EPISODE 35 | Analysis

Scaling Agentic AI in Fraud & AML

SEPTEMBER 08, 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 So, um, imagine you hire like fifty absolute geniuses to run your bank's fraud department.

Emily Okay, that sounds pretty good so far, right?

Jeff They're they're literally the smartest people on the planet. Yeah, but, uh, there is a catch.

Emily There's always a catch.

Jeff Always. So none of them speak the same language. Um, they don't have a manager. They're all sitting in totally separate rooms.

Emily Oh, man.

Jeff And here's the kicker. They all have the unilateral power to just freeze a customer's bank account.

Emily That is a recipe for complete disaster, right.

Jeff And in the world of artificial intelligence, that scenario is not like some abstract hypothetical anymore. It's actually a very real, very dangerous phenomenon. And it's called orchestration sprawl.

Emily Yeah. And it's, it's the exact trap that a lot of countless really financial institutions are falling into right now is they kind of rush to adopt AI because we are way past the era of just asking, you know, whether an artificial intelligence can write a fraud rule or flag a suspicious wire transfer.

Jeff Like the basic.

Emily Stuff. Exactly. The technology has matured. The new crisis is what happens when you have an entire fleet of these specialized AI agents trying to operate all at the same time across a massive, highly regulated financial system.

Jeff And that is exactly the massive challenge that we are unpacking on this deep dive today. We're looking at a really fascinating framework that was laid out by the fraud prevention company DataVisor.

Emily Yes.

Jeff Specifically, we are diving into their recent trusted scale webinar and a companion blueprint they put together. It's all about moving from one AI agent to many, right? So our mission today is really to figure out how organizations can graduate from having a single, isolated AI assistant to deploying a fully coordinated multi-agent digital workforce, right?

Emily And how to do that without unleashing total operational chaos.

Jeff Exactly. Total chaos. So where do we even begin with this?

Emily Well, to get there, we really have to break this persistent illusion that kind of dominates the corporate conversation around AI.

Jeff Okay. What illusion is.

Emily That? So a lot of leaders are still chasing what we could call the solo co-pilot myth.

Jeff The solo co-pilot.

Emily Yeah. They are looking for one omnipotent digital brain that can just, you know, be plugged into their data stream to silently manage all their risk. And in the context of fraud and anti-money laundering or AML, that single magical bot, it just doesn't exist.

Jeff I have to admit, that was kind of my initial assumption too. Oh, sure. When you hear about these massive language models, you just assume you feed it the data, it finds the bad guys and you call it a day, right?

Emily You just flip a switch.

Jeff Yeah. But looking at the actual architecture of how financial crime is fought, um, it is this incredibly interconnected, multi-stage pipeline. I mean, the job doesn't end just because a model flagged a weird transaction.

Emily No, far from it. If we map out the actual life cycle of fraud and AML operations, it's a very complex sequential journey. Okay, so it begins with raw signals, you know, millions of data points streaming in every second, right? Those signals have to be processed by a detection mechanism, and then detection generates specific decisions or alerts.

Jeff When an alert is really just the beginning of the process.

Emily Right? Exactly. That alert has to trigger an investigation. And the investigation then yields findings that must be transformed into formal regulatory reporting.

Jeff Wow.

Emily And then finally, the outcomes of that entire process have to loop all the way back around to optimize the detection models for tomorrow.

Jeff Okay, let's, let's unpack this because, um, it really reminds me of a brilliant but completely isolated detective.

Emily Oh, I like that. Right?

Jeff Like you had this genius investigator who cracks a massive money laundering case. He figures out exactly how the criminal enterprise works, but then he just completely forgets to tell the precinct.

Emily It just keeps it to himself.

Jeff Yeah. He doesn't file any paperwork. He leaves all the evidence locked in the trunk of his car. Right, right. So the case might technically be solved in his own head, but the system still fundamentally fails because the workflow is broken.

Emily That's a that's a perfect analogy.

Jeff But let me push back a bit here, because if these AI models are getting so unbelievably powerful, why can't we just build one giant super agent to do that entire life cycle? Like, why do we even need specialized agents at all?

Emily Well, that is the trillion dollar question, really. And the answer comes down to the friction between processing power, security, and just the fundamentally different nature of the tasks involved.

Emily Mhm.

