힘든 매우 작은 글씨로 나오게 하여 그 내용을 사람들이 인식하기 어렵게 할 수도 있을 것이다.

우리나라의 경우 딥페이크를 직접 규제하는 법은 없고 국회에 제출된 인공지능산업육성법안이나 인공지능책임법안 등에도 이를 다루고 있지 않다. 정보통신망 이용촉진 및 정보보호 등에 관한 법률 제4조 제2항 제7의2에서는 정보통신망을 통하여 유통되는 정보 중 AI 기술을 이용하여 만든 딥페이크 정보를 식별하는 기술의 개발 및 보급을 위한 시책을 마련할 의무를 정부기관에 부여하고 있을 뿐이다. 기존의 법률로는 자본시장과 금융투자업에 관한 법률(이하 '자본시장법') 제177조의 시세조종 금지 조항, 제178조의 부정거래행위 금지 조항, 가상자산 이용자 보호 등에 관한 법률(이하 '가상자산이용자보호법') 제10조의 불공정거래행위 금지 조항 등을 딥페이크를 이용한 불공정거래행위에 규율할 수 있다. 그리고 전기통신사업법 제22조의3 제1호에 따라 특수유형부가통신사업자(인터넷 사업자 등)에 대하여 정보의 위·변조와 관련하여 딥페이크를 방지하기 위한 기술적 조치의무를 가할 수 있는 여지가 있다.

2. 금융 안정성의 위험

1) 생성형 AI는 금융공학의 발전 양상을 반영하여 새로운 파생상품을 개발하거나 ETF의 구조를 설계하는 데 기여할 수 있다. 또한 자동 주식 매매 프로그램을 통하여 인건비를 최소화시키면서 최대한의 수익을 낼 수 있는 가능성이 있다.

그와 반면에 생성형 AI를 증권이나 가상자산 등의 거래에 활용하는 경우 기존에 비하여 금융시스템적 위험을 더 크게 만들 수 있다. 단순 거래 알고리즘으로 인하여 큰 혼란을 유발한 사건으로 2010년에 미국에서 발생한 플래쉬크래시(Flashcrash) 현상이 있고 알고리즘의 입력 정보를 잘못 눌러서 수백억원의 손해를 입은 사안인

서는 선거와 포르노를 표적으로 삼는 딥페이크를 규제하는 법률만 통과되었을 뿐이다.

우리나라의 한맥투자증권 사건도 있다.[65] 생성형 AI는 이에 더 나아가서 여러 시스템의 기반이 되는 체계로 수많은 핀테크 거래시스템이 이를 기초로 구축되어 있어서 AI 시스템이 무단으로 침해받거나 그 설계 자체에 결함이 있는 경우 여러 거래시스템에 광범위한 영향을 미쳐 대형 플래쉬크래시를 유발할 수 있다. 그뿐만 아니라 AI의 거래시스템이 추세추종형 거래를 하도록 설정되었거나 사실과 다른 오류정보를 생성하고 그것이 입력되어 거래주문이 처리되는 경우 그 파급 효과로 플래쉬크래시 등이 유발될 가능성이 있다.

2) 한편 딥러닝을 기반으로 하는 생성형 AI는 금융투자상품이나 가상자산 시장을 조작하는 방법을 학습하는 AI 거래시스템을 구현할 수 있는 위험이 있다. 생성형 AI로 구현된 거래시스템이 어떠한 제한을 부과하지 않고 이익만을 달성하도록 설계된 경우 허수주문을 하거나[66] 가장매매를 하는 등 시장을 조작할 수 있다. AI 거래시스템이 미공개중요정보를 이용하거나 허위의 정보를 유포하여 이익을 취할 수도 있을 것이다. 그리고 고객의 거래 데이터를 이용하여 금융기관의 거래를 선행적으로 실행하여 이익을 취하거나 거래로 일어난 손실을 고객에게 떠넘길 가능성도 있다고 보인다.

그럼에도 불구하고 인간의 고의 또는 과실이 개입되지 않고 AI가 자율적으로 투자전략을 조정하여 시장을 조작하는 방법을 발견하여 실행하였을 뿐인 경우 이를 자본시장법 제176조 이하나 가상자산이용자보호법 제10조의 각 불공정거래 규제를 적용할 수 없어 문제될 수 있다. 이에 대하여 AI의 시장조작을 미리 막기 위하여 학습 데이터나 알고리즘에 공정성 기준을 부과하고 사용자 등에게 사전적 또는 사후적으로 관리, 감독을 수행하고 이를 미이행한 경우 민사상 사용자책임과 같은 과실책임을 입법하는 방안, AI 개발자 등에게 제조물책임법과 같은 무과실책임을 부과하는 방안 등을 생각해 볼 수 있다.

3. 투자자보호 관련 위험

1) AI가 인간을 배제하고 개별적으로 투자 판단을 하는 경우가 적지 않다. 특

65) 대법원 2023. 4. 27. 선고 2017다238486(본소), 2017다238493(반소) 판결.

66) Enrique Martínez-Miranda *et al.*, "Learning Unfair Trading: A Market Manipulation Analysis from the Reinforcement Learning Perspective", 2016 IEEE Conference on evolving and adaptive intelligent SYS(2016), pp,107-108에서는 인간의 사전적 의사 없이 AI가 스스로 학습하여 허수주문을 할 수 있음을 보여 주었다.

히 생성형 AI는 전통적인 방법으로는 생성하기 어려운 투자 방법을 개발하거나 투자설명서를 만들어 내는 능력을 가지고 있다. AI의 투자 판단 및 투자자 보호와 관련하여, 집합투자업과 투자일임업은 가장 쟁점이 되는 분야로[67] 투자권유 단계와 운용 단계로 나누어 살펴볼 수 있다.

2) 투자권유 단계에 있어서는 금융소비자 보호에 관한 법률(이하 '금융소비자보호법')에서 규정하는 설명의무, 적합성 원칙과 적정성 원칙 등은 AI가 인간 직원을 배제하고 투자결정을 하는 경우에도 적용된다. 자산운용계약을 체결할 때 집합투자업자는 고객에게 투자대상, 투자에 대한 위험, 위험 등급, 금융상품의 구조, 기대수익 등 중요한 사항에 대하여 설명하면서[68] 고객의 여러 정보를 고려한 적합하고 적정한 상품을 권유하여야 하고,[69] 이는 AI가 투자 판단에 일정 부분 관여하는 경우에도 마찬가지이다. 만약 투자를 권유하는 과정에서 제시하는 금융상품 설명서나 요약서, 보조자료가 생성형 AI에 의하여 작성되었다면 자문인력은 이에 대한 검토를 수행하여야 하고, 고객에 대한 투자권유 과정에서도 상당 부분이 AI에 의한 분석 내용임을 소개하면서 이에 대한 자신의 검토 내용도 언급해야 할 것이다.

한편 투자 판단의 전부 또는 일부에 대하여 위임받는 투자일임업자에 있어서는, 보통 인간 직원이 투자 여부에 대한 결정을 내리는 경우에는 어떠한 투자 요소를 가중치로 두고 평가할지, 투자 시점이나 투자 액수에 대한 판단 근거는 어떠한 것인지에 대하여 고객에게 설명할 의무가 발생하는 것이 원칙이다. 그러나 AI가 투자 판단을 하는 경우 복잡한 매개변수와 수많은 입력 정보가 출력(투자 판단)에 영향을 미치는 등 설명하기 어려워 설명의무가 발생하지 않는다고 봄이 타당하다.[70] 다만, AI 시스템을 활용하여 투자결정을 내리는 경우 그것이 만능인 것처럼 고객이 오해하거나 AI를 이용한 투자전략에 위험이 있음에도 고객에게 이를 이해시키지 않는 경우 설명의무 위반이 발생할 수 있다.[71]

---

67) 금융위원회는 2016. 3. ① 투자자 성향분석 및 포트폴리오 구성, ② 고객정보 보호, ③ 해킹 방지 및 재해 대비 등에 대한 보안성, ④ 공개 테스트 시행의 요건을 갖춘 로보어드바이저에게 비대면 투자자문 및 일임업무를 허용하였다.

