5장

생성형AI로 증폭되는 은행업계 딥페이크와 금융사기 위험

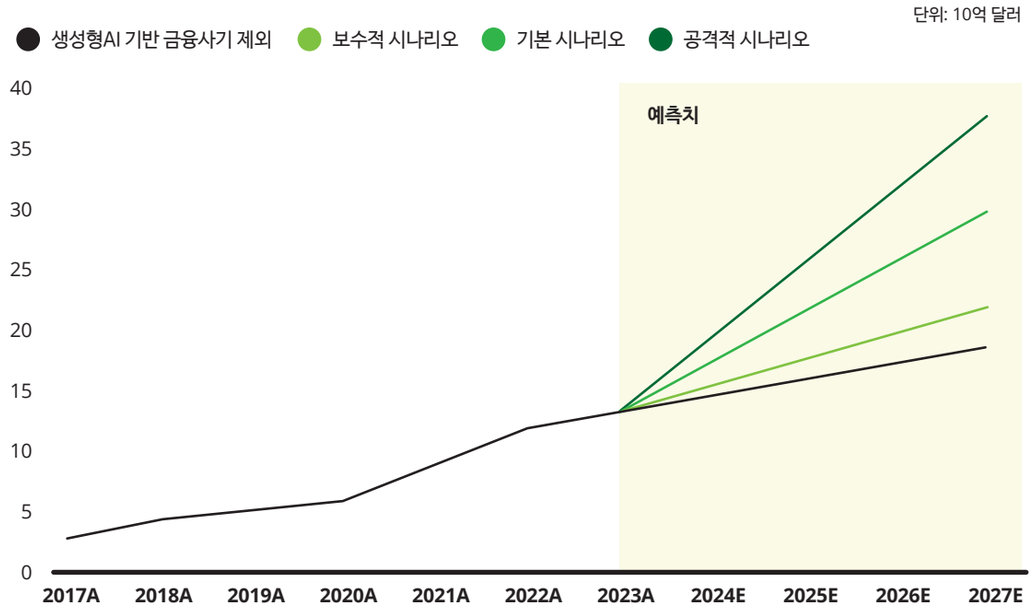
가짜 콘텐츠를 과거 어느 때보다 손쉽게 조작할 수 있게 됐지만, 이를 잡아내는 일은 과거 어느 때보다 어려워졌다. 은행들은 금융사기 위험이 증폭하는 환경에서 사기행위를 포착하고 손실을 예방하기 위해 인공지능(AI) 등 첨단기술에 공격적으로 투자해야 한다.

2024년 1월 홍콩 소재 한 기업의 직원이 사기꾼들에게 2,500만 달러를 송금한 사건이 있었다. 직원은 본인과 동료 직원이 참석한 화상회의에서 최고재무책임자(CFO)가 내린 지시대로 송금을 한 것이지만, 알고 봤더니 동료 직원과 CFO 모두 화상회의에는 참석한 적도 없었다. 사기꾼들이 이들을 모방한 딥페이크(deepfake, AI 기반 인간 이미지 합성 기술)를 만들어 송금하라는 지시를 내린 것이다.[1]

생성형AI(generative AI) 툴이 갈수록 정교해지고 사용이 용이해지는 만큼, 향후 수 년간 은행과 고객을 상대로 한 이러한 사기행위가 횡행할 것으로 우려된다.[2] 딜로이트는 미국 내 생성형AI 기반 금융사기 피해액이 2023년 미화 123억 달러에서 2027년 400억 달러로 연평균 32% 증가할 것으로 전망한다(그림 1).

1. Heather Chen and Kathleen Magramo, “Finance worker pays out $25 million after video call with deepfake ‘chief financial officer,” CNN, February 4, 2024.
2. Jon Bateman, Deepfakes and synthetic media in the financial system: Assessing threat scenarios, Carnegie Endowment for International Peace, July 8, 2020.

그림 1. 미국 내 생성형AI 기반 금융사기 피해액 전망



* ‘A’는 추정치, ‘E’는 예측치임.

출처: The FBI’s Internet Crime Complaint Center; Deloitte Center for Financial Services.

생성형AI로 더 손쉬워진 금융사기

생성형AI 때문에 금융기관과 이들의 고객을 상대로 한 사기행위의 유형과 범위가 무한대로 증폭되고 있다. 사기행위도 창의력만 있다면 얼마든지 새로운 수법이 가능해진 것이다.

