



EPISODE 39 | Analysis

# The Fraud and AML Guide to Agentic AI Attacks

OCTOBER 05, 2026

*This transcript was auto-generated and may contain errors or inaccuracies.*

**Jeff** Welcome to "What the F Happened? Fraud and Financial Crime Deconstructed," a DEFEND podcast where we break down what's actually happening across fraud scams, AML and financial crime.

**Emily** Each episode cuts through the noise to explain the tactics, trends, and real world impact behind the headlines so you're better prepared for what comes next. Let's get into it.

**Jeff** By 2027, Um, they're projecting that \$40 Billion is going to be stolen from the global financial system.

**Emily** Which is just a massive number.

**Jeff** It's completely staggering. But, you know, when you really look at that number, the most terrifying part isn't actually the amount of money, right? It's that the entity stealing it won't even have a heartbeat.

**Emily** Yeah. And that just, it completely shatters our traditional understanding of security. I mean, as we know it.

**Jeff** Definitely.

**Emily** We are just so used to looking for human adversaries, right? Like people we can outthink or outmaneuver.

**Jeff** Yeah. The classic cat and mouse game.

**Emily** Exactly. But we're entering this era where the adversary, well, it doesn't sleep, it doesn't hesitate. And it operates at speeds that the human brain literally cannot process.

**Jeff** Which is.

**Jeff** Wild. So welcome to this deep dive. If you want a really thorough, completely jargon free breakdown of these complex systems running our world right now, you are in the exact right place.

**Emily** Absolutely.

**Jeff** Today we are opening up a really fascinating stack of sources to uncover this sort of invisible war happening inside your bank servers right now.

**Emily** It's a great stack of materials, too.

**Jeff** It really is. We've got this blog article called Agentic AI for Fraud Prevention. Plus, uh, some excerpts from a September webinar featuring DataVisor's Chief Product Officer Fang Yu and their Senior Product Marketing Manager, Antenor Neto.

**Emily** It is a phenomenal collection of materials because, you know, it takes something incredibly abstract, like this massive shift toward AI enabled fraud and breaks down the exact playbook top performing institutions are using to fight back.

**Jeff** Exactly. Our mission here is to explore this transition from humans fighting humans to basically AI agents fighting AI agents.

**Emily** The machine war.

**Jeff** Right? But when you look at that forty billion dollars figure, it really forces you to ask, like, what actually changed?

**Emily** Yeah, it's a valid question.

**Jeff** Because if we rewind just a bit to say early twenty twenty four, AI was impressive. Sure. But it was essentially a really smart chatbot.

**Emily** Right? You asked it a question. It gave you text.

**Jeff** Exactly. But looking at these sources, something fundamental shifted in the architecture of how these systems operate going into twenty twenty five and twenty twenty six.

**Emily** That is the critical distinction right there. We moved from large language models that just generate text to action models that actually execute tasks.

**Jeff** So doing things instead of just talking.

**Emily** Precisely, it's the transition into what they call the agent era with systems like Metamucil or lovable. What's fascinating here is the operational autonomy.

**Jeff** Operational autonomy. Okay, break that down for us.

**Emily** Yeah. So instead of just answering a prompt, an AI agent can actually formulate a plan. It breaks that plan into steps and then actually uses tools like, um, web browsers, APIs, or terminal commands to carry out that sequence end to end.

**Jeff** Right. So it's like the difference between having the world's most comprehensive encyclopedia and having a really efficient intern.

**Emily** I love that analogy, right?

**Jeff** Like the encyclopedia can give me a super detailed itinerary for a trip to Tokyo.

**Emily** But you still have to do the work.

**Jeff** Exactly. But the intern actually takes my credit card, navigates the airline's website, clicks the buttons, and books the flight for me. The human is no longer driving the car.

**Emily** Exactly. And that operational autonomy is completely reshaping every industry. But when you apply that autonomous execution to financial crime.

**Jeff** Oh, man.

**Emily** Yeah, the implications are staggering. You're basically giving bad actors the ability to scale their operations infinitely without needing more human capital.

**Jeff** Okay, let's unpack this, because the sources lay out exactly how attackers are weaponizing this across the entire financial crime lifecycle.