68) 금융소비자보호법 제19조 제1항, 같은 법 시행령 제13조 제4항, 금융소비자 보호에 관한 감독규정 제12조 제4항 [별표 3] 제2호.

69) 금융소비자보호법 제17조, 제18조.

70) 집합투자업의 경우에도 투자비율 등을 미세 조정하는 경우가 발생하고 이는 금융환경 변화에 따라 달라지는데, AI의 투자 판단에 있어서도 그 미세 조정 근거를 설명하기 쉽지 않다고 보인다.

71) 鹿島 みかり・千葉 誠・関口 健太, "投資判断におけるアルゴリズム・AI の利用と法的責任: アルゴリズム・AIの利用を巡る法律問題研究会報告書の概要", 金融研究, 日本銀行金融研究所(2019), 13頁.

3) 자산운용 단계에 있어서는 일반적으로 집합투자업자나 투자일임업자의 투자 판단에 대하여 각각 정도는 다르나 일정한 재량이 인정되므로,[72] 고객의 투자목적 등에 비추어 볼 때 과도한 위험을 초래하는 거래행위에 해당하는지 여부에 따라 선관주의의무 위반인지 검토할 수 있다. 다만, AI가 직접 투자 판단을 행하는 경우 인간이 그 판단의 근거와 프로세스를 이해할 수 없다면 그 당부에 대한 판단을 하기 어렵다. 따라서 AI가 전적으로 투자결정을 하는 경우에는 개별적인 투자결정의 합리성보다는 '해당 AI를 이용하여' 투자를 하는 것이 합리적인지 여부를 고려하여야 할 것으로 보인다.[73]

이에 따라 인간을 배제하고 AI 시스템을 활용한 투자 판단을 실시하는 것 자체가 합리적인지, 특정 모델의 AI 시스템을 활용한 것이 타당한지 등에 대하여 들여다보아야 할 것으로 보인다. 특정 모델의 AI를 이용하는 것이 합리적인지 판단하는 것 자체가 쉽지 않은 경우가 있을 수도 있는데, 그 책임을 확실하게 회피하기 위하여는 해당 AI의 성질에 대하여 고객에게 설명하고 이에 따른 책임에 대하여 미리 협의할 필요가 있다.[74]

나아가 금융투자업자는 AI의 훈련 데이터 세트에 문제는 없는지, 데이터와 알고리즘에 편향은 없는지 등을 살펴 보아야 할 것이다. 금융투자업자는 AI 투지시스템에 데이터 편향 등 문제가 없는지[75] 정기적으로 검사하고 외부 침입이 없도록 보안감사를 실시하여야 한다. 만약 검사를 소홀히 하여 데이터 세트의 문제점이나 편향 등을 인지할 수 있음에도 이를 감지하지 못하였다면 선관주의의무 위반 책임을 질 수 있다. 또한 AI는 판단 시점 이전의 데이터를 학습하고 판단하므로 관련 법령이나 제도의 변경이 있는 등 상황이 변화한 경우 합리적인 판단이 이루어지지 않을 가능성이 있으므로 이를 지속적으로 업데이트하고 관리할 수 있는 전문가의 채용이 필요하다.[76]

---

72) 대법원 2008. 9. 11. 선고 2006다53856 판결.

73) 鹿島 みかり・千葉 誠・関口 健太, 同前, 17頁.

74) 同前.

75) 이와 관련하여, 생성형 AI의 환각(hallucination) 효과가 금융 분야에서 무시할 수 없는 문제로 대두된다. 생성형 AI가 종종 신뢰할 수 없거나 타당하지 않은 결과를 도출할 수 있고 이를 '환각적 출력'이라고 하는데 앞에서 적시한 데이터 세트, 편향 문제 외에도 매개변수의 문제, 훈련 데이터의 노이즈, 해킹 등 다양한 원인에 의하여 발생할 수 있다. 특히 금융 투자에서 생성형 AI의 환각으로 인하여 부적절한 투자 판단을 하거나 분별력 없는 고객에게 잘못된 상품을 추천할 수 있다. 따라서 생성형 AI를 사용하는 금융기관의 법률 분쟁과 평판 저하의 위험을 야기할 수 있고 자칫하다가는 금융시스템적 위기까지 비화될 수 있다.

76) 鹿島 みかり・千葉 誠・関口 健太, 同前, 20頁.

만약 금융투자업자에게 선관주의의무 위반 책임이 인정되기 어렵다면 금융투자업자가 AI 시스템을 개발·운영 등을 외부기관에 따로 위탁한 경우에 한하여 고객은 외부기관에 직접 제조물책임법상 책임[77]이나 불법행위책임을 따로 물을 수 있을 것이다. 금융투자업자의 고객에 대한 선관주의의무 위반 책임이 인정된다면 금융투자업자가 위 외부기관에 위 제조물책임법상 책임이나 하자담보책임을 물을 수도 있을 것이다. 이를 방지하기 위하여 위 외부기관은 전문적인 지식과 경험이 있으므로 금융투자업자에게 해당 AI의 구조와 특징, 한계 등을 정확하고 상세하게 설명하고 관련 자료를 교부하는 것이 바람직하다.

4. 데이터 유출 위험

1) 금융기관들은 생성형 AI 시스템을 활용하여 운영의 효율성을 기하기 위하여 노력한다. 예를 들어, 금융 시스템에 보안 결함이 있는지 AI에게 소스 코드를 검토하도록 할 수 있고 고객 정보를 통합하여 보고서 작성 및 편집을 하도록 시킬 수 있다. 또는 금융기관의 이사회 또는 직원 회의에서 있었던 회의 내용이 녹음된 파일을 생성형 AI로 하여금 회의록으로 작성하도록 하기 위하여 업로드할 수도 있다. 여기서 발생할 수 있는 문제는 그와 같은 과정에서 고객의 금융정보가 유출될 수 있다는 점이다.

금융기관은 고객의 재산 현황, 채무액수, 신용점수 등 매우 민감하고 보안이 요구되는 정보[78]들을 보유하고 있다. 따라서 금융 분야에서의 고객정보 유출은 심각한 상황을 야기할 수 있다. 이는 개인의 사생활 침해뿐만 아니라 보이스피싱 등 2차적 사고로 이어져 금전적 피해를 유발할 수 있고 나아가 금융시장 전반에 대한 신뢰를 저하시킬 수 있다.[79]

공개된 생성형 AI를 이용하면 민감한 금융정보가 외부로 유출될 위험이 크다. 이에 많은 금융기관들은 챗GPT 등 공개 생성형 AI 사용을 금지하고 자체적으로 더

77) AI 프로그램이나 소프트웨어 자체는 '제조물'로 볼 수 없지만, AI가 특정 기기에 설치되면 AI를 포함한 전체 기기가 제조물을 구성하여 제조물책임법이 적용될 여지가 있다(서울중앙지방법원 2006. 11. 3. 선고 2003가합32082 판결).

78) 현재 비금융 개인 데이터는 개인정보 보호법에 따라, 금융 데이터는 신용정보법에 따라 규율을 받고 있다.