기술 혁신이 놀라운 속도로 이뤄지는 만큼, 은행들이 범죄 수법을 미리 파악하기가 갈수록 어려워질 것이다. AI 기반 딥페이크의 ‘자기 학습’ 시스템은 끊임없는 업데이트로 컴퓨터 기반 탐지 시스템을 무용지물로 만든다.[3]

특히 새로운 생성형AI 툴은 싼 가격에 누구나 이용할 수 있기 때문에, 딥페이크 영상과 음성, 문서를 만들기가 더욱 쉬워졌다. 이미 다크 웹(dark web)에는 20달러 짜리부터 수천 달러 짜리까지 사기행위에 이용할 수 있는 소프트웨어를 사고파는 시장(cottage industry)이 형성되어 있다.[4] 범죄용 소프트웨어가 이처럼 무작위로 확산되면서 기존 사기방지용 툴이 무력해지고 있다.[5]

이러한 상황에서 생성형AI 기반 사기행위까지 확산돼 금융서비스 기업들의 우려가 한층 심화되고 있다. 최근 조사에 따르면, 핀테크 부문 딥페이크 사기 건수가 2023년 한 해에만 700% 증가한 것으로 나타났다.[6] 음성 딥페이크만 보면 사기방지 기술이 가짜 콘텐츠 생성 기술을 따라잡지 못하고 있는 실정이다.[7]

3. Alakananda Mitra, Saraju P. Mohanty, and Elias Kougianos, “The world of generative AI: Deepfakes and large language models,” Arxiv.org, February 8, 2024.
4. Nabila Ahmed et al., “Deepfake imposter scams are driving a new wave of fraud,” Bloomberg, August 21, 2023.
5. Hannah Murphy, “Deepfakes make banks keep it real,” Financial Times, September 20, 2023.
6. Isabelle Bousquette, “Deepfakes are coming for the financial sector,” Wall Street Journal, April 3, 2024.
7. Huo Jingnan, “Using AI to detect AI-generated deepfakes can work for audio — but not always,” NPR, April 5, 2024.

특히 이메일 피싱은 생성형AI로 인해 더욱 기승을 부릴 수 있다. 26가지 유형의 인터넷 범죄를 추적하는 미국 연방수사국(FBI) 인터넷범죄신고센터(ICCC)에 따르면, 비즈니스 이메일 피싱이 가장 흔한 인터넷 범죄이자 심각한 재정적 피해를 초래하는 유형으로 나타났다.[8] 지난 수 년간 해커들은 소셜 엔지니어링(social engineering) 공격 기법으로 개인 및 비즈니스 이메일 계정을 해킹해 불법 자금 이체 범죄를 저질렀다. 하지만 생성형AI를 이용하면 이보다 적은 비용을 들이고도 다수의 피해자를 노린 대규모 사기 행위가 가능해진다. FBI에 따르면, 2022년 한 해에만 비즈니스 이메일 피싱 건수가 2만1,832건, 피해액은 약 27억 달러에 달했다. 딜로이트가 제시하는 '공격적 시나리오'는 2027년 생성형AI 기반 이메일 피싱 피해액이 115억 달러에 달할 것으로 전망한다. 은행 업계는 수 십년간 사기행위에 대응하면서 여타 부문에 비해 선제적으로 첨단기술을 활용해 왔다. 하지만 미국 재무부는 보고서에서 "기존 리스크 관리 체제는 첨단 AI 기술을 관리하기에 적절하지 않다"고 지적했다.[9] 과거의 사기방지 시스템은 기업 내 규정과 의사결정 체제를 거쳐야 했지만, 오늘날 금융기관들은 AI와 머신러닝 툴로 위협 행위를 탐지, 경고, 대응하고 있다. 일부 은행들은 AI를 활용해 사기행위를 감지하고 담당 팀에 조사 결과를 전송하는 프로세스를 자동화하고 있다.[10] 또 JP모건(JPMorgan) 등 일부 은행들은 이미 대규모언어모델(LLM)을 도입해 이메일 탈취 행위 등에 대응하고 있다.[11] 마스터카드(Mastercard)는 의사결정지능(Decision Intelligence) 툴을 개발해 수 조개의 데이터 포인트를 스캔해 거래행위의 진위를 파악하는 방식으로 사기행위를 방지하고 있다.[12]

새로운 금융사기의 시대, 은행들의 대응책은?

