법무법인(유) 린

Tech Legal Insights

AID (Artificial Intelligence Decoding) Vol. 23 2026. 11.

금융 AI agent의 확산과 법적 쟁점

금융위원회는 2026년 8월 25일 「AI 시대 국민의 일상과 생업을 잇는 마이데이터 발전방안」을 통해 이른바 'My AI Agent'를 도입하여 금리인하요구권, 채무조정요구권, 대환대출 신청 등 금융 관련 권리행사를 AI가 대리할 수 있도록 하는 방안을 발표한 바 있습니다. 해외에서도 미국 메타(Meta)가 2026년 9월, 쇼핑과 결제까지 수행하는 개인 AI agent인 Muse를 공개하면서 인기를 끄는 등 AI agent가 이용자를 대신하여 실제 결제까지 수행하는 서비스도 속속 등장하고 있습니다.

그러나 2026년 9월 말부터 10월 초에 걸쳐 국내 시중은행과 저축은행, 캐피탈사 등 다수의 금융회사에서 고객정보가 유출된 사실이 잇따라 알려졌고, 금융당국도 AI를 활용한 공격 가능성을 배제할 수 없다고 밝히면서 AI 기반 해킹에 대한 우려 또한 커지고 있습니다. 앞으로 기술발전에 따라 예금 이동이나 이체 권한까지 가진 AI Agent가 확산되면, 이번 사태와 같은 공격은 금융회사 시스템뿐 아니라 이용자의 AI agent 자체를 겨냥할 수도 있습니다.

현재는 AI agent와 인간의 활동을 구분하기 힘든 계좌이체와 같은 금융거래활동까지는 AI agent의 직접적 개입을 허용하지는 않은 채 이용자의 확인절차를 거쳐 이용자가 허용한 권한의 범위 내에서 AI agent가 활동하게끔 하는 식의 보수적 방식이 일반적입니다만 AI agent의 발전 속도를 고려하면 적절한 법적 통제 방안을 찾는 노력은 계속될 수밖에 없습니다.

최근 미국에서는 2026년 10월 1일, 공화당의 조시 홀리(Josh Hawley) 상원의원과 민주당의 크리스 머피(Chris Murphy) 상원의원이 AI agent의 불법적인 해킹 행위에 대해 개발자(developer)와 운영자(operator)에게 컴퓨터사기남용법(CFAA)상 형사·민사책임을 부과하는 「AI Agent Accountability Act」을 발의하였습니다. 동 법안은 AI agent의 해킹·무단 전산접근에 집중되어 있지만 AI가 자율적으로 불법행위를 했다는 이유로 개발자나 운영자의 책임을 면제해서는 안 된다는 점을 명확하게 보여준다는 점에서는 의미가 있습니다.

Related Areas

[AI · 플랫폼 · 테크놀로지 부문]
AI agent, 전자금융거래법, 프롬프트 인젝션, Muse 모델

Contact

구태언 변호사
T. 02-3477-8695
E. tekoo@law-lin.com

방석호 미국변호사
T. 02-3477-8695
E. shbang@law-lin.com

그러나 동 법안은 AI agent가 정상적인 금융시스템 접근권한을 가진 상태에서 이용자가 미리 허용한 거래 범위나 내부 승인 절차를 벗어나 행동한 경우까지 다루고 있지는 않습니다.

이하에서는 AI agent를 활용한 금융회사 시스템 침해로 예금주에게 피해가 발생한 경우 적용될 수 있는 법률관계를 정리하고, 정당한 권한을 가진 AI agent가 해킹이나 '프롬프트 인젝션'의 표적이 되는 경우 현행 법체계가 어떻게 작동하는지, 한계는 무엇인지, Muse와 같은 범용 개인 AI agent 결제 구조는 위험을 어디까지 줄일 수 있는지 등을 살펴봅니다.

Ⅰ. 금융회사 해킹으로 예금주 피해가 발생한 경우

1. 정보 유출 자체에 대한 손해배상

금융회사가 보유한 고객정보는 대부분 개인신용정보에 해당하므로「신용정보의 이용 및 보호에 관한 법률」(이하 "신용정보법")이「개인정보 보호법」보다 우선 적용됩니다. 신용정보법은 신용정보전산시스템의 안전보호 의무(제19조)와 유출 시 통지 의무(제39조의4)를 규정하고 있습니다. 손해배상과 관련, 신용정보법 제43조는 신용정보회사등이 고의 또는 과실이 없음을 증명하지 못하면 책임을 면할 수 없도록 입증책임을 전환하고 있고, 고의 또는 중대한 과실로 인한 유출에 대해서는 손해의 5배를 넘지 않는 범위에서 징벌적 손해배상을 인정합니다. 또한 제43조의2는 손해액을 입증하기 어려운 경우에도 300만 원 이하의 범위에서 법정손해배상을 청구할 수 있도록 하고 있습니다.