79) 그 위험성은 2012년부터 2013년까지 사이에 우리나라에 있었던 일부 신용카드회사의 고객정보 유출 사건에서 충분히 알 수 있다.

작고 전문화된 생성형 AI 시스템을 개발하고 있다.[80] 이와 같이 금융서비스 분야는 다른 분야에 비하여 외부에 공개되는 생성형 AI에 대하여 엄격한 입장이라고 평가된다. 그럼에도 불구하고 금융기관들 중 1/5만이 공개 생성형 AI 이용에 대한 금지조치를 시행하고 있다.[81] 또한 금지조치를 시행하여 자체 개발하더라도 여전히 부분적으로 금융정보 보호 문제가 계속될 가능성이 높다. 이 문제는 근본적으로 생성형 AI가 인터넷이나 소셜미디어 등 외부에서 데이터를 수집하는 특성을 가지고 있기 때문이다.

2) (자체 개발된) 생성형 AI에 있어 주로 나타나는 금융기관의 고객정보 유출은 ① AI 훈련용 데이터 세트인 원데이터의 외부 유출, ② 가명 또는 익명화된 데이터의 실명화, ③ 프롬프트 노출에 따른 개인정보 유출과 같은 모습으로 나타날 수 있다. AI 훈련용 데이터 세트(Data set)의 외부 유출은 이름, 주소 등과 같이 개인의 식별정보나 신용카드 정보와 같은 민감한 정보를 타인에게 노출시킨다. 이에 따라 고객 데이터의 암시장 판매(다크웹에서의 경매)가 일어날 수 있고 공격자가 피해 회사에 협박을 하거나 경쟁업체가 유출된 데이터를 이용함으로써 경쟁 우위가 저하될 가능성을 일으키게 된다.

생성형 AI의 훈련에 사용된 가명 또는 익명화된 데이터가 실명화되어 외부에 노출되는 경우도 상정할 수 있다.[82] 가명정보[83]는 상업적 이용을 포함한 통계 작성, 연구 등의 목적이라면 개인 동의 없이 사용할 수 있다. 익명정보는 제한 없이 자유롭게 사용할 수 있다. 데이터를 가명화 또는 익명화하면 개인정보가 이와 연결될 수 없으므로 민감한 정보가 노출될 우려가 적어진다. 그럼에도 불구하고 여러 데이터 세트에 대한 결합,[84] 빅데이터를 활용한 데이터 마이닝,[85] 데이터 연결 기술 등을 통하여 개인 신원을 밝힐 수 있기 때문에 여전히 가명화 또는 익명화된 데이

80) 조선 Biz, "[파워금융人]⑥ 김선우 우리은행 AI사업부장 "생성형 AI 활용한 PB 곧 나온다… 재무관리도 AI가 척척"", 2023. 11. 19. 해외에서도 JPMorgan Chase, Bank of America, Goldman Sachs 등은 모두 직원의 ChatGPT 사용을 제한하고 있다.

81) Netskope Threat Labs, "Cloud and Threat Report: AI Apps in the Enterprise"(2023).

82) 개인정보의 익명화 처리를 제대로 하지 않은 AI에 대한 사례로는 조선 Biz, "AI 챗봇 '이루다', 개인정보유출 논란 속 결국 사실상 폐기", 2021. 1. 15.

83) 신용정보의 이용 및 보호에 관한 법률 제2조 제16호, 개인정보 보호법 제2조 제1호 다.목에서 규정.

84) 금융 결합데이터의 재사용 허용에 따른 유출 위험성을 다룬 논문으로 Dániel Kondor *et al.*, "Towards matching user mobility traces in large-scale datasets", *IEEE Transactions on Big Data*, Vol. 6(2020)이 있는데, 여기서는 익명의 신용카드 데이터가 공개적으로 이용 가능한 다른 데이터와 결합되어 개인을 매우 정확하게 식별할 수 있음을 보여 주었다.

85) 대용량의 데이터를 추출하여 거기서 숨겨져 있거나 알려지지 않는 패턴을 식별하는 것을 의미한다.

터에서 개인을 재식별할 수 있는 위험이 적지 않다고 보인다.

한편 생성형 AI에서 프롬프트를 입력함에 따른 정보 유출 문제도 적지 않은 위험성을 내재하고 있다. 생성형 AI를 이용할 때 이용자는 프롬프트에 해당하는 질문을 입력한다. 그 질문은 단순할 수도 있지만 매우 복잡하여 여러 페이지 길이가 될 수 있다. 그 과정에서 프롬프트에 고객정보가 들어갈 수 있고 이에 따른 외부 유출 위험은 무시할 수 없는 수준이라고 보인다.

3) 이러한 데이터 외부 유출 위험을 극복하기 위한 가장 유력한 방안으로 합성데이터의 사용이 거론된다. 합성데이터는 실제 데이터 세트인 원데이터를 사용하여 인위적으로 생성한 학습데이터이다. 합성데이터는 기존 데이터의 통계적 특성과 패턴을 그대로 재현하고 있으므로 실제 데이터인 원데이터를 사용하는 경우와 동일한 출력을 낼 수 있다. 앞서 본 익명정보나 가명정보의 사용은 원데이터가 존재하는 이상 개인정보 유출의 위험성이 존재하나 합성데이터에서는 개인정보 유출의 위험성이 훨씬 적어진다. 금융업계에서는 익명화된 정보의 재식별 위험 또는 실제 데이터에 대한 낮은 접근성 등을 극복하기 위하여 합성데이터를 사용하는 경우가 늘어나고 있다.

생성형 AI는 거시적인 경제 정보, 시장지표, 기업들의 회계정보를 모아 분석하고 금융시장의 흐름을 정확하게 인지할 수 있는 합성데이터를 생성할 수 있다. 이는 투자 결정, 위험 회피 전략 마련에 있어 중요한 도움을 줄 수 있다. 무엇보다 대량의 금융정보를 신속하게 판별하므로 빠르게 진화하는 금융환경에서 경쟁 우위를 확보할 수 있다.

그럼에도 불구하고 문제가 전혀 없는 것은 아니다. 합성데이터가 기존의 원데이터 세트를 제대로 대표하지 못하면 잘못된 금융상품을 설계하거나 오류가 있는 재무보고서를 생성할 수 있다. 이에 따라 금융투자자들이 투자손실을 입어 이들에게 막대한 손해배상을 해야 할 수 있고 금융감독당국으로부터 조사를 받을 수 있다. 합성데이터의 생성에 사용된 원데이터가 개인정보 침해물이나 저작권 침해물 또는 영업비밀에서 수집한 데이터 등을 기반으로 구축될 수도 있어 관련 법 위반 여지도 있다.

게다가 합성데이터를 역추론하여 원데이터를 드러내는 방식으로 개인의 데이터가 유출될 수 있는 방법이 있어 주목된다. 이를 '회원 추론 공격'(Membership Inference Attacks)이라고 한다.[86] 이는 합성데이터를 사용하더라도 원데이터에 포함

86) Ryan Webster *et al.*, "This P erson (Probably) Exists. Identity Membership Attacks Against GAN

되어 있는 특정 개인의 대출금 연체나 도산 여부 등의 민감한 정보가 드러나 외부로 유출될 위험이 있음을 시사한다.[87] 예를 들어, 대출과 관련하여 훈련 데이터에 사용된 구성원을 추론하는 데 성공한다면 공격자는 특정 개인의 대출금 연체 여부를 알게 될 것이고 개인을 대상으로 한 표적범죄에 노출시킬 수 있다. 또한 이로 인하여 유출된 정보로 차별적 의사결정에 악용될 수 있는데, 예를 들어 고용, 복지, 교육 등에 있어 그 대상에서 부당하게 제외되는 경우가 있을 수 있다.[88]

5. 불공정의 위험: 편향, XAI, 자동화된 의사결정

(1) 금융서비스 제공에 있어서의 편향 문제

금융업계에서는 다수 고객들의 대출, 신용카드 발급, 신용점수 평가, 로그인 및 결제서비스, 보험료 및 보험금 산정 등을 자동화 또는 간소하게 처리하기 위하여 생성형 AI를 더 많이 활용할 수 있다. 이러한 경우 AI 모델 편향 문제가 발생할 수 있다. AI 모델의 편향은 잘못된 결정을 하게 만드는 체계적인 오류를 의미하는 것으로, 데이터 편향, 알고리즘의 편향 등이 있다.