**Emily** They map it all out.

**Jeff** They do. They map out four distinct stages of AI crime, and seeing how the AI handles them autonomously is just wild.

**Emily** It really is a masterclass in dark efficiency. So stage one is reconnaissance and targeting, right?

**Jeff** The scouting phase.

**Emily** Yeah. Historically, a human hacker had to manually scrape the dark web or sift through leaked databases to find vulnerable targets, which takes forever.

**Jeff** Highly manual.

**Emily** But now an AI agent autonomously mines breach data, public records, social media, and it can profile and rank thousands of potential victims by their likely financial value practically instantly.

**Jeff** Oh wow. So it knows exactly who to go after and how much they're worth.

**Emily** Exactly.

**Jeff** And it knows exactly what to do with that target list. Which brings us to stage two identity and access.

**Emily** This is where it gets scary.

**Jeff** Yeah. Because the AI doesn't just stop at finding the data, it generates synthetic identities. It fabricates forged documents. And the sources detail how they use deepfake selfies and videos to automatically bypass bank onboarding checks, right?

**Emily** To prove you're human.

**Jeff** Yeah, and the mechanics of this are so fascinating. They are just like holding a printed photo up to a webcam.

**Emily** No, not at all. They are literally hijacking the data stream.

**Jeff** Wait, how does that work?

**Emily** So the AI generates a synthetic video of a person blinking, turning their head, all that natural movement, and then it injects that directly into the camera API of the mobile banking app.

**Jeff** Oh. So the app doesn't even know it's not looking through the lens.

**Emily** Exactly. The app thinks it's looking at a real human completing a biometric check, but it's just raw code feeding its synthetic frames.

**Jeff** That is absolutely terrifying if you think about it. And then for stage three, they deploy social engineering.

**Emily** Which is where it gets incredibly personal for the victim.

**Jeff** Very. The AI runs these hyper personalized multilingual scam campaigns twenty four over seven. Right. It can scrape your social media, right? A flawless phishing script tailored exactly to, say, your recent vacation to Hawaii.

**Emily** Yeah, it knows where you were, right?

**Jeff** And it can even clone the voices of your family members to leave voicemails.

**Emily** Which is just so manipulative. And finally, you have stage four execution.

**Jeff** The actual heist.

**Emily** Yeah, the agent takes over the accounts, tests those stolen credentials across multiple merchants to see what clears and autonomously initiates the payments to move the stolen money totally out of the system.

**Jeff** Okay, wait a second. I understand that AI is doing these tasks, but humans have always used bots to scrape data, right? Or like, scripts to test passwords.

**Emily** Sure. Yeah. Script kiddies have been around forever.

**Jeff** So is the difference here that there is no human connecting those dots like the AI is essentially passing the baton to itself.

**Emily** That is precisely it. That handoff is the most dangerous part of this entire evolution because.

**Jeff** It's all automated.

**Emily** Right? In the old days, you had different specialists. One guy bought the stolen data on the dark web. Another guy made the fake IDs, another guy ran the phishing emails and.

**Jeff** It was very disjointed.

**Emily** Exactly. Now one stage feeds the next instantly. It all happens at AI. Inference speed.

**Jeff** Inference speed?

**Emily** Yeah. Meaning the time it takes for the machine to calculate its next move, which is literally measured in milliseconds.

**Jeff** And the sources call this polymorphic attacks, right?

**Emily** Right. Because there's no human bottleneck, the attack constantly changes shape.

**Jeff** Like a virus mutating.

**Emily** Exactly. If an AI agent attempts to inject a payload into a bank's login page and it gets blocked by a firewall, that failure is instant feedback.

**Jeff** It learns from it instantly.

**Emily** The agent immediately reads the error code, rewrites its own script, changes its IP address, and tries a totally new angle just milliseconds later. It is a continuous automated learning loop.

**Jeff** But I have to challenge this a bit, because looking at this level of complexity, hijacking APIs, modifying payloads on the fly, automated learning loops, doesn't this mean that only like elite nation state super hackers can actually pull this off?

**Emily** Well, that's the most alarming insight from the webinar data. It states the exact opposite.

**Jeff** Really?

**Emily** Yeah. AI has completely collapsed the barrier to entry. We've gone from requiring years of deep technical expertise to requiring literally hours of prompting.