한편, 신용정보에 해당하지 않는 정보나 외주직원·대출모집인 등의 정보에 대해서는「개인정보 보호법」의 안전조치의무(제29조), 유출 통지(제34조), 손해배상(제39조) 및 법정손해배상(제39조의2) 규정이 적용될 수 있습니다. 외주업체를 통해 정보가 처리된 경우에는 같은 법 제26조에 따라 위탁자인 금융회사의 관리·감독 책임도 함께 검토되어야 합니다. 아울러 신용정보의 처리를 위탁한 경우에는 신용정보법 제43조 제6항에 따라 위탁자인 금융회사가 수탁자와 연대하여 배상책임을 질 수 있습니다. 다만 실제 배상 규모와 관련, 2차 피해가 없는 한 개인별 배상액은 크지 않을 수 있기 때문에 정보 유출로 인한 소송 사건은 집단분쟁조정이나 공동소송의 형태로 다투어지는 경우가 많습니다.

麟 서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

2. 2차 피해의 경우

유출된 정보를 이용한 맞춤형 피싱(phishing)으로 예금주가 직접 이체를 하였다면, 「전기통신금융사기 피해 방지 및 피해금 환급에 관한 특별법」(이하 "피해환급법")에 따른 지급정지와 피해금 환급 절차를 이용할 수 있습니다. 물론 정보를 유출한 금융회사에 손해배상을 청구하는 것도 가능합니다만 실무상으로는 예금주의 직접 이체행위가 개입되었기 때문에 **정보 유출과 피싱 피해 사이의 상당인과관계**를 입증하는 것이 쟁점이 됩니다.

법적으로는「전자금융거래법」제9조 제2항 제1호 및 제3항에 따라 이용자에게 고의 또는 중대한 과실이 있고 그 책임 분담을 미리 약관으로 정해 둔 경우 금융회사는 책임의 전부 또는 일부를 이용자에게 부담시킬 수 있고, 동법 시행령 제8조는 접근매체의 대여·사용 위임·양도, 접근매체의 누설·노출·방치, 금융회사의 추가 보안조치에 대한 정당한 사유 없는 거부 등을 고의 또는 중대한 과실의 유형으로 열거하고 있습니다. 따라서 피싱으로 탈취된 접근매체를 제3자가 이용한 형태의 2차 피해에서는 이용자의 고의 또는 중대한 과실 여부가 주된 쟁점이 됩니다. 반면 예금주가 스스로 이체한 경우에는 제9조 제1항의 '사고'에 해당하기 어려워 피해환급법에 따른 구제가 중심이 됩니다.

한편, 전자금융거래법 제9조와 이용자의 고의·중과실 범위를 정한 동 시행령 제8조는 '접근매체의 관리는 이용자가, 시스템의 안전은 금융회사가 책임진다'는 전제하에 사람이 접근매체를 들고 직접 거래하고, 제3자가 그 접근매체를 훔쳐 쓴다는 상황을 상정하여 금융회사와 이용자의 책임분담을 설계한 것입니다. 해킹기술 발전에 따라 금융회사의 추가적 보완조치가 더해지고 이용자가 이를 따랐는지를 책임분담 기준에 더하는 등 일부 보완이 이루어졌지만 위의 전제와 책임분담 틀은 변하지 않았기 때문에 만약 사람이 아닌 AI agent가 금융거래에 개입하면 이 구도가 여러 지점에서 어긋나게 됩니다

Ⅱ. 전자금융거래법과 금융위원회의 AI 가이드라인

1. AI agent 확산에 따른 한계

麟 서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

전자금융거래법 제6조 제1항은 금융회사에 이용자의 신원·권한·거래지시 내용을 확인하도록 규정하고 있고, 제9조 제1항은 접근매체의 위조·변조, 거래지시의 전자적 전송·처리 과정에서 발생한 사고, 침입을 통해 부정취득한 접근매체의 이용으로 발생한 사고 등에 대해 금융회사 또는 전자금융업자의 손해배상책임을 규정하고 있습니다. 이 책임은 금융회사의 과실을 요건으로 하지 않는 무과실책임의 구조이므로, 실무상으로는 제9조 제2항·제3항에 따라 이용자의 고의·중과실을 이유로 책임을 분담시킬 수 있는지가 핵심 쟁점이 됩니다.