Emily Well, think about the difference between sifting through, say, ten million raw credit card swipes in a fraction of a second to find a velocity anomaly versus the task of drafting a highly nuanced, legally compliant suspicious activity report for a human regulator to actually sit down and read.

Jeff Oh, I see those require totally different skill sets. Exactly. Like one is heavy math and speed, and the other is nuance in language and legal compliance.

Emily Precisely that if you try to engineer one single AI agent to handle both extremes. You end up with a system that is far too slow to catch real time fraud, right? And far too generalized to write a competent legal report.

Jeff That makes total sense.

Emily Furthermore, and this is big, you introduce a catastrophic security vulnerability.

Jeff Oh. Because of access?

Emily Yes.

Emily To do everything that single agent would require unrestricted access to every database, every customer record, and every action protocol across the entire institution. Oh, wow. And in a regulated environment, you absolutely cannot have a single point of failure with God level access.

Jeff Ah, so specialization is really a structural necessity just to keep the system fast, accurate and secure.

Emily Yes, absolutely.

Jeff We basically need the math agent and the investigation agent and the reporting agent.

Emily Correct.

Jeff But that brings up the obvious next hurdle. If you have all these highly specialized, siloed agents doing their own specific jobs.

Emily Mhm.

Jeff Um, how do they actually communicate safely?

Emily That's the challenge.

Jeff Like, is there some standardized language they have to speak to prevent dropping the ball?

Emily There has to be. And it is arguably the most critical component of a multi-agent ecosystem. Okay. A system is only as strong as its handoffs. Moving data from a detection bot to an investigation bot isn't just about forwarding an email, right?

Jeff It's more complex, much more.

Emily It requires what the industry calls a handoff contract.

Jeff The handoff contract. Okay. So I was originally picturing a relay race, you know, where a baton is being passed. Sure. But based on the complexity here, it sounds much more like a surgical operating room.

Emily Oh, interesting. How so?

Jeff Well, like the triage nurse in the emergency room is basically your detection agent. Okay. Yeah. When they hand a patient over to the attending surgeon, who is your investigation agent, they don't just shove the patient through the double doors and say, hey, this guy is sick. Good luck.

Emily Right. They lose their license.

Jeff Exactly. They provide a highly structured, standardized medical chart.

Emily That is a phenomenal way to visualize it. Yes. If a detection agent spots a sophisticated fraud ring and just passes a generic high risk alert to the downstream investigation agent, that second agent is essentially walking into the operating room blind, right?

Jeff It doesn't know what it's looking at exactly.

Emily It has to start its analysis completely from scratch, which wastes computing power and time. So to prevent that, a governance hand-off contract dictates five explicit questions that every single agent to agent transfer must answer.

Jeff Okay, let's walk through those using that medical chart analogy, because it really kind of clarifies the mechanism for me.

Emily Sure, let's do it.

Jeff So the first requirement is context.

Emily Yes. Context determines exactly what information is being transferred. Okay. In our fraud scenario, are we just passing the alert or are we passing the full entity history, the linked accounts and the IP addresses that triggered the flag? Right. The downstream agent needs the full history of the event to make a logical next step.

Jeff Which leads to the second requirement on the chart, which is evidence. Mhm. And I guess it's not enough to just say this is fraud.

Emily Never. We cannot tolerate unexplained AI hallucinations in finance.

Jeff Yeah, that would be bad.

Emily The hand off contract forces the upstream agent to show its math. What source events, specific data points and internal logic prompts led to this conclusion.

Jeff So the investigation agent needs to see the vital signs and the lab results, not just the diagnosis. Okay. And the third piece is authority. And this one feels particularly crucial for security.

Emily It is the ultimate guardrail, honestly. Authority explicitly defines what the receiving agent is actually permitted to decide or execute based on the file it just received. Oh, yeah. Just because an investigation agent receives a comprehensive fraud dossier does not necessarily mean it has the authority to permanently close a customer's checking account.

Jeff Oh I.

Emily See. It's authority might be strictly limited to just drafting a summary and freezing outbound transfers for twenty four hours.

Jeff Right. And if it wants to do more than that. We hit the fourth requirement, which is approval. Yes. So the contract has to specify exactly where in the workflow a human being is required to step in and basically look things over.

Emily You always need to map the human in the loop checkpoints. Where does a human need to review, edit, approve or reject the AI's proposed action?

Jeff Okay. Makes sense.

