bank und markt

55. Jahrgang | Fritz Knapp Verlag | die-digitale-bank.de

9

15. September 2026

DIE DIGITALE BANK

Digitaler Sonderdruck

FACETTEN DER SICHERHEIT

Agentic Fraud – eine neue Betrugskategorie

Adam Davies

In dieser Ausgabe

KARTEN
cards | cartes NEWS

Foto: AdobeStock/SteezyFX

Agentic Fraud – eine neue Betrugskategorie

Von Adam Davies

Foto: AdobeStock/Hanif

Agentenbasierte Zahlungssysteme und KI-gesteuerte Kunden-Agenten verändern die Betrugsrisiken im Banking. Das Zeitfenster für die Erkennung schrumpft, etablierte Modelle zur Betrugsprävention stoßen an ihre Grenzen, nicht zuletzt, weil Betrugsangriffe sich durch KI-Agenten enorm skalieren lassen. Auch die Regulierung ist an die neue Betrugskategorie noch nicht angepasst, warnt Adam Davies. Offen sind beispielsweise Haftungsfragen. Banken stehen deshalb vor der Aufgabe, ihre Betrugsprävention auf eine Welt vorzubereiten, in der Maschinen nicht nur Zahlungen ausführen, sondern auch andere Maschinen zu täuschen versuchen. Red.

Seit rund drei Jahrzehnten beruht die Betrugserkennung in der Finanzdienstleistungsbranche auf einer weitgehend stabilen Annahme: Menschen initiieren eine Transaktion. Das Verhalten dieser Menschen kann beobachtet, modelliert und hinterfragt werden, wenn es von gängigen Nutzungsmustern abweicht, und bildet die Basis der gängigen Betrugsprävention bei Banken. Menschen sind berechenbare Wesen und nichtmenschliche Verhaltenssignale gelten als deutliche Hinweise auf ein Betrugsrisiko. Predictive AI ist hierbei zum bewährten Rückgrat der Echtzeitüberwachung und Mustererkennung geworden. Machine-Learning-Modelle verarbeiten dabei Tausende von Transaktionen pro Sekunde, um Auffälligkeiten zu erkennen, bevor Verluste entstehen.

Heute sehen sich Banken allerdings mit einer neuen Entwicklung konfrontiert: Nicht-menschliche Akteure, die im Auftrag von Konsumenten oder Unternehmen handeln, werden zu eigenständigen Wirtschaftsteilnehmern mit Zahlungsbefugnis. Die technologische Grundlage, auf die sich die Finanzbranche seit 30 Jahren verlässt, muss damit eine Realität abbilden, für die sie nie konzipiert wurde: eine Welt, in der hinter einer Transaktion nicht mehr zwangsläufig ein Mensch steht.

Schon heute übernehmen Maschinen vielfältige Aufgaben mit finanzieller Relevanz: Beispielsweise bestellen und bezahlen vernetzte Kühlschränke selbstständig Lebensmittel, oder Autos übernehmen selbstständig die Bezahlung von Parkgebühren, Maut und Ladevorgängen, ohne dass der Fahrer aktiv eingreift. KI-Agenten verwalten Abonnements, vergleichen Angebote und wickeln Einkäufe eigenständig ab. Weitere Pilotanwendungen im Handel, in der Reisebranche und in Ökosystemen vernetzter Geräte laufen bereits. Das Marktpotenzial ist beträchtlich: Prognosen zufolge könnte das von KI-Agenten ausgelöste Transaktionsvolumen bis 2030 weltweit 1,7 Billionen US-Dollar erreichen.* Andere Schätzungen gehen sogar von 3 bis 5 Billionen US-Dollar aus. Der Zahlungsverkehr würde folglich zunehmend nicht mehr von Menschen, sondern von Maschinen initiiert.