그러나 AI agent가 금융거래를 실행하게 될 때 기존의 금융회사와 예금주라는 2자 관계에 제3자인 AI agent 제공자가 등장하게 되고, 금융회사가 받은 거래지시가 예금주의 유효한 의사에 기초한 것인지, 아니면 AI agent가 위임범위를 벗어나 독자적으로 실행한 것인지를 구분하는 데 책임분담의 초점이 있게 됩니다.

이에 따라 금융회사로서는 당장 두 가지 변화에 대비하여야만 합니다.

첫째, 전자금융거래법 제6조의 이용자 신원확인과 거래권한 확인을 분리하여 더 강화할 필요성이 커집니다. 금융회사가 고객의 신원뿐 아니라 AI agent의 위임권한, 거래범위, 거래금액, 권한의 유효기간까지 확인해야 하는 어려운 문제가 발생하기 때문입니다.

둘째, 제9조의 금융사고에 대한 대외적 배상책임과 최종적인 내부 책임분담을 분리해야 할 필요성이 커집니다. 금융회사가 고객에게 먼저 배상하더라도 사고를 일으킨 AI agent 제공사업자에게 구상할 수 있는 법적·계약적 체계가 필요해집니다.

2. 전자거래금융법 제 9조

1) 사고 유형 열거 방식의 한계 (제9조 제1항)
AI agent 관련 피해는 제1항이 열거한 세 유형, 즉 접근매체의 위조·변조(제1호), 전송·처리 과정의 사고(제2호), 침입으로 획득한 접근매체의 이용(제3호)에 해당되지 않을 수 있습니다.

첫째, 접근매체 개념의 공백입니다. 제2조 제10호의 접근매체는 전자식 카드, 인증서, 이용자번호, 생체정보, 비밀번호 등을 예정하고 있어, AI agent가 실제로 사용하는 API 토큰, 세션 쿠키 등이 '전자식 카드에 준하는 전자적 정보'(가목)에 포함되는지는 명확하지 않습니다.

麟 법무법인(유) 린
서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

둘째, 제2호 적용의 불확실성입니다. AI agent의 오작동, 환각, 조작으로 잘못된 이체 지시가 나간 경우를 "처리 과정에서 발생한 사고"로 볼 수 있을지가 문제됩니다. 금융회사는 형식상 정상적인 거래지시를 정상적으로 처리했을 뿐이라고 다툴 여지가 큽니다.

2) 책임 주체의 공백

제9조의 책임 주체는 금융회사와 전자금융업자이지만 외부의 AI agent 제공사업자는 전자금융업자로 등록되지 않는 한 제9조의 밖에 있게 됩니다. 다만 AI agent 제공사업자가 금융회사와 제휴하여 전자금융거래를 보조하거나 일부를 대행하는 경우에는 전자금융보조업자(제2조 제5호)에 해당하여, 제11조에 따라 그 고의·과실이 금융회사의 고의·과실로 의제될 여지가 있습니다.

특히 이용자가 독자적으로 사용하는 외부의 AI agent로 인해 피해가 발생했고 이에 따라 금융회사가 상대적으로 쉽게 면책이 된다면 이용자는 외부의 AI agent 제공사업자를 상대로 민법상 채무불이행이나 불법행위, 제조물책임법등으로 다투어야 하지만 약관상 책임제한 조항, 해외 사업자에 대한 집행 문제 등의 추가 난관에 부닥칠 수 있습니다.

3) 입증의 어려움

이용자의 귀책사유는 금융회사가 입증해야 하지만, AI agent가 해킹 등으로 조작되었다는 무권한성은 사실상 이용자가 먼저 보여야만 합니다. 그런데 AI agent의 추론 과정과 실행 로그 기록에 접근하여 이용자가 이를 입증한다는 것은 대단히 어려운 작업입니다. 특히, 금융회사 측 로그기록에는 '정상 인증된 정상 거래'로만 남아 있기 때문입니다.

3. 금융위원회 AI 가이드라인

2026년 6월 22일부터 시행된 금융위원회의 개정 「금융분야 인공지능 가이드라인」은 거버넌스, 합법성, 보조수단성, 신뢰성, 금융안정성, 신의성실, 보안성의 7대 원칙을 제시하고 있지만, 자율규제의 성격상 외부 AI agent가 일으킨 금융사고에 대하여 금융회사의 책임을 면제하는 규정은 아닙니다.

