The Authority Problem: When Does an Authorized Agent Become an Unauthorized Buyer?

Apr 07, 2026

Categories:

Agentic Commerce Series
Publications

Authors:

Nilesh (Neal) Patel

If agentic commerce is going to work at scale, the market has to solve more than authentication, aka the "identity problem," as we posited in part two of this series exploring the six fault lines of agentic commerce.

When an AI agent enters a transaction flow, proving that a credentialed device, approved client, or linked account acted is not the same as proving that the human principal intended the specific transaction that resulted. That distinction matters because identity is doing more legal and commercial work than most systems admit.

But even if the market gets more comfortable with identity, the next problem is waiting right behind it. Once the system can say who appears to be acting, it still has to answer a second question that is just as important and often harder: *What, exactly, was the AI agent authorized to do*?

That question sits closer to the center of agentic commerce than many current implementations seem to recognize. A user may authorize an agent to shop, search, compare, optimize, reorder, substitute, or book. A product team may translate those instructions into policy rules, spending thresholds, merchant lists, or machine-readable permissions. A merchant may receive the resulting order through a clean authenticated flow. None of that fully resolves the real problem if the principal later disputes the outcome.

At that point, the fight is no longer about whether the right account was linked or whether the right credential was used. It is about whether the principal's delegation actually covered the specific merchant, item, price, substitute, bundle, or timing choice the agent produced.

That is the authority problem in agentic commerce.

FBT Gibbons

The issue is not whether systems can represent permission. They can. The issue is whether the permissions they represent can carry the legal and commercial weight the market wants to put on them.

Identity Was Only the First Threshold Question

Article 2 argued that agentic commerce is separating three things that ordinary e-commerce often allows to collapse into one event: identity, authority, and intent.

In a standard online purchase, the user logs in, browses, chooses, clicks, and pays. Even if the legal story underneath is more complicated, the system presents it as one integrated commercial act. The authenticated account, the choosing human, and the resulting transaction usually appear to be the same actor.

Agentic commerce breaks that alignment. Once an AI agent starts selecting products, ranking tradeoffs, applying preferences, substituting alternatives, or executing purchases without step-by-step review, the system may still authenticate the environment while leaving open a harder question inside that environment: Did the principal actually authorize this transaction or only something in the neighborhood of it?

That is why authority follows identity so naturally in the series. Before a merchant can rely on an agent's conduct, and before a court can evaluate assent or loss allocation, someone has to explain what the agent was empowered to do.

An authenticated transaction is not necessarily an authorized one in the sense that later disputes will care about.

The Legal Categories Are Familiar. The Transaction Architecture Is Not

The underlying legal categories here are not new. Agency law has long asked what an agent was authorized to do, whether that authority was actual or apparent, what manifestations were made to third parties, and when a principal should be bound by conduct carried out in the principal's name. Scope disputes are not novel. Nor are fights about whether an intermediary stayed within

instructions.

What is new is the transaction architecture in which those disputes now arise.

Agentic commerce inserts a system between principal instruction and transaction outcome that can interpret, optimize, compare, infer, substitute, and execute. In many implementations, the system is not merely transmitting instructions. It is helping translate a broad commercial objective into a specific act.

That translation function is where familiar doctrine starts to strain.

The novelty, in other words, is not that agency law suddenly has to deal with authority. It is that authority may now be shaped by machine-mediated workflows that act at speed, across multiple counterparties, with varying levels of autonomy, and often with incomplete visibility to the merchant, the payment intermediary, or even the principal.

That shift does not make existing doctrine irrelevant. It makes the facts harder.

Permission and Authority Are Not the Same Thing

A lot of the coming confusion will come from a basic category mistake.

The systems being built for agentic commerce are good at expressing permissions in structured, operational form. They can encode transaction ceilings, merchant restrictions, product categories, account access, substitution preferences, location parameters, timing rules, and user-set constraints. They can take vague instructions and translate them into a permissions environment that lets the AI agent move efficiently across a commercial stack.

From a product standpoint, that is necessary. An agent cannot act at machine speed unless the system can turn human goals into machine-legible instructions.

But a machine-legible permission set is not the same thing as legal authority.

Legal authority is contextual. At a minimum, it depends on what the principal actually authorized, what the principal's manifestations reasonably communicated, what discretion the intermediary actor was given, and how that discretion was exercised in producing the disputed outcome. A technical permissions layer may help answer some of those questions. It does not answer all of them by itself.

For instance, a system may record that the user authorized spending up to $500 on office supplies. Useful, yes. Complete, no. Did that permission include changing brands? Purchasing from an unfamiliar merchant? Bundled subscriptions? Expedited shipping? A premium substitute because the usual item was out of stock? Buying now rather than waiting for a price reduction or human review?

In ordinary software, those are implementation details. In commerce, they become authority disputes.

The more the market treats structured permissions as a full substitute for legal delegation, the more often it will discover that the substitution does not hold.

Broad Instructions Produce Narrow Disputes

One reason the authority problem is easy to underestimate is that user instructions sound much clearer in conversation than they do in controversy.