Perspektivwechsel im Betrugsmanagement

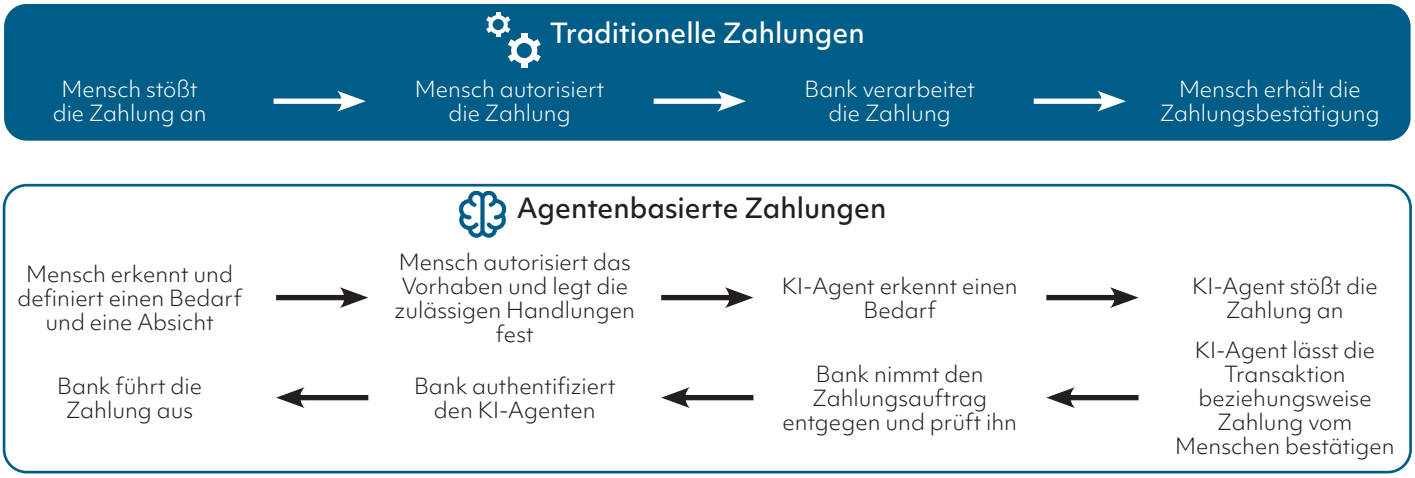
Für das Betrugsmanagement bedeutet diese Entwicklung einen Perspektivwechsel. Bisher stand vor allem die Frage im Mittelpunkt: „Wer ist der Kunde?“ In Zukunft gilt es zusätzlich festzustellen, welcher KI-Agent handelt, von wem er autorisiert wurde und ob die jeweilige Transaktion dem erteilten Auftrag entspricht. Dafür reicht es nicht aus, bestehende Betrugsverfahren einfach nur zu beschleunigen. Einer zunehmend KI-gesteuerten Bedrohung lässt sich auf Dauer nur mit einer strukturell anderen, entsprechend automatisierten und agentenbasierten Betrugsabwehr wirksam begegnen.

Denn der Übergang von menschlich ausgelösten zu agentenbasierten Zah-

Foto: Fico

Adam Davies, Vice President, Fraud Product Management, Fair Isaac Corporation (FICO), Bozeman (Montana)

Abbildung 1: Traditionelle Zahlungen versus Agentic Payments



Quelle: Fico

lungen verändert den gesamten Ablauf einer Transaktion. Bisher stellte ein Mensch einen Bedarf fest, suchte und kaufte das entsprechende Produkt, autorisierte die Zahlung und erhielt nach der Verarbeitung der Zahlung eine Bestätigung seiner Bank. Bei einer agentenbasierten Transaktion gibt der Kunde dagegen zunächst ein Ziel vor, erteilt die erforderlichen Vollmachten und definiert klare Handlungsgrenzen. Innerhalb dieses vordefinierten Rahmens kann der KI-Agent selbstständig einen Bedarf erkennen, eine Aktion wie beispielsweise eine Bestellung durchführen und die entsprechende Zahlung auslösen. Der Kunde wird unter Umständen erst über den Kauf informiert, nachdem die Bank den KI-Agenten authentifiziert und die Transaktion bereits verarbeitet hat.