데이터 편향이란 AI 모델을 학습시키는 데 사용하는 데이터가 사람이나 사회가 가지는 편견을 포함하는 것을 의미한다. 데이터 편향의 유형으로 크게 데이터 선택 편향과 데이터 측정 편향으로 나뉘는데, 구체적으로는 의도적으로 오류가 있는 데이터나 조작된 데이터를 삽입한 경우, 특정 집단에 대한 데이터가 누락되거나 일정한 특성을 강조하여 수집되어[89] 데이터 자체에 불평등과 편견적 요소가 있는 경우 등이 있다. 예를 들어, 얼굴 인식 금융 결제 시스템에서 동북아시아인만을 과도하게

Generated Faces", arXiv:2107.06018v1 [cs.CV](2021)〈https://arxiv.org/abs/2107.06018〉.

87) 이에 대응하여 등장한 'Machine Learning Privacy Meter'는 AI가 회원 추론 공격에 얼마나 취약한지 분석하는 도구로 AI 모델에 대한 공격을 가정적으로 진행하여 유출되는 정보를 파악하고 그 대응책을 마련하는 것을 목적으로 한다[Sasi Kumar Murakonda, Reza Shokri, "ML Privacy Meter: Aiding Regulatory Compliance by Quantifying the Privacy Risks of Machine Learning", CoRR abs/2007.09339(2020)].

88) 그 밖에 데이터 외부 유출을 방지하기 위한 또 다른 방안으로, 훈련용 데이터 세트와 생성된 데이터에 대한 마스킹·토큰화·암호화하는 방안, 데이터 세트에 잡음(Noise)을 추가하는 방안, AI 플랫폼에 공유되는 데이터 양을 최소화하는 방안 등이 제안되고 있다(Forbes June 22th, 2023 article, "How Companies Can Use Generative AI And Maintain Data Privacy").

89) 특정한 문화, 집단에서 가져온 데이터로 훈련하는 경우 그에 따른 편견과 사상, 생각 등이 반영될 수 있다.

대표하는 훈련데이터를 사용하여 백인, 흑인이나 동남아시아인이 결제를 시도할 때 오류를 일으킬 수 있다.[90] 한편 훈련데이터가 금융시장 상황의 변화, 새로운 금융상품의 등장 등으로 인하여 미래에는 유용하지 않거나 무의미할 수 있어 이 또한 편향을 유발할 수 있는데 이를 데이터 드리프트(Data Drift)라고 한다.

알고리즘 편향이란 개발자가 고의로 또는 무의식적으로 자신의 사상이나 선호를 AI 알고리즘에 반영하여 특정 집단에 체계적으로 불리한 결정을 내리는 것을 말한다. 알고리즘 편향에 대하여 이를 직접적으로 다루는 법률이나 규정은 없으나, AI 운영 가이드라인 제5항 나.목에서는 알고리즘 편향을 피하기 위한 목적 등으로 금융회사들로 하여금 공정성 판단 지표를 마련하면서 고객에 대한 잠재적인 피해를 최소화하도록 하고 있다.

이와 같은 AI 모델의 편향은 대출, 신용카드 발급 등에 있어서 불공정하고 차별적인 결과를 일으킬 수 있고 이는 금융기관들에게 손해배상 및 평판 위험이 생길 여지가 있다. 특히 대출에 있어서 적격 차용인에 대한 대출이 거부되거나 부적격 차용인에 대한 대출이 승인될 수 있다. 예를 들어, 4~50대 남성의 신용점수가 높은 경우가 많다는 데이터를 학습한 AI가 신용점수 측정 및 대출에 있어서 20대 남성이나 여성에 대한 대출을 거부하거나 대출을 하더라도 이들에게 부당하게 높은 대출금리를 적용할 수 있다.[91]

따라서 AI의 알고리즘에 있어 오류 가능성이 전혀 없는 것이 아니기 때문에 고객을 잘못 분류할 경우 인간이 수정할 수 있도록 추가적인 조치가 필요하다.[92] 또한 사후적으로도 AI 활용에 있어 차별, 낙인효과가 발생하였는지 면밀하게 측정하고[93] 감독할 수 있어야 한다.[94] 그와 별개로 투명성 조치와 윤리적 데이터 수집은 데이

90) Daniel McDuff *et al.*, "Characterizing Bias in Classifiers using Generative Models", *Advances in Neural Information Processing Systems* 32(2019)의 연구 결과 참조, 위 논문은 생성적 적대 신경망을 활용하였다.

91) 보험의 경우도 마찬가지 위험이 있을 수 있는데, 예를 들어 특정 직업을 가진 자, 특정 인종에 대하여 짧은 기대수명을 가진다는 점을 학습한 AI가 소수 집단에 보험상품을 판매하는 것을 거부하는 결정을 할 수 있다. 이와 관련하여 Alexander Pomerantz *et al.*, "Racial and Ethnic Differences in Insurer Classification of Nonemergent Pediatric Emergency Department Visits"(2023)은 흑인 및 히스패닉계 아동의 응급실 방문에 대하여 AI가 백인 아동에 비하여 비응급으로 식별하여, 보험환급금 삭감이 이루어질 가능성이 높다고 밝히고 있다.

92) Alžběta Krausová, "Intersections between Law and Artificial Intelligence", *International Journal of Computer* 27(1)(2017), 무엇보다 AI 시스템에서 '공정성'의 내용과 기준을 학습시켜 인종, 성별, 연령, 지역, 종교, 교육 수준 등 다양한 분야를 고려하도록 하여야 한다.

93) 예를 들어, AI의 편향을 탐지할 수 있는 소프트웨어로 AI Fairness 360과 IBM Fairness 360 등이 있다.

94) AI 운영 가이드라인에서는 개인에 대한 부당한 차별 등 개인의 권익과 안전 등에 중대한 위험을 초

터 편향 문제에 대한 해결책으로 등장하고 있다. 투명성 조치는 데이터 수집 및 처리 과정을 공개하여 출력 결과에 대한 신뢰를 높이는 방법이다. 그리고 윤리적 데이터 수집은 특정 집단에 차별적이거나 민감한 영향을 미칠 수 있는 데이터를 처리할 때 윤리적인 원칙을 준수하는 것을 말한다.

한편 AI 훈련 및 데이터 처리, 생성, 알고리즘 생성, 작용 등 모든 절차가 잠재적으로 편견의 영향을 받을 수 있으므로 해당 절차를 상세하게 기록하여야 한다. 그러나 생성형 AI와 같이 복잡한 알고리즘을 사용하면 그 기록을 수집하더라도 이를 이해하는 것이 불가능할 수 있다. 따라서 뒤에서 보는 설명가능한 AI(XAI) 기술 등을 적용하여 맥락에 맞는 설명이 도출될 수 있도록 조치할 필요가 있다.