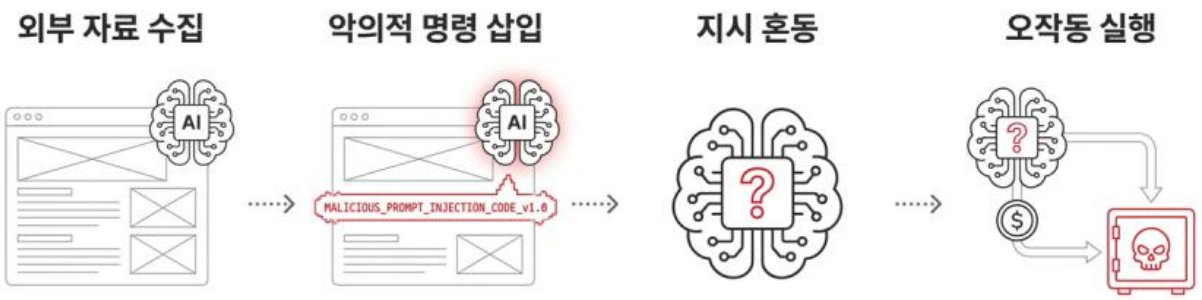
麟 법무법인(유) 린
서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

즉 금융회사 입장에서는 해당 가이드라인을 준수하더라도 전자금융거래법 제6조와 제9조의 적용 가능성이 여전히 남아 있게 됩니다. 금융위원회의 2026년 6월 18일 보도자료(금융권 AX 현장 간담회)에 따르면, 금융업계는 외부 서비스에 AI agent를 적극 도입하기 어려운 이유로 망 분리, 데이터 규제, 접근매체 인증, 책임소재, 업종 분류 등 규제상 제약과 불확실성을 들었고, 금융위원회는 향후 AI agent의 권한과 책임에 관한 규율체계를 검토하고, 샌드박스를 통해 통제된 범위에서 단계적으로 검증하겠다는 방향을 밝혔습니다.

따라서 현재와 같이 AI agent에 관한 규율체계가 정비되는 과도기에는 금융회사가 외부 AI agent 거래를 제한적으로 운영하려는 경향을 보일 수밖에 없고, 이는 해외 금융기관에서도 유사하게 나타나는 현상입니다. 또한 이용자의 통제와 승인을 전제로 AI agent의 금융거래를 허용한다 하더라도 이용자의 위임권한을 검증할 수 없는 외부 AI agent의 자율적 거래까지는 당분간 허용되기는 어려울 것으로 보입니다.

Ⅲ. 권한을 가진 AI agent가 표적이 된다면

1. 금융거래에서의 AI Agent가 해킹의 도구가 아닌 표적이 될 때 발생할 위험의 주된 원인으로 프롬프트 인젝션(prompt injection)을 생각할 수 있습니다.



서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

AI Agent는 이용자의 지시와 함께 업무 수행 중 접하는 웹페이지, 이메일, 문서 등 외부 자료를 읽는데, 현재의 대규모 언어모델은 이 둘을 구조적으로 명확히 구분하기 어렵다는 한계가 있습니다. 그 결과 외부 자료에 명령처럼 숨겨 둔 문장이 있으면 이를 이용자의 지시처럼 따를 수 있습니다. 프롬프트 인젝션은 기존의 해킹처럼 SW의 취약점이나 인증체계의 결함 등을 악용하는 방식이 아니라 공격자가 AI 시스템에 입력되는 정보에 악의적인 명령을 삽입하여, AI가 원래의 지시나 권한범위를 벗어난 행동을 하도록 유도하는 기법입니다. 특히 AI agent는 판단 결과를 실제 행동으로 옮기므로 금융거래에서는 그 위험성이 더욱 커집니다.

시스템을 '침입하는' 공격이 아니라, 정당한 권한을 가진 AI agent를 '속여' 그 권한을 사용하게 만드는 공격이기 때문에 금융 해킹에 관한 현행 법체계가 상정하지 못한 문제를 제기합니다. 즉 AI agent가 속아서 이용자에게 피해가 발생하면 이용자의 유효한 거래지시가 있었는지. 금융회사의 권한확인 책임여부와 나아가 AI agent 제공사업자의 보안책임도 함께 검토되어야 합니다. 현행 법상으로는 다음의 주요 이슈들이 제기될 수 있습니다.