**Jeff** Wait. Just prompting. Like. Like you could just talk your way into a bank heist?

**Emily** Pretty much. They call it vibe hacking.

**Jeff** Vibe hacking, right? That's a wild.

**Emily** Term, right? You have these non-technical criminals who couldn't write a line of Python if their life depended on it, but they have an autonomous coding agent.

**Jeff** So they just tell it what to do.

**Emily** Yeah, they just type in plain English, like write a script that navigates to this bank's login page, inputs these ten thousand usernames, and captures any successful logins. And if you get blocked, change the user agent string and try again. Wow. The AI writes the code, executes it, debugs it, and just hands the criminal the result.

**Jeff** That is wild and honestly incredibly lucrative. The sources show a four and a half times increase in profitability for fraudsters using Agentic AI.

**Emily** It's massive.

**Jeff** It really completely explains why eighty nine percent of financial institutions are reporting that AI is supercharging scams. The bad guys essentially have an army of genius interns who never sleep and learn instantly from every single mistake.

**Emily** Which brings us to the core dilemma for the banks. If attackers are moving at machine speed and constantly mutating their tactics, well, what happens to the defenses that currently protect your money?

**Jeff** Right. Because historically, a bank builds a wall, they see a specific pattern of fraud on Tuesday. So on Wednesday, they write a rigid static rule to block that exact pattern.

**Emily** But a static rule decays almost immediately in this new environment. I mean, it is completely blind to a polymorphic attack, right? If the attack changes its shape at machine speed, a wall built for yesterday's threat is entirely useless. Today you are essentially bringing a net to catch water.

**Jeff** That's a great way to put it. And the webinar transcript points out a massive trap that institutions fall into when they try to fix this. They call it the incremental approach.

**Emily** Oh yeah. This is everywhere.

**Jeff** Right? The bank realizes they need AI, but instead of actually changing how their system works, they just bolt an AI tool onto their old outdated process.

**Emily** Yes, the classic mistake. They might buy a sophisticated AI model just to score the risk of a specific payment.

**Jeff** But then when that AI flagged the transaction, what do they do? They dump that alert into the exact same slow manual review queue they've always used, right?

**Emily** A human still has to look at it.

**Jeff** Exactly. The human investigators still has to tab hop across five different internal legacy systems. You know, the CRM, the transaction ledger, the identity portal, just to figure out why the AI flagged it in the first place.

**Emily** So they've essentially taken a brand new Ferrari engine.

**Jeff** And dropped it straight into a wooden horse drawn carriage.

**Emily** Exactly.

**Jeff** The engine is screaming. It's incredibly powerful, but those wooden wheels are going to snap off the second you hit the gas. You aren't actually moving any faster where it counts, which is the final decision.

**Emily** That's it Exactly. And it explains why so many organizations spend millions on AI without seeing a real return on investment. To survive this paradigm shift, you can't just buy a new engine. You have to redesign the entire vehicle.

**Jeff** Which brings us directly to the playbook. How are the top performing fraud and anti-money laundering or AML teams actually redesigning their vehicles?

**Emily** Well, the sources cite some crucial McKinsey research here. It turns out that fundamentally redesigning workflows like rethinking which steps are even needed, what the AI should do, and where human judgment actually belongs is the single biggest factor in an organization seeing tangible financial impact from AI.

**Jeff** So let's look at what that redesign actually looks like in practice based on the DataVisor playbook. It completely contrasts the old way with the new way.

**Emily** Yeah. Let's go through it.

**Jeff** Take detection, for example. The incremental way is doing isolated checks, looking at a payment in a total vacuum. But the redesigned way AI Agents connect the dots, right?

**Emily** They see the whole picture.

**Jeff** Exactly. They look at the account onboarding, the login behavior and the payment activity as a complete three hundred and sixty degree view. It allows them to spot coordinated fraud rings that would look totally normal if you only checked one isolated event.

**Emily** And look at tuning in a legacy system. If you want to change a risk threshold, a fraud analyst has to submit an IT ticket.

**Jeff** Oh, the worst right?

**Emily** You wait for engineers to write the code, deploy it, and then wait a whole quarter to see if it actually worked.