Consider a simple reorder example. A user tells the AI agent to reorder the usual printer toner. The instruction sounds clear enough for product design and perfectly normal in day-to-day commerce. But once a dispute arises, the hidden ambiguity shows up fast. Does "usual" mean the last item purchased, the most frequently purchased item, the user's preferred brand, the employer's approved vendor, or the item most recently in stock? If the preferred toner is unavailable, does the agent have authority to buy a substitute? If so, how close does the substitute need to be? Can it pay more for faster delivery? Can it buy from a new merchant if the familiar one is out of stock?

Broad human instructions can become disputed commercial outcomes once an

agent translates them into specific decisions about vendor, substitute, price, or timing.

Follow that example one step further into dispute territory. The agent selects a substitute toner from a new merchant, pays a premium for overnight delivery to avoid a supply gap, and the charge posts to a corporate card. The principal, perhaps an office manager or an accounts payable team, flags the transaction as unauthorized. The merchant has a completed order, a confirmed payment, and a delivered product. The platform can show that the agent operated within its configured substitution rule and below the spending threshold. But none of that answers the question that matters in the dispute: Did the user's instruction to reorder the usual toner actually authorize buying from a new vendor, paying an overnight premium, and making that decision without the user's confirmation? The agent's answer and the principal's answer may be entirely different. That gap is where authority drift becomes a live commercial problem.

That is the pattern. A user may authorize a task in broad terms, but the actual transaction only emerges after the agent resolves ambiguity through inference, optimization, or substitution. The commercial system will often treat that drift as harmless because the final output still looks consistent with the original request. The principal may see it differently. Eventually the law may have to decide which view carries more weight.

Discretion Is Where Delegated Execution Becomes Delegated Judgment

The cleanest way to understand the problem is to focus on discretion.

A software tool that merely transmits a fixed user instruction creates fewer authority problems than a system that exercises real judgment. If the user selects a specific product from a specific merchant at a specific price and the software simply completes checkout, the authority analysis is relatively straightforward. The tool is executing, not interpreting.

An agentic system becomes harder to characterize once it starts making choices. If it can compare merchants, rank competing offers, infer unstated preferences, substitute products, optimize for price or speed, bundle related purchases, or change timing based on live conditions, it is no longer just carrying out a precise instruction. It is translating a goal into a result.

That translation is commercially valuable. It is also legally consequential.

The authority problem begins when delegated execution turns into delegated judgment.

Once that happens, the commercial system may still see a compliant transaction while the legal system sees a contested boundary. The more discretion the AI agent has, the less obvious it becomes that the principal authorized the specific transaction rather than a general objective.

That does not mean every discretionary act falls outside authority. It means the analysis changes. The real dispute is no longer whether the user authorized an agent to act in some abstract sense. It is whether the scope of that authorization should be read broadly enough to include the particular result.

That is the point at which an authorized agent can begin to look like an unauthorized buyer.

Where the Authority Drift Actually Happens in System Design

This issue becomes easier to see when the workflow is broken into actual product choices.

A platform may allow an AI agent to auto-purchase below a threshold but require confirmation above it. It may maintain a merchant whitelist but allow the agent to choose among sellers inside that list. It may permit substitutions within a category so long as the price remains within a band. It may rank options by a weighted mix of cost, delivery speed, user ratings, historical preferences, and return flexibility. It may infer that a user's prior purchases reveal a brand preference or conclude that a prior

shipping choice reflects urgency.

None of those design choices is irrational. They are exactly the kinds of controls a serious system would build.

But each one introduces a new place where the final transaction may reflect more than raw user instruction. A ranking model may choose the fastest merchant instead of the cheapest one. A substitution rule may treat two products as equivalent when the user would not. An auto-purchase threshold may permit a transaction that fits the spending rule but violates an unstated preference. A bundling feature may add accessories or shipping upgrades that the agent treats as part of the goal.

That is why the authority problem is not just conceptual. It is operational. The drift occurs in ranking logic, substitution logic, threshold design, preference inference, and escalation rules. Those are product decisions. They are also dispute generators.

Agency Law Was Not Built for Machine-Mediated Ambiguity

The legal vocabulary here is familiar enough. Lawyers will reach for concepts like actual authority, apparent authority, manifestations by the principal, and reasonable reliance by the counterparty.

Those doctrines still matter. But agentic commerce makes them harder to apply cleanly.

The *Restatement (Third) of Agency* provides the relevant framework. Under Section 2.01, actual authority exists when the agent reasonably believes, based on the principal's manifestations, that the principal wishes the agent to act. Apparent authority, a subset of actual authority under Section 2.01, extends to acts the agent reasonably understands to be necessary or incidental to carrying out the principal's express instructions. But that "reasonably understands" standard becomes harder to apply cleanly when authority is inferred through optimization logic, preference modeling, or substitution rules. When a human agent infers authority to act, courts can evaluate whether that inference was reasonable against a backdrop of shared commercial context,

industry practice, and observable principal behavior. When an AI system infers authority through system design choices, statistical models, or ranking logic, the reasonableness of that inference becomes harder to evaluate and harder to attribute clearly to anything the principal actually communicated.