(2) 의사결정 이유의 설명 문제

AI 알고리즘 결과를 그대로 따를 경우뿐만 아니라 이를 단순히 참고만 한 경우에도 고객들에게 특정 의사결정을 내린 이유를 설명할 필요가 있다. 그러나 생성형 AI에는 다수의 매개변수가 존재하고 이러한 매개변수들이 개별적으로 AI 기능을 조정할 뿐만 아니라 수십, 수백 개가 이어진 입출력 정보가 결과에 미치면서 내부적 의사결정이 최종적인 조치에 어떻게 기여하는지 이해하기 난해하다.[95] 따라서 예를 들어 대출 심사 과정에서의 대출 승인 거부, 대출 승인시 금리 결정 등이 자동화된 의사결정으로 이루어진 경우 그 이유를 고객에게 설명하는 것이 매우 어려울 수 있다. 그로 인하여 불이익을 받는 고객들이 이의를 제기할 수 있는 기회를 상실시킬 수 있을 뿐만 아니라 금융기관의 조치에 대한 절차적 공정성과 객관성을 훼손시킬 수 있다.

이와 관련하여 금융서비스 분야에서는 의사결정 과정을 명확하게 설명해 주는 AI를 뜻하는 '설명가능한 AI'(Explainable AI, 이하 'XAI')"가 주목받고 있다.[96] XAI는 AI 프로세스 뒤에 있는 작성자의 의도, AI가 사용하는 모델 계열, AI를 교육하기 전에 작성자가 미리 입력한 매개변수, 작성자가 교육하는 데 사용한 훈련데이터 세트, 출력 데이터의 오류 가능성 등에 대한 설명을 포함할 수 있다.[97] XAI는 고객에게

래할 수 있는 서비스를 '고위험 서비스'라고 지칭하면서 이에 대하여 AI 시스템을 활용하는 경우, 적절한 내부통제 활동 및 승인절차를 마련하고 승인 책임자를 지정하도록 규율하고 있다(제2항 라.목).

95) 이러한 이유로 AI를 '블랙박스'라고 부른다.

96) 이에 대한 설명으로 David Gunning *et al.*, "XAI—Explainable artificial intelligence", *Science Robotics,* 4(37)(2019) 참조.

97) Tanusree De *et al.*, "Explainable AI: A Hybrid Approach to Generate Human-Interpretable Explanation for Deep Learning Prediction", *Procedia Computer Science*, Vol. 168(2020)은 신용카드 채무불이행 가

금융기관의 서비스에 대한 신뢰를 구축하고 불공정하거나 편향된 결과가 나타날 위험을 줄이는 데 중요한 역할을 할 수 있다. 그뿐만 아니라 나중에 법적 분쟁이 발생할 경우 재판부는 금융기관에 XAI를 활용한 설명을 요구할 수 있고 그 설명을 기초로 금융기관의 결정이 사실에 기반하였는지 여부, 재량권을 남용하였는지 여부 등에 대하여 판단할 수 있다.[98]

다만 XAI와 관련하여 제도적인 개선이 필요한데, 먼저 사용자들이 그 설명을 쉽고 간명하게 인식할 수 있도록 하여야 한다. 그러나 생성형 AI 등은 기술적, 기능적으로 복잡한 구조를 가지고 있어 이와 상충되는 문제가 있다. 따라서 어떠한 표현 방식을 구현하고 내용을 공개할지와 관련하여, 그 표준적 정의와 방식을 정하는 것이 필요하다. 또한 일정한 현상이나 결과에 대한 설명을 시간과 장소에 구애받지 않고 일관적으로 표현할 수 있는 성능 요건을 제시하고 그 감독체계를 구현할 필요가 있다. 그리고 XAI를 사용하는 과정에서 개인정보나 영업비밀 등이 유출될 수 있으므로 이를 방지하기 위한 제도적 보완이 필요하다.

(3) 자동화된 의사결정의 문제

자동화된 의사결정(Automated Decision Making)이란 사람의 개입 없이 자동화된 수단으로 의사결정을 내리는 것을 말한다. 금융 분야는 AI에 의한 의사결정의 자동화가 이루어지기 쉬운 대표적인 분야로,[99] 대출이나 신용카드 발급의 승인 또는 거부, 신용한도의 설정, 대출이자율의 결정, 신용점수 산정, 보험료 및 보험금의 산정 등에서 활용할 수 있다.

생성형 AI는 자동화된 의사결정에 있어서 탁월한 기능을 발휘할 수 있다. 생성

능성이 높은 고객을 식별하기 위하여 성별, 교육 수준, 결혼 여부, 연령 등이 나열된 30,000명의 인구통계 데이터와 특정 6개월 동안의 신용카드 청구금액 데이터 및 6개월 간의 결제금액 데이터로 AI를 훈련시킨 다음, Cluster-TREPAN 의사결정트리를 이용하여 사람에게 설명할 수 있는 의사결정 논리구조를 만들려고 하였다. 그중 일부 논리구조를 각색하면 '25세 이상 60세 미만인지'[pd_25≦ age〈 60 : 비해당(false)] → '전체 6개월 동안 1회 이상 연체되었는지 여부'[pd_a6〉1 : 해당(true)] → '처음 3개월보다 지난 3개월 동안 연체 횟수가 1.25배 많은지 여부'[ratio_pd_l3_f3〉1.25 : 비해당(false)] → '연체확률 낮다'(Node2 : 0) 정도로 소개할 수 있다. 이에 따라 위 논문은 심층신경망 내의 정보 흐름을 시각화함으로써 신용카드 채무불이행 예측 결과에 대하여 사람이 이해할 수 있는 수준의 설명(또는 코드)을 생성하려 하였다.

98) Ashley Deeks, "The Judicial Demand for Explainable Artificial Intelligence", Virginia Public Law and Legal Theory Research Paper No. 2019-51(2019), pp.1839-1840.

99) Juliana Hadjitchoneva, "Efficient Automation of Decision-making Processes in Financial Industry: Case Study and Generalised Model"(2019), p.2.

형 AI는 통상의 AI보다 방대한 양의 데이터, 예를 들어 대출신청인의 휴대전화, 소셜네트워크, 인터넷 검색, 쇼핑 기록,[100] 소비 습관, 생활스타일, 위치정보 등을 수집하여 직원의 개입 없이 자동으로 대출 승인 및 대출이자율 결정 등을 할 수 있다.[101] 생성형 AI를 활용하여 의사결정의 자동화를 달성함으로써 금융 분야에서는 업무의 효율성이 증진되고 고객의 입장에서도 빠르게 결정을 받을 수 있다.

그러나 자동화된 의사결정은 비록 AI 알고리즘에 따라 사람의 편견이나 부패의 개입 여지 없이 이루어진다고 하더라도[102] 그 자체로 다음과 같은 잠재적인 위험이 있어 완전히 신뢰하기는 어렵다. ① 앞에서 살펴본 데이터 및 알고리즘 편향의 문제는 자동화된 의사결정에 대하여 정확성을 담보하지 않고 오히려 불공정하거나 차별적인 결과를 낼 수 있음을 시사한다.[103] ② 사람들은 위 의사결정의 내부적 절차가 어떠한지 알 수 없어 절차적 공정성이 침해될 여지도 있다. ③ 나아가 고객이 프로파일링에 활용되는 자신의 데이터에 대하여 정보 제공 및 사용 동의를 하지 않았거나 그 수집 및 처리 방법을 예상하지 못하였을 수 있다.