Geschwindigkeit kein verlässlicher Hinweis mehr für Betrug

Der Zahlungsprozess wird damit komplexer und für Banken wesentlich komprimierter. Gleichzeitig stehen weniger der bisher üblichen Verhaltensmerkmale für die Betrugserkennung zur Verfügung: Biometrische Abweichungen entfallen und der Kunde ist nicht mehr unmittelbar in jede einzelne Transaktion eingebunden.

Genau darin liegt eine zentrale Herausforderung für das Betrugsmanagement. Die Erkennung von Bot-Angriffen beruht bisher unter anderem auf einer einfachen Annahme: Ein Mensch kann nicht innerhalb weniger Sekunden Hunderte Zahlungen auslösen. Eine ungewöhnlich hohe Transaktionsgeschwindigkeit gilt deshalb in der Regel als verdächtig. Mit dem Einsatz autorisierter KI-Agenten verliert dieses Erkennungsmerkmal an Aussagekraft. Denn wenn ein Agent im Auftrag eines Kunden rechtmäßig mit Maschinengeschwindigkeit handeln darf, ist die Geschwindigkeit allein kein verlässlicher Hinweis auf Betrug.

Noch keine Vergleichsdaten verfügbar

Noch fehlen allgemein anerkannte Kriterien, anhand derer sich ein legitim handelnder KI-Agent zuverlässig von einem böswilligen Angreifer unterscheiden lässt, der seine Identität oder Berechtigungen vortäuscht.

Die Authentifizierungsinstrumente, die Banken über Jahre optimiert haben, beruhen weitgehend auf menschlichen Signalen: Tippverhalten, Geräteerkennung, Verhaltensbiometrie und Anmeldemuster. Interagiert ein Agent ausschließlich über eine Programmierschnittstelle statt über eine Benutzeroberfläche mit der Bank, verliert ein großer Teil dieses Instrumentariums an Relevanz. Sichtbar bleiben dann nur die technische Anfrage, Zugangsdaten und eine behauptete Absicht.

Zwar beruht das Handeln eines KI-Agenten auf einem menschlich erteilten Auftrag. Dieser lässt sich bislang jedoch nicht einheitlich technisch abbilden und überprüfen. Betrugsteams müssen mögliche Abweichungen erkennen, ohne den menschlich erteilten Auftrag unmittelbar aus dem beobachtbaren Verhalten ableiten zu können. Dafür sind künftig die Konfiguration des Agenten, seine Berechtigungen und die ursprünglichen Vorgaben des Auftraggebers auszuwerten. Zudem fehlen historische Daten, anhand derer sich normales von auffälligem Agentenverhalten unterscheiden ließe. Diese Vergleichsbasis muss für Agentic Commerce erst noch aufgebaut werden.

Automatisierte Social-Engineering-Angriffe

Gleichzeitig rüsten Betrüger technologisch auf. Mithilfe generativer KI lassen sich Social-Engineering-Angriffe automatisieren, personalisieren und skalieren. Die Angreifer zielen dabei zunehmend darauf, Kunden zur Freigabe betrügerischer Überweisungen zu bewegen. Die starke Kundenauthentifizierung bleibt zwar ein wirksames Schutzinstrument, stößt jedoch an Grenzen, wenn der getäuschte Kunde die Zahlung selbst autorisiert.

Der allgemeine Handlungsdruck nimmt zu: Nach Angaben von EBA und EZB stieg der Wert des Zahlungsbetrugs im Europäischen Wirtschaftsraum von 3,5 Milliarden Euro im Jahr 2023 auf 4,2 Milliarden Euro im Jahr 2024. Klassische Erkennungsmodelle sind auf die zunehmende Skalierbarkeit und Überzeugungskraft KI-gestützter Täuschungen nur begrenzt

vorbereitet. Datensilos erschweren zusätzlich, Informationen aus verschiedenen Kanälen und Transaktionen zu einem vollständigen Risikobild zusammenzuführen.