Emily And finally, the fifth requirement is auditability. Auditability, right? What data is being permanently recorded about this specific handoff for future review regulatory compliance and and machine learning.

Jeff Because if the regulators show up like six months later asking why an account was frozen, you can't just shrug and say, well, the bots told us to.

Emily Yeah, they won't accept that answer.

Jeff You need the entire immutable chart of that digital transaction.

Emily Exactly.

Jeff Okay, that makes total sense in theory. But let's look at the reality of the tech boom we're in right now. Okay. What happens when a bank buys the hype? They scale up way too quickly. They ignore all this strict contract stuff and just start dropping AI bots into every single department.

Emily Well, that is when you trigger the orchestration sprawl we talked about at the very beginning, right?

Jeff The scroll.

Emily And what's fascinating here is that a highly successful, isolated pilot program can actually mask the incoming disaster.

Jeff Wait, really? How?

Emily Well, a single AI agent might work flawlessly in a sandbox environment. But when you scale up to dozens of agents without a shared operating model, the failure modes are very aggressive and they are entirely predictable.

Jeff Okay, let's make those failure modes tangible because they really illustrate why this orchestration is so hard. Sure, the first risk mentioned is context loss.

Emily Right? So imagine an agent handling account takeovers. It flags a suspicious login from a foreign country. Okay. If it suffers context loss during the handoff, the downstream agent only sees a bad login. Oh, it doesn't see that this particular customer travels to that specific country every single summer for the past ten years. Oh, no. Right. So because the historical context was dropped between the silos, the system incorrectly freezes a loyal customer's account, which creates a terrible user experience.

Jeff That is bad. Yeah. So the second failure mode is conflicting decisions, which just sounds like an absolute nightmare for a compliance officer.

Emily Oh, it is the definition of operational chaos. You could easily have an AI agent in the credit card vision, operating against an entirely different set of risk thresholds than an AI agent in the mortgage division. So one bot is actively improving a massive loan increase for a customer, while another bot at the exact same institution has just flagged that same customer for suspected money laundering.

Jeff Oh my gosh, you're just fighting yourself at that point.

Emily Exactly.

Jeff Okay, I'll take it a step further with the third risk, which is permission drift. Mhm. This sounds like an agent somehow quietly promoting itself.

Emily Yeah. Permission drift happens when complex systems become tangled and an agent inadvertently gains the ability to access or modify more sensitive data than its actual role requires. Wow. Without strict authority boundaries, a bot meant only to read transaction history might find a back door to start altering customer communication preferences.

Jeff That is terrifying.

Emily It is. And add to that the fourth risk, which is evidence gaps.

Jeff Okay. Evidence gaps.

Emily This is when outputs simply cannot be traced back to their source events. The bot makes a massive claim, but the audit trail is broken, rendering the decision totally unusable for regulatory purposes.

Jeff Which defeats the whole point. Right? And the final, predictable failure is cue in latency risk.

Emily Because a multi-agent system is deeply interconnected. If just one specialized agent say, um, the bot responsible for translating foreign transaction notes, okay, if it becomes slow, unavailable, or starts churning out low confidence results, it becomes a bottleneck.

Jeff Oh I.

Emily See. It paralyzes the entire end to end workflow. The system's speed is dictated by its slowest, most confused agent.

Jeff So if the core issue isn't whether one bot works in a vacuum, but how a whole, you know, society of bots responds when things inevitably break or delay or drift. Right? How do you manage that? Like if the problem is fifty geniuses running wild without a manager, what does the manager actually look like?

Emily Here in enterprise architecture, that manager is called the AI control plane.

Jeff The AI control plane?

Emily Yes. It is the central governing mechanism that transforms a chaotic collection of individual AI tools into a dependable shared operating model.

Jeff Okay, I have to stop you there. I here control plane. And when you look at the framework, it talks about building four distinct layers and enforcing five foundations of governance. Honestly, it sounds incredibly heavy.

Emily It does sound like a lot.

Jeff It sounds like we are building a massive, slow moving digital department of motor vehicles.

Emily The DMV of AI.

Jeff Exactly. Like, aren't we just creating a giant bureaucracy that completely kills the agility and speed that we bought the AI for in the first place?

Emily Well, it is completely natural to view governance as friction. I heard that a lot. Yeah, but think about the brakes on a formula One race car. Okay. The brakes are not there to make the car go slow. The brakes are there so the driver can confidently take a corner at one hundred and fifty miles per hour without crashing into a wall.