이에 따라 EU의 GDPR 제22조는 프로파일링을 포함한 정보주체에게 중요한 권리·의무에 영향을 미치는 자동화된 의사결정 자체를 금지하고 있고 계약 체결 또는 이행을 위하여 필요한 경우, 정보주체의 동의가 있는 경우 등에 한하여 예외적으로 허용하고 있다. 우리나라의 개인정보 보호법 제38조의2에서는 자동화된 의사결정을 완전히 금지하지는 않고 정보주체에게 자신의 권리 또는 의무에 중대한 영향을 미치는 경우에 한하여 이를 거부하고 이에 대한 설명, 인적 개입에 의한 재처리를 요구할 수 있는 권리를 부여하였다.

앞서 본 대출이나 신용카드 발급, 보험계약 체결의 승인 또는 거부, 신용한도의

---

100) 예를 들면 신용카드 발급을 신청한 자의 인터넷 검색 데이터와 쇼핑 기록 등을 수집하고 신청인의 소비 패턴을 파악하여 신용카드 채무불이행의 위험 정도를 산정함으로써 신용카드 발급 여부를 결정할 수 있을 것이다.

101) 이를 프로파일링(Profiling)이라고 하는데, EU의 GDPR 제4조 제4항은 이에 대하여 "특히 자연인의 업무 성과, 경제적 상황, 건강, 개인적 선호, 관심사, 신뢰도, 행동 등을 분석하거나 예측하기 위하여 행해지는 경우로, 자연인과 관련된 개인적인 특정 측면을 평가하기 위하여 개인정보를 사용하여 이루어지는 자동화된 개인정보의 처리를 의미한다"고 규정하고 있다.

102) Andrew Kumiega, Ben Van Vliet, "Automated Finance: The Assumptions and Behavioral Aspects of Algorithmic Trading", *Journal of Behavioral Finance* 13(1)(2012).

103) 예를 들어, 건설회사 직원이 근무 중 사고를 당할 확률이 평균보다 더 많다는 것을 발견한 AI가 건설회사에서 행정직으로 근무하는 직원에게 더 높은 상해보험료를 지불하도록 하는 경우, 직업의 문제라기보다는 보직의 문제일 수 있기 때문에 다른 회사의 행정직 직원들보다 보험료를 더 높게 받는 것이 불합리한 차별이 될 여지가 있다.

설정, 신용점수 산정[104]은 오류가 있다면 그 신청인에게 심각한 부작용을 초래할 수 있으므로 위 법률 등에서 규율하는 '정보주체에게 중요한 권리·의무에 영향을 미치는 경우'에 해당한다고 볼 수 있다.[105] [106] 이와 같은 경우 최소한 AI의 결론은 참고만 하고 최종적으로 인간이 결정하도록 인적 개입을 규정할 필요가 있다. 나아가 AI의 결론을 참조하여 최종적으로 인간이 결정한 경우 투명성 조치 또는 XAI와 같은 기술적 조치에 따라 AI가 특정한 결과를 도출한 이유와 그 기술적 메커니즘에 대하여 고객에게 충분히 설명할 수 있어야 할 것이다.[107]

6. 기타 쟁점

(1) 고객 상담시 투명성 위험

고객들은 생성형 AI를 사용하는 대화에이전트나 챗봇과 대화하면서 실제 인간과 구별하기 어려울 정도로 인간 상담원의 응답과 유사한 대답을 받을 수 있다. 이에 따라 고객들은 자신들이 인간과 대화하고 있다고 오해할 수 있다. 고객이 받은 이메일 또한 생성형 AI에 의하여 작성된 것임에도 인간이 작성하였다고 생각할 수도 있다.[108]

---

104) EU 사법재판소(Court of Justice of the European Union: CJEU)는 2023. 12. 7. OQ v. Land Hessen & SCHUFA Holding 사건(Case C-634/21)에서 독일의 신용평가기관(피고)이 개인(원고)의 신용점수를 산정한 행위에 대하여 자동화된 의사결정을 원칙적으로 금지하는 GDPR 제22조가 적용된다고 설시하였다. 이에 대하여 피고는 자신이 산정한 신용점수를 토대로 은행이 최종적으로 대출 신청을 거부한 것이고 피고가 대출 신청을 거부한 것은 아니므로 신용점수 산정 자체로는 데이터 주체에 중요한 영향을 미치지 않는다고 주장하였으나, 위 재판소는 신용점수 산정이 대출 승인 여부에 중요한 역할을 하였을 뿐만 아니라 은행은 신용점수 생성에 사용된 자동화된 절차에 대한 세부적인 정보를 가지고 있지 않아 신용평가기관이 GDPR 15(1)(h)조에 따른 정보 의무를 준수할 수 있는 더 나은 위치에 있다는 이유로 피고의 주장을 배척하였다.

105) 다만, 기계적 시스템의 도움이 필수적인 반복적이고 구조화된 작업의 경우, 예를 들어 신용카드 한도와 신용 한도의 미세 조정 등에 있어서는 완전 자동화가 필요하다고 보이고, 다만 그 내용은 휴대전화 어플리케이션이나 이메일 등을 통하여 고객에게 통지하여 이의를 제기하거나 설명을 구할 권한을 부여하여야 할 것이다.

106) 대출이자율, 보험료 및 보험금도 그 액수에 따라 이에 해당할 여지가 있을 것이다.

107) Max Schemmer, Niklas Kühl & Gerhard Satzger, "Intelligent Decision Assistance Versus Automated Decision-Making: Enhancing Knowledge Work Through Explainable Artificial Intelligence", Proceedings of the 55th Hawaii International Conference on System Sciences(2022)에서는 AI가 도출한 결론을 참고만 하고 사람이 최종적으로 결정하는 경우를 '지능형 의사결정 지원'(Intelligent Decision Assistance: IDA)이라고 명명하면서, 이를 인간의 개입이 없는 완전히 자동화된 의사결정의 단점을 메우면서 데이터에 대한 통찰력을 제공할 수 있는 좋은 대안으로 보았고, 대신에 XAI와 같은 기술적 조치가 필요함을 논증하였다.

108) Government of Canada, "Guide on the use of Generative AI".

아직까지 관련 기술이 크게 발전하지는 않았지만 AI 상담원과의 전화 연결 또는 줌(Zoom)과 같은 화상회의 서비스를 통하여 보고 듣는 화상, 음성 또한 마찬가지이다. 이 경우 고객들이 혼란을 겪지 않고 금융시장에 대한 신뢰를 유지할 수 있도록 투명성 조치를 하는 것이 필수적이라고 생각한다. 여기서의 투명성 조치는 고객들이 생성형 AI 시스템을 사용하는 대화에이전트 등과 상호작용하고 있는 것을 알 수 있도록 알리는 것을 말한다. 예를 들어 AI가 생성한 이메일에 대하여 고객들이 식별할 수 있도록 워터마크를 사용하거나 위 생성형 AI의 작동 방식, 사용 이유, 위 AI에 의하여 생성된 내용의 품질 보증에 대하여 고객들이 이해하기 쉬운 방식으로 게시할 수 있을 것이다.