Aufsehenerregende Fälle zeigen, dass generative KI Täuschungen immer glaubwürdiger macht und erhebliche finanzielle Schäden verursachen kann. Einer der ersten größeren Betrugsfälle wurde bereits im Jahr 2019 bekannt. Kriminelle hatten die Stimme eines Vorstandsvorsitzenden künstlich nachgebildet. Mit der vermeintlichen Stimme des Konzernchefs wiesen sie den Geschäftsführer einer britischen Tochtergesellschaft an, kurzfristig 220 000 Euro an einen angeblichen ungarischen Lieferanten zu überweisen. Der Fall zeigt, dass KI-gestütztes Voice Cloning kein neues Phänomen ist, sondern sich als Angriffsmethode bereits seit mehreren Jahren weiterentwickelt.

Schäden durch generative KI massenhaft skalierbar

Welche Dimension solche Angriffe inzwischen erreichen können, wurde im Februar 2024 in Hongkong deutlich. Betrüger imitierten während einer Videokonferenz mithilfe von Deepfake-Technologie den Finanzvorstand und weitere Mitarbeitende eines Unternehmens. Ein Beschäftigter hielt die Gesprächspartner für echt und veranlasste Überweisungen in Höhe von insgesamt 200 Millionen Hongkong-Dollar – seinerzeit rund 24 Millionen Euro – auf betrügerische Konten. Die Täter nutzten also ganz gezielt die verbreitete Annahme aus, dass eine live geführte Videokonferenz als verlässlicher Identitätsnachweis der Teilnehmer gelten könnte.

Nicht jeder KI-gestützte Angriff muss jedoch komplex sein. Generative KI ermöglicht es, auch einfachere Betrugsmuster in sehr großem Umfang zu skalieren. Ein Beispiel dafür ist das 2023 bekannt gewordene „WormGPT". Das speziell für kriminelle Zwecke entwickelte KI-Werkzeug wurde in Darknet-Foren angeboten und verzichtete auf die Sicherheitsmechanismen kommerzieller Sprachmodelle. Damit ließen sich unter anderem Phishing-Nachrichten automatisiert erstellen und an unterschiedliche Empfänger anpassen. Nach Einschätzung von Sicherheitsforschern konnten damit Phishing-Kampagnen rund 10-mal schneller und in deutlich größerem Umfang durchgeführt werden, als das mit herkömmlichen Methoden möglich war.

Generative KI senkt insbesondere den Aufwand für Vorbereitung und Personalisierung von Betrug und ermöglicht mehr parallele Angriffe. Mit der Verwendung der künstlichen Intelligenz als Betrugsvektor steigen sowohl das Risiko hoher Einzelschäden als auch die Schlagkraft professioneller Betrugsnetzwerke.