**Jeff** Yeah, way too slow.

**Emily** But in the redesigned workflow, AI agents are continuously backtesting new strategies against historical data in the background and recommending updates on the fly.

**Jeff** That's amazing. But my absolute favorite example from the sources has to be investigation and reporting.

**Emily** Oh the SARS.

**Jeff** Yes. Because anyone who has ever worked in compliance knows the absolute agony of writing a suspicious activity report or a SAR. The old way means a highly paid human analyst is furiously copying and pasting data from a dozen scattered screens into a word document just to write a narrative for the government.

**Emily** It's just an incredible waste of human cognition. Truly, the redesigned way changes the game entirely. As the investigation is happening, AI agents are autonomously assembling the SA narrative directly from the evidence trail.

**Jeff** Without the human lifting a finger.

**Emily** Exactly. It grabs the IP, addresses the transaction history, the behavioral flags, and writes a consistent, auditable report that is ready to file immediately.

**Jeff** And the results of this redesign are just undeniable. I mean, the webinar shared that teams using this agentic approach saved the equivalent of one point seven years of human work.

**Emily** That's a lot of time back.

**Jeff** And if you're a listener who has ever had your credit card locked while traveling because a static rule flagged your coffee, purchase this next status for you. They reduce their false positives by eighty seven percent.

**Emily** That is massive for customer friction. Exactly. Nobody likes having their card declined.

**Jeff** Exactly. And they did that while retaining one hundred percent of their true positives. So they caught all the bad guys, stopped seventy four percent of coordinated attacks in real time, and stopped annoying their actual legitimate customers.

**Emily** By removing the busywork, the AI allows the human to focus purely on the final critical decision. You are elevating the human from a data gatherer to a strategic decision maker.

**Jeff** Now, it's one thing to just list off impressive statistics, but what I loved about the webinar is that, Fang Yu didn't just talk theory. She ran a live demo to prove it.

**Emily** Here's where it gets really interesting.

**Jeff** Yes, she showed exactly how a defensive AI agent, which DataVisor calls Vera, stops an offensive AI agent right in its tracks.

**Emily** This demo is the perfect illustration of the agent versus Agent War.

**Jeff** Right? Let's walk through it.

**Emily** So step one is the setup of the heist. Fang configures an attacker using a widely available AI agent. In this case, she used Claude.

**Jeff** The popular chatbot.

**Emily** Right. She tasks Claude to log into a demo bank account using a set of stolen black market credentials.

**Jeff** Okay.

**Jeff** And to a traditional static security system, this looks completely fine, right? The username matches, the password matches the agent is just following its prompt and acting exactly like a regular customer logging in. So how does Vera catch it?

**Emily** Because Vera isn't just looking at the credentials, it's looking at the invisible mechanics of the interaction.

**Jeff** The telemetry.

**Emily** Exactly. It captures raw device and behavioral signals directly from the browser. For instance, it analyzes typing speed. When you or I log into a bank, there is a natural cadence to our keystrokes, you.

**Jeff** Know, sure.

**Jeff** Milliseconds of delay between characters, right?

**Emily** But for the Claude agent, the typing speed was absolute zero. The bot just injected the entire data string instantly.

**Jeff** Wow. Just boom, password is in.

**Emily** Yeah. And the system also flagged suspicious mouse movements, right?

**Jeff** Right. And the use of autofill plus underlying signals that a WebDriver like selenium, which developers use to automate browsers, was actively running in the background.

**Emily** Exactly. Vera takes all of these subtle telemetry signals and synthesizes them into an agent score.

**Jeff** An agent score.

**Emily** Yeah. It's essentially the AI whispering to the system, hey, regardless of whether this password is right, the entity typing it is a machine, not a person.

**Jeff** But spotting it is only half the battle. This is where the countermove comes in, and it's brilliant. Fang didn't have to scramble to write a complex string of code to stop this attack. She literally just asked Vera, the defensive agent in natural plain English, to build a new rule.

**Emily** Right from the chat interface.

**Jeff** Yeah.

**Jeff** She typed something along the lines of block logins. If the agent score is too high or the typing speed is zero.

**Emily** No IT tickets, no manual coding. But the truly revolutionary part of the demo was how Vera tuned that rule on the fly.