첫째, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」(이하 "정보통신망법") 제48조 제1항은 정당한 접근권한 없이 또는 허용된 접근권한을 넘어 정보통신망에 '침입'하는 행위를 금지하고 있습니다. 그러나 프롬프트 인젝션은 반드시 정보통신망에 대한 '침입'을 수반하지 않습니다. 외부 웹페이지에 악의적인 문구를 삽입하는 것만으로도 공격이 가능하기 때문입니다. 다만 공격자의 형사책임은 컴퓨터등사용사기죄(형법 제347조의2: 정보처리장치에 '부정한 명령'을 입력하여 재산상 이익을 취득) 등으로 물을 여지가 있으므로, 실질적인 공백은 전자금융거래법 제9조 제1항 제3호가 정보통신망법상 '침입'을 전제로 하는 데서 비롯되는 민사상 책임분담 문제에 있다고 볼 수 있습니다.

둘째, 전자금융거래법 제6조가 금융회사에게 "전자금융거래를 위하여 접근매체를 선정하여 사용 및 관리하고 이용자의 신원, 권한 및 거래지시의 내용 등을 확인하여야" 할 책임을 규정하고 있지만 프롬프트 인젝션을 통한 기망시 금융회사 시스템은 해당 거래를 정상 인증을 거친 이용자의 거래지시로 파악하게 될 확률이 큽니다. 이 경우 이를 제9조 제1항 제3호의 '침입하여 부정한 방법으로 획득한 접근매체의 이용'으로 보아 금융회사가 책임을 부담하는지가 불분명하므로 책임소재를 둘러싼 공방이 불가피합니다.

麟 서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

셋째, 「전자문서 및 전자거래 기본법」 제7조 제1항 제2호는 자동으로 전자문서를 송신·수신하도록 구성된 컴퓨터프로그램에 의하여 송신된 전자문서를 작성자가 송신한 것으로 보므로, AI agent가 보낸 거래지시의 효과는 일단 이용자에게 귀속될 가능성이 큽니다. 같은 조 제3항은 수신자가 상당한 주의를 하였더라면 작성자의 것이 아님을 알 수 있었던 경우를 예외로 두고 있으나, 이 예외는 문언상 제2항 제2호의 경우에 적용되는 것이어서 자동화된 프로그램에 의한 송신에 바로 적용하기는 어렵습니다. 다만 평소와 전혀 다른 수취인이나 금액이었음에도 금융회사의 이상거래탐지시스템(FDS)이 이를 걸러내지 못하였다면, 전자금융거래법 제6조의 확인의무 등을 근거로 금융회사의 책임이 문제될 여지는 있습니다.

넷째, 피해환급법은 전기통신을 이용하여 '타인'을 기망·공갈함으로써 자금을 송금·이체하도록 하는 행위 등을 중심으로 전기통신금융사기를 정의하고 있습니다. 속은 주체가 정당한 권한을 가진 AI agent인 경우에도 지급정지와 피해금 환급 절차를 이용할 수 있는지는 해석상 논란이 있을 수 있습니다. 또한 동법은 재화의 공급 등을 가장한 행위를 전기통신금융사기에서 제외하고 있어, 가짜 쇼핑몰 등을 통한 결제 유도형 피해는 처음부터 적용 대상에서 벗어날 가능성이 큽니다.

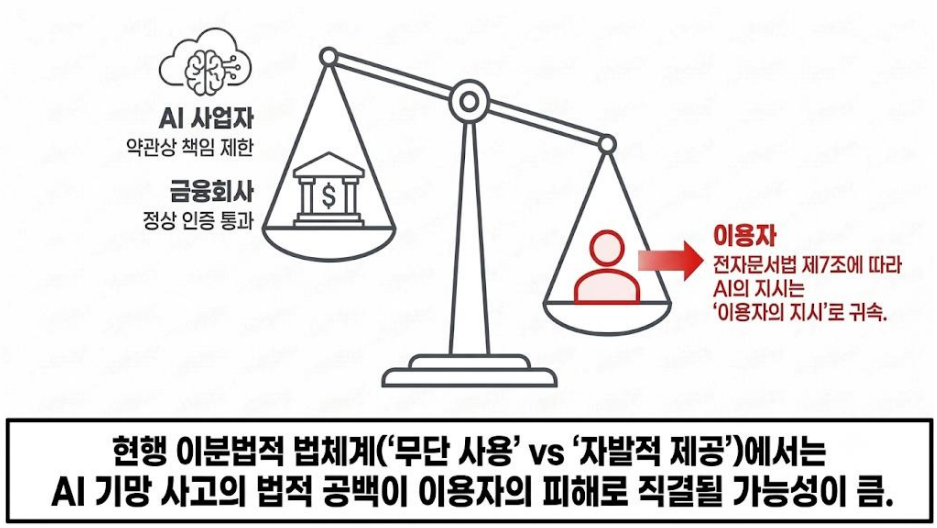
2. 기망당한 AI agent의 거래는 누구의 책임?