은행들은 생성형AI 기반 사기에 대한 대응책 마련에 초점을 맞춰야 경쟁력을 유지할 수 있다. 첨단기술과 인간의 직관력을 융합해 사기행위에 선제적으로 대응할 수 있는 기술 활용법을 모색해야 한다. 만병통치약은 없으므로, 사기방지팀은 다양한 범죄 수법을 미리 파악해 대응할 수 있도록 끊임없이 학습 속도를 가속화해야 한다. 이와 동시에 새로운 금융사기의 시대에 발맞춰, 조직 전체의 전략, 거버넌스, 자원 활용 방식 등도 재편할 필요가 있다.

기술 발전의 속도가 워낙 가파른 만큼, 은행들은 사기방지 툴을 개발하는 외부 기관과의 협력도 강화할 필요가 있다. 특정 기업에 대한 위협은 곧 여러 기업과 경제 전반의 위협으로 확산되므로, 은행 리더들은 은행 업계 내외부 기관들과 협업 전략을 수립해 생성형AI 기반 사기행위에 대응해야 한다. 은행들간 협력도 필요하지만, 충분한 역량을 갖추고 신뢰할 수 있는 테크 기업들과도 협력해 사기방지를 위해 각각의 책임 분야를 설정하는 전략을 공동 수립하면 도움이 될 수 있다.

고객들도 사기방지를 위한 파트너로 활용할 수 있다. 하지만 사기 피해가 고객의 행동 때문인지 금융기관의 실수 때문인지에 따라 고객 관계가 시험대에 오를 수 있다는 점을 유의해야 한다. 고객들은 은행 서비스 이용 시 효율성과 안전성을 기대하는데, 생성형AI 딥페이크가 이 두가지를 모두 위협할 수 있다. 이를 방지하기 위해 은행들은 잠재적 리스크의 특징과 은행의 리스크 관리 방식에 대해 고객들에게 충분한 정보를 제공해야 한다. 은행 앱의 푸시 알림을 통해 정보를 제공하는 등 주기적인 커뮤니케이션 접점이 필요하다.

8. Federal Bureau of Investigation, Internet crime report 2023, April 4, 2024.
9. US Department of the Treasury, Managing artificial intelligence-specific cybersecurity risks in the financial services sector, March 2024.
10. Edmund Lawler, "Banks face the twin-edged sword of generative AI," BAI, March 4, 2024.
11. Penny Crosman, "JPMorgan Chase using advance AI to detect fraud," American Banker, July 3, 2023.
12. Mastercard, "Mastercard supercharges consumer protection with gen AI," press release, February 1, 2024.

규제당국들도 생성형AI의 잠재력과 위협을 주시하고 있으므로, 은행들은 새로운 산업 표준을 수립하는 데 적극적으로 동참할 필요가 있다. 또한 신기술 발전 초기 단계에서 컴플라이언스 프로세스와 시스템을 마련해 놓으면, 실제 규제 컴플라이언스가 필요하게 될 때 발 빠르게 대처할 수 있다.

마지막으로 AI 기반 사기행위를 탐지, 대응, 보고할 수 있도록 신규 인력을 확보하고 기존 인력을 훈련하는 데 투자해야 한다. 현재 일부 은행들이 비용 감축을 우선시하는 상황에서 비용과 시간이 많이 드는 이러한 투자에 나서는 것이 쉽지 않은 일이다. 하지만 금융사기에 대비하려면 인력에 대한 대대적 훈련이 우선시돼야 한다. 특히 내부 엔지니어링팀뿐 아니라 외부 벤더 업체와 용역 등을 활용해 사기방지 소프트웨어를 개발함과 동시에 학습과 수정을 지속하는 조직문화를 수립하는 데 초점을 맞춰야 한다.

생성형AI 기반 사기행위는 갈수록 정교해지고 확산되고 있다. 2027년에 이르면 이로 인해 은행들과 고객들이 최대 400억 달러의 피해를 입을 것으로 예상된다. 증폭하는 위협에 대응하려면 더욱 민첩한 대응이 가능하도록 투자를 강화해야 한다.

연구 방법론

생성형AI 기반 금융사기 전망은 과거 추세와 딜로이트의 금융 사기 및 리스크 전문가들이 제시한 데이터를 기반으로 도출했다. 또한 FBI의 ICCC가 추적하는 26가지 유형의 인터넷 범죄 각각에 '생성형AI 사기 리스크' 점수를 집계한 후, 생성형AI 도입 양상을 '보수적 시나리오', '기본 시나리오', '공격적 시나리오'로 구분해 2027년까지 다양한 사기 수법의 증가율을 예측했다. 예측치에는 각 사기 유형의 차이점도 반영했다.