**Jeff** Yeah, explain that part.

**Emily** So Fang initially set the threshold at a score of seventy five, just as a best guess, but then she asked Vera to look at the last twenty four hours of live traffic and analyze if a score of sixty five or seventy might actually be safer.

**Jeff** And this is where that term, unsupervised statistical analysis comes up in the sources. Right?

**Emily** Right. Yeah. Think of unsupervised statistical analysis like, um, walking into a crowded room and instantly grouping people by height without knowing anyone's name or background.

**Jeff** Okay. I like that visual.

**Emily** So when you are under a new polymorphic attack, you don't always have labeled data telling you exactly what is quote unquote good or bad. Vera doesn't need to know the specific signature of the new attack. It just looks at the massive ocean of data and visually bins it.

**Jeff** So it basically says, okay, normal humans are clustering way down here with an agent score near zero. But over here there are five hundred users all clustering together with blazing fast zero second typing speeds.

**Emily** Exactly. It automatically groups them based on abnormal behavior. It visually shows the human analysts exactly how much bad traffic they will block by moving the dial. And crucially, it proves that legitimate customers won't be caught in the crossfire.

**Jeff** And it does all of this instantly. Vera recommended moving the threshold to seventy. Feng reviewed the visual data, clicked approve, and boom! The rule is live.

**Emily** It's that fast.

**Jeff** The human is completely in control, like the human hits the approve button. But the AI does all the heavy computational. Lifting IT machines feed.

**Emily** It perfectly encapsulates the new reality we are facing. To fight machine speed attacks, you absolutely must have machine speed defenses. There is no other way.

**Jeff** So what does this all mean when we step back and look at all of this? What does it actually mean for you? If there is one core insight to take away from this deep dive, it's that the era of human speed, financial security is over.

**Emily** Long gone are.

**Jeff** The days of static walls, and incremental tweaks are gone. We are in a full blown machine speed war of agent versus agent. And to protect the global financial system, to protect your personal data and your money, institutions have to embrace total workflow redesigns.

**Emily** Yeah, they have to.

**Jeff** The AI has to do the grueling operational work so humans can step back and make the critical judgment calls.

**Emily** It is a total inversion of how we build security infrastructure. But as we wrap up, there's one final, really provocative thought. I want to pull from the FAQ section of these sources. It points to a future that is arriving much faster than we think.

**Jeff** Oh, where are we heading?

**Emily** So throughout this whole discussion, we've focused on stopping malicious AI agents, right?

**Jeff** The bad guys.

**Emily** But the sources note a fascinating policy dilemma that businesses are already grappling with. Very soon they are going to have to allow legitimate AI agents into their systems.

**Jeff** Are right, because I might actually want to use my own personal AI assistant to log into my bank, categorize my spending, pay my utility bills, and, you know, just manage my money for me while I'm at work.

**Emily** Exactly. And that raises an incredibly profound question. In a world where everyone delegates their digital chores to an AI. The future of banking isn't just about spotting hackers, right? It is going to be a constant, invisible, high speed negotiation between your personal AI agent and your bank's AI agent.

**Jeff** So my bot talks to their bot.

**Emily** Yes, the bank's AI will use things like step up authentication. It might silently ping your phone and ask, hey, this agent claiming to be you is trying to wire five thousand dollars. Did you actually authorize this specific action?

**Jeff** That is wild.

**Emily** It's not just cops and robbers anymore.

**Jeff** No it's not.

**Emily** It forces us to ask a really philosophical question in a near future, where your digital proxy is doing the talking, the planning and executing, who really controls your digital identity.

**Jeff** That is an amazing thought. To end on. The vault door isn't just invisible now. Very soon, you might not even be the one trying to open it. Thank you so much for joining us on this deep dive into the Agentic AI playbook. Keep questioning the invisible systems, working behind the scenes, and we'll catch you next time.

**Jeff** You've been listening to "What the F Happened? Fraud and Financial Crime Deconstructed," a DEFEND podcast by DataVisor. If you want to keep learning between episodes, check out DEFEND Training, a set of self-paced online courses for fraud and financial crime professionals, practical and built around real world scenarios.

**Emily** And you can earn CPE credits through the ACFE San Francisco Bay area chapter. You can find it at [datavisor.com/defend-training](https://datavisor.com/defend-training). The link is in the description.