(2) 반경쟁 위험

금융 영역에 있어서 생성형 AI의 사용은 경쟁 관련 문제도 낳을 수 있다. 복수의 금융기관들이 동일한 알고리즘을 운영하는 AI 모델을 사용하는 경우 동일한 사안에 대하여 유사한 의사결정을 생성할 수 있다.[109] 이는 경쟁 관련 문제와 연관될 수 있는데, 예를 들어 어느 대출신청인에 대하여 금융기관들이 모두 동일한 비율로 연 금리를 과도하게 높게 유지하여 대출신청인이 선택의 여지 없이 같은 금리로 대출을 받아야 할 수 있다. 보험의 경우에도 마찬가지로, 예를 들어 교통사고 발생 빈도가 2년에 3차례에 이른다고 하여 보험회사들이 같은 비율로 보험료를 할증하거나 자동차보험 가입을 거부할 수도 있을 것이다. 이는 경쟁제한성의 문제를 초래하여 공정거래법상 문제로 비화될 수도 있다. 다만 현행법상으로는 명시적으로 담합을 한 것이 아니거나 실제 의사소통이 있는지 증명하기 어렵기 때문에 규제당국에서 부당한 공동행위로 규율하기 어렵다고 보인다.[110] 그럼에도 위와 같은 경우 금융소비자들에게 선택의 제한을 초래하고 공정한 경쟁을 저해하는 결과를 나오게 할 수 있어[111]

---

109) 이를 '알고리즘의 묵시적 담합'(Algorithmic tacit collusion)이라고 한다[Alessio Azzutti, WolfGeorg Ringe & H. Siegfried Stiehl, "Machine Learning, Market Manipulation, and Collusion on Capital Markets: Why the "Black Box" Matters", 43 *U. Pa. J. Int'l L.* 79(2021), p.110].

110) 주진열, "AI 알고리즘 가격설정과 이른바 '알고리즘 묵시적 담합' 문제에 대한 고찰", 『경쟁법연구』, 제41권, 한국경쟁법학회(2020), 332-368쪽.

111) 금융기관들 사이의 직·간접적인 의사소통이 없더라도 금융상품의 가격이나 이율이 고정되어 공동행위가 존재한다는 공통적인 인식이 있다고 볼 수 있는 경우가 있을 수 있다. 예를 들어 다수의 금융기관이 동일한 업체가 제공하는 AI 알고리즘을 사용하고 있고 이를 금융기관들이 인식하는 경우를 들 수 있다. 이와 별개로 생성형 AI가 경쟁 금융기관의 경영전략을 습득하여 모방하는 경우 마치 담합과 같은 효과가 발생할 가능성이 있는데 이에 따른 경쟁제한성 문제를 어떻게 해결할지 문

이에 대한 연구가 필요하다고 보인다.[112]

Ⅴ. 결 론

이 글에서는 금융 분야에서의 생성형 AI 사용 현황을 살펴보고 이로써 발생할 수 있는 법적 쟁점과 과제에 대하여 논증해 보았다. 특히 법적 쟁점에 관하여 딥페이크 등 조작 위험, 금융 안정성의 위험, 투자자 보호 관련 위험, 데이터 유출 위험, 불공정의 위험으로 나누어 살펴보았다. 나아가 간단하게 고객 상담시 투명성 위험, 반경쟁 위험에 관하여도 짚어보았다.

마지막으로 여기서 살펴볼 점은 금융 영역에서 생성형 AI를 비롯한 AI의 규제 방향을 어떻게 정할지에 관한 것이다. 앞서 본 EU의 AI법은 용도별 위험성의 정도에 따라 AI를 분류하여 고위험 AI에 규제를 집중하도록 하였다. 이는 규제력의 낭비를 예방하고 효율성을 추구한다는 점에서 타당하다고 보인다. 이는 우리나라 국회에 발의된 AI 법안들이 취하고 있는 기본 규제 방향이기도 하다. 또한 EU AI법에서 생성형 AI를 비롯한 범용 AI를 따로 규율하면서 기본적으로 규제를 최소화하되 고위험 용도로 사용될 가능성도 염두에 두는 점도 참고할 만하다. 다만, 위 법은 고위험 AI에 해당하는 분야에 금융 영역을 명시하지 않아 금융 용도 AI를 제한적 위험이 있는 AI로 판단하는 것으로 보이나 세부적인 영역, 예를 들어 대출, 신용평가, 보험계약 등에 있어 개인의 평가수치가 왜곡되면서 그 생활에 심각한 영향을 줄 수 있는 경우도 있어서 이 경우에는 고위험 분야로 규제하는 것이 바람직해 보인다.

나머지 금융 영역에 있어서는 AI의 활용을 최대한 자유롭게 하고 사후적으로 규제하는 방식을 취하는 것이 타당해 보인다. 특히 금융업계에서는 비용 절감을 위

제될 수도 있다(Alessio Azzutti, WolfGeorg Ringe, H. Siegfried Stiehl, *supra* note 109, p.109).

112) 이와 관련하여 EU 의회는 2017. 2. AI 알고리즘에 대하여 '전자인격'(Electronic persons)을 도입하는 것에 대한 영향 평가와 법적 분석을 하라는 취지의 결의안을 채택하였다. 전자인격을 인정한다면 AI 자체에 대하여 부당한 공동행위의 수범자로 인정할 수 있으나, 현재의 약한 AI로는 단순 알고리즘에 불과하여 인간과 같은 정도의 인지능력이 결여되어 있을 뿐만 아니라 법인과 같은 또 다른 인격을 인정하는 것은 기존 법체계와 조화되기 어렵다는 점에서 부정하는 것이 바람직해 보인다[이영철, "인공지능 알고리즘에 의한 불공정 거래행위의 법적 규제 - 공정거래법상 부당한 공동행위를 중심으로 -", 『상사법연구』, 통권 제110호(2021), 231-232쪽].

한 목적으로 생성형 AI를 개발하여 이를 활용하려 하는 경우가 많은데, 새롭게 규제를 가함으로써 발생하는 비용이 AI를 사용함으로 인하여 절감되는 비용을 초과한다면 금융 AI 산업의 위축을 초래하게 될 것이다. 따라서 규제를 최소화하면서 유연하고 시장을 존중하는 방향으로 법적 규율이 이루어지는 것이 필요하다고 보인다.

■ 참고문헌

이영철, "인공지능 알고리즘에 의한 불공정 거래행위의 법적 규제 - 공정거래법상 부당한 공동행위를 중심으로 -", 『상사법연구』, 통권 제110호(2021).

주진열, "AI 알고리즘 가격설정과 이른바 '알고리즘 묵시적 담합' 문제에 대한 고찰", 『경쟁법연구』, 제41권, 한국경쟁법학회(2020).

채명식·조유리, "2023 인공지능", KISTEP 한국과학기술기획평가원(2023).

금융위원회, "금융분야 인공지능의 신뢰를 높인다", 보도자료, 2023. 4. 17.

조선 Biz, "AI 챗봇 '이루다', 개인정보유출 논란 속 결국 사실상 폐기", 2021. 1. 15.

_____, "[파워금융人]⑥ 김선우 우리은행 AI사업부장 "생성형 AI 활용한 PB 곧 나온다… 재무관리도 AI가 척척"", 2023. 11. 19.

Alessio Azzutti, Wolf Georg Ringe & H. Siegfried Stiehl, "Machine Learning, Market Manipulation, and Collusion on Capital Markets: Why the "Black Box" Matters", 43 *U. Pa. J. Int'l L.* 79(2021).

Alexander Pomerantz *et al.*, "Racial and Ethnic Differences in Insurer Classification of Nonemergent Pediatric Emergency Department Visits"(2023).

Alžběta Krausová, "Intersections between Law and Artificial Intelligence", *International Journal of Computer* 27(1)(2017).

Andrew Kumiega, Ben Van Vliet, "Automated Finance: The Assumptions and Behavioral Aspects of Algorithmic Trading", *Journal of Behavioral Finance* 13(1)(2012)

Antonia Creswell *et al.*, "Generative Adversarial Networks: An Overview", *IEEE Signal Process Magazine* 35(1)(2018).

Ashish Vaswani *et al.*, "Attention Is All You Need", *Neural Information Processing Systems*(2017).

Ashley Deeks, "The Judicial Demand for Explainable Artificial Intelligence", Virginia Public Law and Legal Theory Research Paper No. 2019-51(2019).