Jeff Oh, wow. I really like that.

Emily In financial crime, moving fast and breaking things is just not an option. You are dealing with highly sensitive, personally identifiable information, customer life savings, and severe federal penalties.

Jeff I see your point. The governance actually enables the speed because without it, you're basically just driving blind.

Emily Exactly.

Jeff So how does this control plane actually function as those F1 brakes?

Emily Well, it creates a structured hierarchy. At the bottom you have the data layer, which is your raw signals and events. Okay. Above that sits the agent layer, which is where your specialized digital workers actually live and analyze.

Jeff Right to bots.

Emily The third level is the orchestration layer. And this is where the magic happens. Okay, this layer handles the routing. It enforces the hand off contracts and manages permissions. Got it. And finally, at the very top, you have the human and regulatory interface, which provides the dashboard for ultimate oversight.

Jeff But the architecture alone isn't enough, right? The framework outlines five foundations of governance that have to be baked into that orchestration layer to prevent the F1 car from crashing.

Emily Yes, those foundations are the bedrock of systemic trust. Number one is identity and access.

Jeff Which means.

Emily It is not just about logging in, it's about cryptographic certainty regarding who or what other system has the authority to invoke an agent, change its prompt, or publish its results.

Jeff Okay, that makes sense.

Emily Number two is context management, which we discussed earlier. The control plane acts as a persistent memory bank, ensuring that the entire narrative of a transaction follows the file from detection all the way to reporting.

Jeff Right. So we don't lose the plot halfway through.

Emily Exactly.

Jeff And the third foundation is policy enforcement, which draws a hard, unbreakable line between what an AI agent is allowed to recommend versus what it is allowed to actually execute.

Emily Yes, that boundary is non-negotiable. The fourth foundation is observability. You know, you cannot manage what you cannot see. True. The control plane must constantly monitor the system for latency spikes, quality degradation, and exception rates. But more importantly, it monitors for model drift.

Jeff Okay, let's pause and define model drift for a second, because I think that's a really fascinating vulnerability. Sure. That's when the AI doesn't break, but the world around it changes.

Emily Exactly. An agent that was, say, ninety nine percent accurate at spotting wire fraud last month might suddenly drop to sixty percent accuracy today. Wow. Not because the code changed, but because the fraudsters invented a brand new tactic that the model hasn't learned yet. Oh, why? Observability catches that drift in real time. And that requires the fifth Foundation, which is change management.

Jeff Okay. Change management.

Emily When you spot that drift, you need robust testing environments, full auditing capabilities and, crucially, the ability to instantly roll back an agents update if a new prompt causes unexpected chaos in production.

Jeff Right. And by enforcing these foundations, you ensure data isolation so customer data never leaks into the wrong bots memory. Exactly. So here's where it gets really interesting, because looking at an abstract architecture is one thing, but DataVisor actually provided a concrete real world application of this using their AI agent, which is called Vera.

Emily Yes. Vera.

Jeff And they applied this entire multi-agent orchestration model to tackle synthetic identity fraud.

Emily Synthetic identity fraud is arguably the most complex challenge in modern banking.

Jeff Really? Why is.

Emily That? Well, fraudsters are not just stealing a single person's credit card. They are harvesting a real social security number from a child, combining it with a fake name, a fabricated address, and a burner phone number. Oh, and they are slowly building a completely synthetic Frankenstein human being over several years to eventually bust out massive lines of credit.

Jeff That is wild. Catching a ghost like that is incredibly difficult because they look like perfect customers right up until they steal a million dollars.

Emily Precisely.

Jeff So how does a coordinated digital workforce tackle that?

Emily It starts with a strategy agent. This bot acts as a high level analyst. It continuously monitors production data across the bank, actively searching for blind spots in the current defense rules where these synthetic identities might be hiding. Okay, when it finds a new pattern, it proposes a new detection rule.

Jeff And then relying on that secure handoff contract, it passes its findings to the investigation agent.

Emily Yes.

Jeff And this is the part that totally blew my mind. Yeah. The investigation agent doesn't just read the alert.

Emily No it doesn't.

Jeff It takes the strategy context and it actually designs the downstream workflow for the human investigators.

Emily It's amazing.