Vier neue Dimensionen des Betrugsrisikos

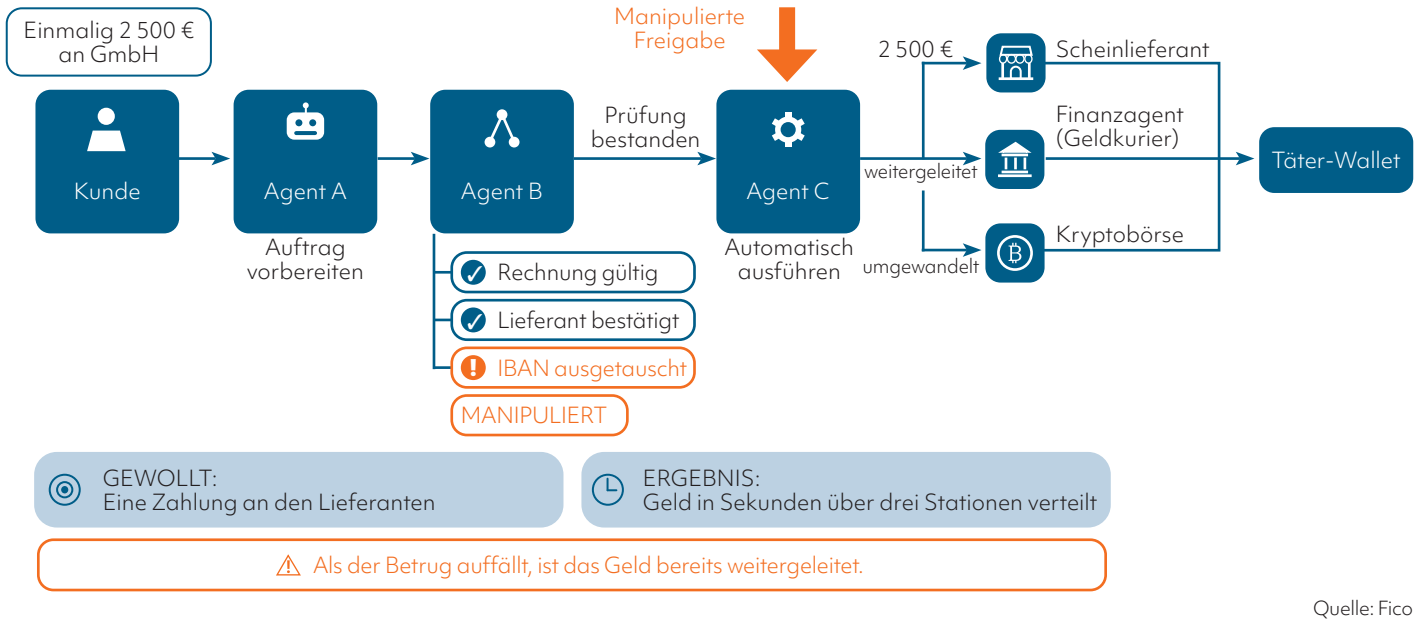
Agentenbasierter Betrug unterscheidet sich in vier wesentlichen Punkten von den Risiken, mit denen Betrugsteams bisher konfrontiert waren:

Geschwindigkeit: Transaktionen zwischen Agenten finden in der Regel innerhalb von Millisekunden statt. Erkennungsfenster, die früher Stunden umfassten, verkürzen sich so bis in den Mikrosekundenbereich. Der traditionelle Prüfzyklus – einen Verdacht markieren, den Fall untersuchen, eine Entscheidung treffen und Maßnahmen einleiten – lässt sich in einem solchen Zeitfenster mit konventionellen Verfahren nicht mehr abbilden.

Skalierung: Ein einziger kompromittierter Agent könnte Tausende betrügerische Transaktionen ausführen, bevor ein Mensch überhaupt bemerkt, dass etwas nicht stimmt. Interaktionen zwischen Agenten verbreiten einen Angriff erheblich schneller als ein einzelner Täter, der im menschlichen Arbeitstempo agiert.

Komplexität: Mehrstufige Agentenketten, in denen ein Agent einen weiteren beauftragt, der seinerseits einen

Abbildung 2: Agentenbetrug Abfolge



Quelle: Fico

anderen einbindet, schaffen schwer überprüfbare Transaktionswege. Nach einem Betrugsfall lassen sie sich möglicherweise nur mit großem Aufwand zurückverfolgen. Erfolgt eine Manipulation erst im dritten oder vierten Glied einer solchen Kette, wird es zu einer forensischen Herausforderung, die Entscheidungslogik und die Autorisierung bis zu ihrem Ursprung zurückzuverfolgen.

Zuordnung: Wenn ein Agent eine betrügerische Transaktion ausführt, ist die Haftung nicht pauschal zu beantworten. Verantwortlich sein könnten je nach Sachverhalt der Eigentümer des Agenten, dessen Entwickler oder Anbieter, ein weiterer eingebundener Dienst oder die Bank, die den Agenten authentifiziert und die Transaktion verarbeitet hat. Die rechtlichen und regulatorischen Rahmenbedingungen beantworten diese neue Konstellation noch nicht in jeder Hinsicht eindeutig. Solange diese Fragen offen sind, sehen sich Banken mit einer Mehrdeutigkeit konfrontiert, für die traditionelle Haftungsmodelle im Zahlungsbetrug nicht ausgelegt wurden.