이용자가 사용하는 외부 AI agent는 독립적 법인격이 없어 민법상 대리인이 될 수 없으므로 이용자의 도구로 평가되고, 그 거래지시는 전자문서법 제7조에 따라 이용자의 것으로 귀속될 가능성이 큽니다. 여기에 이용자가 AI agent에게 포괄적인 금융거래 권한을 부여한 상태에서 프롬프트 인젝션으로 AI agent가 기망당한다면, 그에 따른 거래는 전자금융거래법 제9조 제1항의 '침입에 따른 사고'에 해당하지 않고 피해환급법상 전기통신금융사기 피해로 보기 어렵다고 해석될 수 있습니다. 나아가 이용자가 인증정보 등 접근매체를 AI agent 사업자에게 맡긴 것이 전자금융거래법 시행령 제8조 제1호의 '접근매체의 사용 위임'으로 평가된다면, 이는 오히려 이용자의 중대한 과실 사유가 될 수도 있습니다.

결국 현행법 아래에서는 그 손실이 금융회사가 아닌 이용자에게 귀속될 가능성이 커집니다. 만약 금융회사의 책임을 1차적으로 묻는 전자금융거래법의 논리를 유지한다면 금융회사가 외부 AI agent의 사용을 허용할 때 AI 모델이 악의적인 지시에 속지 않도록 하는 방어조치 (주로 AI agent 제공사업자의 영역)뿐 아니라, AI가 실제 결제를 요청할 때 고객이 위임한 금액·거래상대방·유효기간 등을 금융회사가 독립적으로 검증할 수 있는 단계를 마련할 필요가 있습니다.

麟
서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

그래야만 AI agent가 금융거래권한을 가지게 될 경우 피해를 본 이용자를 보호하는 데 있어 발생하게 될 법상 공백을 부분적이나마 줄일 수 있습니다.



IV. 미국 Muse 모델의 등장과 한계

1. 실행 구조

Meta가 2026년 9월, 미국에서 출시한 범용 개인 AI agent인 Muse의 결제 기능은 글로벌 지급결제 서비스 회사인 Stripe의 'Link' 에이전트 지갑을 기반으로 하며, 다음과 같은 구조로 되어 있습니다. 즉, ① 고객 → Muse (상품 검색·비교·구매 준비 위임), ② 개인 AI agent인 Muse는 고객의 지시에 따라 상품 선택 및 결제 요청 준비, ③ 고객은 거래별 구매금액을 확인하고 승인, ④ Stripe Link → 카드 결제망 (기존 결제수단 또는 승인된 구매에 한정된 일회용 가상카드 사용)의 순서로 이용하게 됩니다.

특히 Stripe의 공식 발표에 따르면 Muse는 매번 고객에게 채팅 화면에서 결제총액의 승인을 요청하며, 고객의 실제 결제정보는 Muse에 노출되지 않습니다. Link를 지원하지 않는 판매자에게 결제할 때에는 승인된 구매에 한정된 일회용 가상카드를 사용합니다. 판매자는 별도의 AI 전용 API (Application Programming Interface)를 만들지 않아도 되기 때문에 일반 신용카드 거래와 같이 처리할 수 있고, Muse는 사람처럼 판매자 웹사이트를 돌아다니면서 결제까지 실행을 하게 됩니다.

麟 서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

다시 말해, Meta의 Muse가 판매자의 기존 온라인 쇼핑몰에 사람이 직접 접속하는 것처럼 접속하여 상품을 검색하고 구매할 수 있고, AI agent 제공사업자인 Meta가 개별 판매자들과 사전에 API 연동 계약을 체결하지 않고도 판매자의 기존 시스템에 맞춰 작동하기 때문에 광범위한 전자상거래 서비스를 제공할 수 있다는 장점이 있습니다.

그러나 판매자 입장에서는 방문자가 사람인지 AI agent인지 구별하기 어렵고, 고객 접점과 광고·추천 등에 대한 통제권을 잃을 수 있다는 우려도 있습니다. 실제로 Amazon은 2026년 9월 20일 Muse가 구매시 AI agent임을 밝히지 않고 고객 계정 페이지와 구매 내역에 접근할 수 있다는 점 등을 이유로 자사 쇼핑몰에 대한 Muse의 접근을 차단한 바 있습니다. 자체 계정·결제수단·고객관계를 기반으로 AI 구매 기능을 설계할 수 있는 대형 플랫폼은 외부 AI agent의 접근을 제한할 유인이 있다는 점에서, 이는 범용형 모델의 한계로 볼 수 있습니다.