Jeff It automatically sets up the alert queues. It defines the triage processes. It even customizes the user interface, building out risk scorecards and generating the specific step by step checklists. The human analysts will need to investigate this brand new type of fraud.

Emily Yes, it is literally structuring the human work based on its own analytical findings.

Jeff That's incredible.

Emily It is. And once the human completes the investigation, the workflow triggers the reporting agent.

Jeff Okay. The final step, right?

Emily This bot takes the approved verified findings and automatically generates business intelligence dashboards and regulatory summaries, allowing stakeholders to track how efficiently this new synthetic fraud ring is being dismantled.

Jeff You know, you just highlighted the single most important caveat in this entire deep dive. What's that? Throughout this incredibly sophisticated automated workflow, the human element never disappears.

Emily No, the human is the ultimate safeguard. Under this governance model, humans remain entirely legally and operationally responsible for consequential decisions. Right. The AI agents act as the ultimate support staff. They propose strategies, they design efficient workflows, they gather evidence, and they recommend dispositions. But a human reviewer must confirm, edit, reject or approve those recommendations before a customer's account is closed or a report is filed. The goal of Agentic AI is to amplify expert human judgment, not replace it.

Jeff Which is really the only sane way to run a bank.

Emily Absolutely.

Jeff I mean, you would never let a team of brilliant interns unilaterally change the company's routing numbers without the CFO signing off.

Emily No. Definitely not.

Jeff But if an organization is listening to this and they realize they are currently just sort of duct taping random AI tools together. Yeah. Um, where do they actually begin building this orchestrated control plane without getting totally overwhelmed?

Emily The overarching advice from the experts is pragmatic. You start small and you focus on friction. Okay. Do not add AI agents simply because a vendor offers them. You deploy an agent only when there is a defined measurable operational bottleneck.

Jeff And the blueprint specifically calls out investigation and reporting as the best places to start, right?

Emily Yes, because those areas are historically plagued by massive amounts of manual, tedious work, right?

Jeff All the copy pasting.

Emily Exactly. Human analysts spend hours copying and pasting data from five different screens just to write a single summary. A single AI agent can automate that data gathering and drafting process immediately. Oh, wow. So the evolution needs to be sequential. You establish a single agent to handle a manual task. Then you build a basic coordinated workflow between two agents. Okay? Once you prove that handoff works, you expand into full governed orchestration using the control plane architecture. And eventually that matures into continuous optimization.

Jeff It is all about earning trust at every single sequential step before you expand the footprint.

Emily Precisely.

Jeff Which really crystallizes the core takeaway for everyone listening today. If your executive board is asking, you know, how many AI agents have we deployed this quarter? They are asking the wrong question completely.

Emily It is a vanity metric. Yeah. The only question that dictates long term survival is can we trust the entire chain? Right. Can the detection bot securely pass immutable evidence to the investigation bot? Does the system preserve the context of a customer's history all the way through to the regulatory report? Mhm.

Jeff And do our human operators have the observability to audit, control and understand the consequential decisions happening at every single step?

Emily And if you can answer yes to those questions, then you have successfully built the control plane. And, you know, whether you are leading a fraud team, running a tech startup, or managing a supply chain, mastering orchestration, rather than just automating random, isolated tasks is how you future proof your organization against sprawl.

Jeff It separates the companies that just bought a bunch of expensive tools from the companies that actually built a secure, dependable digital operating model.

Emily Which leaves us with a truly fascinating and maybe slightly intimidating new reality to ponder.

Jeff Oh yeah.

Emily Yeah. Think about it. If a company successfully orchestrates this environment where a team of specialized bots handles the data gathering, the complex analysis, the workflow design, and the initial reporting, the human workers are no longer just doers. They are no longer the ones manually connecting the dots on a spreadsheet. They fundamentally transition into being managers of a digital workforce, and that requires a completely different cognitive skill set.

Jeff It really does. I mean, how does a legacy corporation rewrite its job descriptions, and how do you retrain a twenty year veteran employee when their primary colleague and effectively their most productive direct report is a machine?

Emily It's a huge shift.

Jeff Yeah, we kicked this off talking about fifty geniuses without a manager. Yeah. It seems like the future of enterprise isn't just about building smarter bots. It's about training an entirely new generation of human leaders who actually know how to manage them.

Emily Absolutely.

Jeff Definitely something to chew on as you look at your own workflows this week. Thanks for taking this deep dive with us.

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. The link is in the description.