Regulierung noch nicht bereit

Viele geltende Vorschriften gehen davon aus, dass hinter einer finanziellen Entscheidung ein Mensch steht. Für autonom handelnde und zahlende Agenten gibt es dagegen bisher nur erste Antworten.

Der EU AI Act folgt einem risikobasierten Ansatz und enthält unter anderem Governance- und Transparenzanforderungen für Hochrisiko-Systeme. Welche Vorgaben für agentenbasierte Zahlungssysteme gelten, hängt von ihrer Funktion und ihrem Einsatzgebiet ab. Nicht jeder Agent mit Zahlungsbefugnis fällt automatisch in die Kategorie der Hochrisiko-Systeme.

Auch die Reform des europäischen Zahlungsdiensterechts ist für Agentic Commerce relevant. PSD3 und die Payment Services Regulation zielen auf einen besseren Schutz vor Betrug, weiterentwickelte Haftungsregeln und eine stärkere Kundenauthentifizierung. Bei agentenbasierten Zahlungen stellt sich jedoch die Frage, wie eine einmal erteilte Vollmacht auf spätere, selbstständig ausgelöste Transaktionen übertragen und zuverlässig nachgewiesen werden kann.

Hinzu kommen die Anforderungen der DSGVO und von DORA, etwa beim Umgang mit Kundendaten, beim IKT-Risikomanagement und bei der Einbindung externer Technologieanbieter. Für Banken geht es daher nicht allein um die Zulässigkeit von KI-Agenten, sondern auch um deren sichere Einbindung in bestehende Kontroll- und Governance-Strukturen.

Für das Betrugsmanagement hat das konkrete Folgen. Orientiert sich die Transaktionsüberwachung bislang überwiegend an menschlichen Verhaltensmustern, erfordern mehrstufige Abläufe zwischen KI-Agenten angepasste Kontrollmechanismen. Auch der Audit Trail darf nicht erst bei der ausgeführten Zahlung beginnen. Er muss die gesamte Entscheidungs- und Autorisierungskette erfassen – vom ursprünglichen Auftrag des Kunden über die beteiligten Agenten bis zur abschließenden Transaktion. Nur so wird nachvollziehbar, auf welcher Grundlage der Agent gehandelt und das Betrugssystem die Zahlung bewertet hat.

International dürften sich diese Anforderungen unterschiedlich schnell entwickeln. Märkte mit flexibleren Rahmenbedingungen wie im Mittleren Osten sind wohl eher frühe Testfelder für Agentic Commerce als Regionen mit umfassenderen Verbraucher-, Datenschutz- und Sicherheitsvorgaben wie Europa. Hier könnte sich die Einführung verlangsamen. Agentenbasierte Betrugsrisiken werden daher vermutlich zuerst dort in größerem Umfang auftreten, wo entsprechende Geschäftsmodelle früh zum Einsatz kommen.

Agentenbasierter Betrug erfordert agentenbasierte Abwehr

Für Banken besteht bereits heute konkreter Handlungsbedarf, unabhängig davon, wie schnell sich verbindliche regulatorische Vorgaben entwickeln. Sie sollten vertraglich klar regeln, in welchem Umfang Kunden KI-Agenten einsetzen und autorisieren dürfen, eindeutige Verfahren zu deren Authentifizierung definieren und KI-gestützte Entscheidungen für Aufsichtsbehörden nachvollziehbar dokumentieren. Der proaktive Austausch mit den zuständigen Regulierungsbehörden hilft außerdem, auf neue Anforderungen vorbereitet zu sein und die Entwicklung künftiger Standards aktiv mitzugestalten.