Binh Le *et al.*, "Why Do Facial Deepfake Detectors Fail?", WDC '23: Proceedings of the 2nd Workshop on Security Implications of Deepfakes and Cheapfakes(2023).

Caosen Xu *et al.*, "A Financial Time-Series Prediction Model Based on Multiplex Attention and Linear Transformer Structure", *Appl. Sci.* 13(8)(2023).

Dániel Kondor *et al.*, "Towards matching user mobility traces in large-scale datasets", *IEEE Transactions on Big Data*, Vol. 6(2020).

Daniel McDuff *et al.*, "Characterizing Bias in Classifiers using Generative Models", *Advances in Neural Information Processing Systems* 32(2019).

David Gunning *et al.*, "XAI—Explainable artificial intelligence", *Science Robotics,* 4(37)(2019).

David S. Krause, "Mitigating Risks for Financial Firms Using Generative AI Tools"(2023).

Edmond Lezmi, Jiali Xu, "Time Series Forecasting with Transformer Models and Application to Asset Management"(2023).

Emilija Strelcenia, Simant Prakoonwit, "A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection", *Mach. Learn. Knowl. Extr.* 5(1)(2023).

Enrique Martínez-Miranda *et al.*, "Learning Unfair Trading: A Market Manipulation Analysis from the Reinforcement Learning Perspective", 2016 IEEE Conference on evolving and adaptive intelligent SYS(2016).

Erik Brynjolfsson, Danielle Li & Lindsey R. Raymond, "Generative AI at work", National bureau of economic research(2023).

Government of Canada, "Guide on the use of Generative AI".

Jinhong Wu, *et al.*, "Interpretation for Variational Autoencoder Used to Generate Financial Synthetic Tabular Data", *Algorithms* 16(2)(2023).

Joerg Osterrieder, "A Primer on Artificial Intelligence and Machine Learning for the Financial Services Industry"(2023).

Juliana Hadjitchoneva, "Efficient Automation of Decision-making Processes in Financial Industry: Case Study and Generalised Model"(2019),

Kelvin J.L. Koa *et al.*, "Diffusion variational autoencoder for tackling stochasticity in multi-step regression stock price prediction", Proceedings of the 32nd ACM international conference on information and knowledge management(2023).

Mahsa Tavakolia *et al.*, "Multi-Modal Deep Learning for Credit Rating Prediction Using Text and Numerical Data Streams"(2023).

MarketResearc report, July 2023, "Global Generative AI In Financial Services Market By Type (Solutions And Services), By Application (Credit Scoring, Fraud Detection, And Other), By Deployment Mode (Cloud And On-Premises), By Region And Companies - Industry Segment Outlook, Market Assessment, Competition Scenario, Trends, And Forecast 2023-2032".

Max Schemmer, Niklas Kühl & Gerhard Satzger, "Intelligent Decision Assistance Versus Automated Decision-Making: Enhancing Knowledge Work Through Explainable Artificial Intelligence", Proceedings of the 55th Hawaii International Conference on System Sciences(2022).

National Secuirty Agency *et al.*, "Contextualizing Deepfake Threats to Organizations"(2023).

Netskope Threat Labs, "Cloud and Threat Report: AI Apps in the Enterprise"(2023).

OECD, "Generative artificial intelligence in finance"(2023).

Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, "Learning representations by back-propagating errors", *Nature*, Vol. 323(1986).

Ryan Webster *et al.*, "This Person (Probably) Exists. Identity Membership Attacks Against GAN Generated Faces", arXiv:2107.06018v1 [cs.CV](2021).

Sasi Kumar Murakonda, Reza Shokri, "ML Privacy Meter: Aiding Regulatory Compliance by Quantifying the Privacy Risks of Machine Learning", CoRR abs/2007.09339(2020).

Sepp Hochreiter, Jürgen Schmidhuber, "Long Short-Term Memory", *Neural Computation*, Vol. 9(1997).

Siti Aishah Binti Mohd Yusof, Fatin Aqilah Binti Mohamad Roslan, "The Impact of Generative AI in Enhancing Credit Risk Modeling and Decision-Making in Banking Institutions", *Emerging Trends in Machine Intelligence and Big Data*, Vol. 15(2023).

Suman Kalia, "Potential impact of generative artificial intelligence(AI) on the financial industry", *IJCI Journal*, Vol. 12(2023).

Tanusree De *et al.*, "Explainable AI: A Hybrid Approach to Generate Human-Interpretable Explanation for Deep Learning Prediction", *Procedia Computer Science*, Vol. 168(2020).

TC Helmus, "Artificial intelligence, deepfakes, and disinformation: A primer", Rand(2022).

Tuna Tuncer *et al.*, "Asset Price and Direction Prediction via Deep 2D Transformer and Convolutional Neural Networks", Proceedings of the Third ACM International Conference on AI in Finance(November 2022).

Yu-Fu Chen, Szu-Hao Huang, "Sentiment-influenced trading system based on multimodal deep reinforcement learning", *Applied Soft Computing* 112(4)(2021).

Ziyue Wang *et al.*, "CNN-and GAN-based classification of malicious code families: A code visualization approach", *International Journal of Intelligent Systems*, Vol. 37(2022).

American Banker August 25th, 2023 article, "How banks can adopt generative AI".

CNBC May 10th, 2023 article, "Google will label fake images created with its A.I."

CNN May 23th, 2023 article, "'Verified' Twitter accounts share fake image of 'explosion' near Pentagon, causing confusion".

Forbes June 22th, 2023 article, "How Companies Can Use Generative AI And Maintain Data Privacy".

Techcircle May 2nd, 2023 article, "Nearly half of Indian internet users faced AI-driven voice scams this year".

The New York Times Aug. 30th, 2023 article "Voice Deepfakes Are Coming for Your Bank Balance".

The Wall Street Journal pro Cybersecurity August 30th,, 2019 article, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case".

鹿島 みかり・千葉 誠・関口 健太, "投資判断におけるアルゴリズム・AI の利用と法的責任: アルゴリズム・AIの利用を巡る法律問題研究会報告書の概要", 金融研究, 日本銀行金融研究所(2019).

The Korean Journal of Securities Law, Vol. 25, No. 1, 2024

A Study on Current Utilization and Legal Issues of Generative Artificial Intelligence in Finance

Lee, Sukjoon

ABSTRACT

Since the release of GPT-3, generative artificial intelligence (hereinafter referred to as 'AI') has played a significant role across various industries worldwide, and the market size is predicted to grow substantially in the future. In the financial sector, generative AI is also bringing about positive changes, contributing significantly to tasks such as customer consultation, customer investment management, risk management including loans, and overall business improvement for financial companies.

Generative AI includes Generative Adversarial Networks (GAN), Transformer models, and Multi-modal AI, each having its own strengths and weaknesses. These AIs excel in tasks such as generating synthetic data, distinguishing fraudulent transactions, non-face-to-face banking, predicting financial flows, sentiment analysis, financial analysis, and more.

However, despite the positive impacts, there are anticipated legal issues in utilizing generative AI in the financial sector. These include manipulation risks (deepfakes), risks to financial stability (flash crashes, financial system risks), investor protection-related risks, data leakage risks, unfairness risks (data and algorithm bias, explainable AI, automated decision-making), among others. This discussion focuses on these points, examining the key issues and legal and technical solutions. Additionally, transparency risks during customer consultations and risks related to anti-competition are also touched upon.

Key Words: Finance, Generative Artificial Intelligence, Deepfakes, Automated Decision Making, Bias, Explainable AI, Data Leakage, Chatbots, Robo-advisors, Generative Adversarial Network(GAN), Transformer Model, Multi-modal AI.