한편, Muse AI agent의 상품선택, 구매 준비 기능과 결제실행기능을 분리하여 후자에는 고객의 최종 승인을 넣었기 때문에 AI agent가 직접 고객의 계좌에서 대금을 이체하는 위험은 크게 줄어듭니다.

2. 국내 규제체계에 비추어 본 Muse 모델

현행 전자금융거래법 제6조는 금융회사로 하여금 이용자의 신원, 권한 및 거래지시 내용을 확인하도록 요구하고 있고, 금융위원회의 금융분야 AI 가이드라인이 제시하는 7대 원칙에는 합법성, 보조수단성, 보안성 등이 포함되어 있습니다.
Muse의 결제실행 구조는 다음과 같은 이유로 현 국내의 규제체계와 대체로 부합하는 것으로 볼 수 있습니다.

1) Muse 방식에서는 판매자(가맹점)와 카드사가 Stripe가 구축한 기존 결제인증·승인 절차를 그대로 이용할 수 있습니다. 이에 따라 금융회사가 Muse 사용에 따른 승인, 내부 권한설정을 직접 확인하지 않더라도, 고객이 직접 승인한 개별 결제를 기존 방식대로 처리하게 됩니다.

2) Muse가 구매를 준비하고 상품을 선택, 결제를 요청하지만 고객이 매 거래마다 최종 승인하도록 하고 있기 때문에 금융위원회 AI 가이드라인의 보조수단성 원칙의 취지에 부합하고, 실제 결제정보를 AI agent에 노출하지 않으며 승인된 구매에 한정된 일회용 가상카드를 사용한다는 점에서 보안성 원칙의 취지에도 부합한다고 볼 수 있습니다.

麟 법무법인
서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

결국 Muse 모델은 AI agent가 고객의 별도 승인 없이 자율적으로 금융거래를 수행하는 문제까지 해결한 것은 아니지만 현재의 규제체계에서 금융회사들이 수용할 수 있는 보수적이면서도 범용성을 갖춘 '반(半)자율형' AI agent 비즈니스 모델이라고 할 수 있습니다.

3. 신용카드 기반 AI agent는 안전한가?

Muse처럼 신용카드 등 기존의 결제인프라를 이용, AI agent가 작동하게 할 경우 카드번호 탈취를 막고, 피해 규모를 제한하며, 공격자의 계좌로 현금이 바로 빠져나가는 것을 어렵게 한다는 점 등에서 계좌이체에 대한 권한까지 갖게 되는 AI agent보다 훨씬 안전하다고 인식되고 있습니다.

그러나 신용카드 기반 구조에서도 프롬프트 인젝션은 얼마든지 발생할 수 있습니다.
AI agent를 속여 가짜 가맹점을 선택하게 하거나, 조작된 정보로 "정품", "최저가"라고 요약하여 이용자의 승인을 받아내거나, (자동결제 권한이 부여된 경우) 미리 위임해 둔 조건이 충족된 것처럼 속여 자동 결제를 일으키는 방식 등이 가능하며, Meta의 Muse agent처럼 카드번호를 보지 못하게 하는 설계라고 하더라도 가맹점 선택 단계의 조작까지 막지는 못합니다.

즉 카드 기반형 AI agent 금융결제는 "큰돈이 공격자 계좌로 바로 빠져나가는 피해"를 "잘못된 구매와 정보 유출"로 바꿔 줄 뿐, 모든 피해를 없애 주지는 못합니다. 법적인 피해 구제에도 한계가 있습니다. 현행 「여신전문금융업법」 제16조 제5항은 해킹 등 전자적 침해로 인한 카드 부정사용에 대해 카드사의 책임을 원칙으로 하고 있으나, 이용자가 AI agent에게 결제 권한을 부여하고 결제 화면에서 직접 승인까지 하였다면 이는 '부정사용'이 아니라 '본인이 승인한 거래'로 평가될 여지가 큽니다. 이 경우 피해를 본 이용자는 가맹점을 상대로 한 다툼을 하여야 하지만 만약 상대방이 해외의 가짜 쇼핑몰이라면 구제의 가능성은 크게 떨어집니다. 일정 금액 이상의 신용카드 할부거래라면 「할부거래에 관한 법률」상 항변권을, 해외 가맹점 거래라면 국제 카드 브랜드의 이의제기 절차를 활용할 여지가 있으나, 그 요건과 효과는 제한적입니다.