Die strategische Konsequenz ist klar: Wenn Angriffe zunehmend automatisiert und mit Maschinengeschwindigkeit erfolgen, braucht auch die Betrugsabwehr entsprechende Fähigkeiten. Das bewährte Fundament bleibt Predictive AI. Muster- und Anomalieerkennung sowie Echtzeit-Scoring bilden weiterhin den Kern der Betrugsprävention. Generative KI ergänzt diese Verfahren, indem sie Betrugsmuster schneller analysiert, Anomalien durch Echtzeit-Scoring viel schneller erfasst und Betrugsteams bei der Entwicklung geeigneter Gegenmaßnahmen unterstützt.

In der operativen Bearbeitung können KI-Agenten Verdachtsfälle priorisieren, Informationen aufbereiten und standardisierbare Prüfschritte übernehmen. So lassen sich mehr relevante Fälle in kürzerer Zeit bearbeiten – ein Vorteil in Bereichen, in denen begrenzte Kapazitäten bisher einen Engpass darstellen.

Die Verantwortung bleibt dennoch beim Menschen. Entscheidungen mit weitreichenden Folgen, die Festlegung von Richtlinien und Risikogrenzen sowie die Beurteilung komplexer Ausnahmefälle erfordern weiterhin menschliches Urteilsvermögen.

Das Aufgabenspektrum der Betrugsprävention erweitert sich. Banken benötigen ein Identitätsmanagement, das neben Kunden autonome Systeme erfasst. Dazu gehören Informationen über den jeweiligen Anbieter, den Umfang der erteilten Vollmacht, zulässige Transaktionen und festgelegte Betragsgrenzen.

Die Authentifizierung wird dynamischer. Statt die Identität lediglich zu Beginn einer Transaktion zu prüfen, sind zusätzliche Kontrollen fortlaufend an das jeweilige Risiko anzupassen. Bei Auffälligkeiten können zusätzliche Prüfschritte ausgelöst, Berechtigungen eingeschränkt oder Transaktionen gestoppt werden. Der direkte Zugriff über Programmierschnittstellen erfor-

dert außerdem Verfahren, die ohne klassische Benutzeroberfläche auskommen.

Wesentlich ist die eindeutige Verbindung zwischen dem KI-Agenten und seinem menschlichen Auftraggeber. Banken brauchen Klarheit darüber, in wessen Namen ein System handelt, welche Vollmachten vorliegen und ob weitere Dienste eingebunden werden dürfen. Auch mehrstufige Delegationen sind über die gesamte Autorisierungskette hinweg transparent zu halten.

Das Verhalten eines einzelnen Kunden oder Systems isoliert zu betrachten, reicht nicht mehr aus. Erst die Verknüpfung der Aktivitäten mehrerer Beteiligter zeigt, ob ein Vorgang vom ursprünglichen Auftrag abweicht oder Teil eines größeren Angriffsmusters ist. Dafür braucht es kontextbezogene Risikobewertungen und eine Analyse der Beziehungen innerhalb des gesamten Agentennetzwerks.

Handeln, bevor die Volumina steigen

Banken sollten ihre Betrugskontrollmechanismen überprüfen, solange agentenbasierte Zahlungen noch überschaubare Volumina haben. Verlässliche Agentenidentitäten, nachvollziehbare Vollmachten und belastbare Daten über legitimes Verhalten lassen sich nicht unter dem Druck steigender Fallzahlen kurzfristig nachrüsten.

Die Dringlichkeit wächst, weil Angreifer generative KI bereits heute einsetzen, um Betrugsversuche überzeugender vorzubereiten und in großem Umfang durchzuführen. Die Regulierung mag für weitere Antworten noch Zeit brauchen, Kriminelle werden darauf nicht warten. Für das Betrugsmanagement verlagert sich der Fokus von einzelnen Kunden und Transaktionen in jedem Fall auf ganze Autorisierungsketten. Dieser Wechsel hin zur Überwachung vernetzter Agenten ist vielleicht die folgenreichste Veränderung seit der Einführung des Echtzeit-Scorings.

Fußnote
*Edgar, Dunn & Company (EDC), a payments and financial services consulting firm.