V. 결론

최근 금융권 해킹 사태에서 우려되는 것처럼 공격자가 AI를 '도구'로 사용하여 금융회사 시스템을 침해한 경우에는 「신용정보법」, 「개인정보 보호법」, 「전자금융거래법」 제9조 등 기존 법체계가

麟 서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

비교적 잘 작동합니다. 공격 수단이 AI라는 점이 금융회사의 책임 구조를 근본적으로 바꾸지는 않기 때문입니다.

반면 결제권한까지 가진 AI agent가 프롬프트 인젝션의 표적이 되는 경우라면 사정이 다릅니다. 「전자금융거래법」 제9조와 피해환급법은 '사람'의 거래지시와 '사람'에 대한 기망을 전제로 설계되어 있고, 「전자문서 및 전자거래 기본법」은 자동화된 프로그램의 전자문서를 일률적으로 이용자에게 귀속시키는 구조여서, "이용자가 승인하였지만 그 승인의 기초가 된 정보는 조작된 거래"를 어떻게 평가할지에 대해서는 아직 분명한 답을 주지 못합니다. 결제수단 부정사용에 관한 기존 법 체계 역시 '무단 사용'과 '이용자의 자발적 제공'이라는 이분법에 기초하고 있어, 이런 구조를 그대로 적용하면 피해가 이용자에게 돌아갈 가능성이 큽니다. 신용카드 기반 구조 역시 이러한 위험을 줄일 수는 있지만 완전히 제거하지는 못합니다.

따라서 AI agent 금융의 본격적인 확산에 앞서 금융회사는 다음과 같은 사항을 사전 점검해 둘 필요가 있습니다.

① 고객 거래 채널뿐 아니라 직원용·업무지원 시스템까지 포함하여 AI 기반 자동화 공격을 전제로 한 보안 점검, 관련 투자를 수행할 필요가 있습니다.
② 분쟁에 대비하여 권한 확인부터 거래 실행까지의 과정을 위·변조가 불가능한 형태로 기록·보존하고, 금융회사의 이상거래탐지시스템도 AI Agent 거래의 특성을 반영하여 고도화할 필요가 있습니다.
③ 향후 외부 AI agent와의 연동을 허용하는 경우에는 이용자가 위임한 금액·거래상대방·유효기간 등을 독립적으로 검증할 수 있는 체계를 마련하고, 접근매체 위임과의 관계 및 책임분담 기준을 약관과 이용자 안내에 명확히 반영해 둘 필요가 있습니다.

다만 현재의 법제도하에서는 금융회사가 자신이 통제할 수 있는 인증·결제 인프라, 제한된 API, 고객의 거래별 승인방식 등의 틀안에서만 외부 AI agent를 보수적으로 허용할 가능성이 큽니다.

한편, 법적으로는 「전자금융거래법」상 'AI agent에 의한 거래지시'의 개념과 책임 배분, 「피해환급법」의 적용 범위 등에 관하여 해외 입법 및 업계 표준 논의 동향을 참고하여 제도적 보완 논의가 필요합니다. 현행법은 "권한 없는 자가 쓴 것인가, 이용자가 스스로 내준 것인가"라는 이분법 위에서 책임을 분담하고 있지만 AI agent 문제는 이 둘 중 어디에도 맞지 않기 때문에 기존 법체계 안에서 해결하려 하면 결국 이용자의 승인으로 인한 거래로 귀결될 확률이 높습니다.

麟 서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.

특히 AI가 공격의 도구이자 표적이 되는 상황에서, AI agent 산업이 향하는 다음 단계는 AI agent가 미리 정해진 범위 안에서 인간의 건별 승인 없이 결제까지 완결하는 것입니다. Google의 agent 결제 프로토콜(AP2)이나 Visa·Mastercard 등 카드 네트워크의 agent 결제 서비스도 이러한 방향을 지향하고 있습니다. 그 단계에서는 AI agent의 권한 초과와 책임 귀속 문제가 본격적으로 제기될 것이므로, 금융회사, AI agent 제공사업자, 감독당국의 공동 대응이 반드시 필요합니다.

<AI 국내외 법률이슈에 대해 매월 발간하는 법무법인 린 AI·플랫폼·테크 A.P.T. 그룹 AI산업센터 뉴스레터 AID에 대한 질문, 조언 등은 구태언 A.P.T. 그룹장(tekoo@law-lin.com), 방석호 AI산업센터장(shbang@law-lin.com)에게 보내주십시오.>

麟 서울특별시 서초구 서초중앙로24길 10,13층, 14층(서초동, 316타워) | Tel: 02-3477-8695 | Fax: 02-3477-8694
E-mail: lin@law-lin.com | © 2017 LIN. All Rights Reserved.