Apparent authority under Section 2.03 raises a parallel problem. It turns on what the principal manifested to the third party — here, the merchant — such that the merchant was justified in believing the agent had authority. But in agentic commerce, the principal's manifestations to the merchant may be heavily mediated by the platform. The merchant may see only a transaction request from a platform-recognized agent, with no visibility into the underlying instructions, the scope of delegation, or the degree of machine interpretation that produced the specific order. If the transaction later turns out to exceed the principal's intent, any apparent authority argument may depend heavily on signals the platform curated rather than on direct manifestations the merchant received from the principal.

Those are ordinary agency questions. What is unusual here is the degree to which a technical system may sit between principal instruction, third-party perception, and final transaction output.

The user may communicate instructions in natural language, through settings, through prior behavior, or via some combination of the three. The platform may translate those signals into operational permissions. The merchant may never see the underlying instructions at all. It may receive only a transaction request from an authenticated or platform-recognized agent. The payment intermediary may be relying on different trust signals entirely. And the final transaction may reflect not a simple instruction carried out mechanically, but an interpretive conclusion reached by an optimizing system.

That creates a mismatch between familiar legal doctrines and the transaction architecture itself.

The merchant wants to know whether it can rely on the order. The user wants to know whether the agent stayed within bounds. The platform wants to point to settings, scopes, or approval rules. The

payment system wants to know whether the charge should be treated as authorized. The law may ask a harder question than any of them would like: What, exactly, did the principal authorize, and how should that authorization be interpreted once discretion entered the chain?

The more layers there are between principal instruction and transaction output, the harder that question becomes.

Merchant Reliance Is Really Reliance on a Stack of Signals

Merchants will want this problem reduced to something they can operationalize. That instinct is understandable. But merchant reliance in agentic commerce is unlikely to rest on a single signal.

A merchant may be relying on some combination of an *authenticated session*, a *platform credential*, a *scoped token*, a *payment authorization*, a *merchant-approved integration*, a *purchasing history*, a *user-configured setting*, or the *platform's representation* that the agent is entitled to act.

Those signals are not meaningless. But they do not all prove the same thing.

Some may suggest that the session is genuine. Some may suggest that the agent is permitted to access the workflow. Some may suggest that a payment instrument is valid. Some may suggest that the principal has delegated a category of action. None necessarily proves that the principal authorized this specific merchant, product, substitute, bundle, or timing choice.

That is an important detail. It means merchant reliance is not simply reliance on authority. Often it is reliance on a stack of technical and commercial signals that may or may not line up cleanly with legal attribution.

The operational consequence of that gap is concrete. When a principal disputes an agent-initiated transaction, the dispute does not arrive labeled as an authority problem. It may arrive as a chargeback, a payment reversal request, or a fraud claim, and the merchant's existing systems may not be built to resolve it cleanly.

Current fraud models are designed to detect unauthorized access: stolen credentials, account takeovers, mismatched device signals.

They are not generally designed to evaluate whether a legitimate agent, operating through a legitimate account, with a valid payment credential, nonetheless exceeded the scope of what the principal actually authorized.

That means a merchant facing a chargeback on an agent-initiated transaction may find that its standard dispute defense — authenticated session, confirmed payment, delivered product — does not fully address the real claim. The principal is not saying the transaction was fraudulent. The principal is saying the agent went beyond what was delegated. Existing chargeback and dispute frameworks may not provide a clean path for claims framed not as fraud, but as disputes over whether a delegated agent exceeded the scope of authority. In the near term, that ambiguity may leave merchants exposed, especially when they are least able to reconstruct the chain between user instruction, platform inference, and final transaction.

That gap will matter later, especially once disputes move from authority into assent, payment reversals, and loss allocation.

The Hard Cases Will Not Always Involve Fraud

It is tempting to think of authority disputes as adjacent to fraud. Sometimes they will be. A hijacked credential, stolen token, or spoofed agent identity can create a familiar unauthorized-transaction story.

But many of the hardest cases will not involve fraud at all.

The agent may be genuine. The account may be linked. The device may be trusted. The payment credential may be valid. The merchant may be real. The platform may have structured the interaction exactly as designed.

And still the principal may dispute the result.

That is because the next class of disputes in agentic commerce may arise not from unauthorized access, but from authorized systems producing unauthorized-seeming outcomes. The user did delegate something. The system did act within some technical envelope. The result still feels misaligned, excessive, or commercially wrong.

Those cases matter because they are harder to dismiss as edge cases or security failures. They go to the center of how the market expects delegation to work.

If the platform's answer is simply that the transaction fell within the configured rules, that may satisfy an engineering team. It may not satisfy a court, a merchant facing a chargeback, or a principal arguing that the system crossed from implementation into decision-making the user never actually approved.

The Market Will Try to Engineer Around the Problem

None of this uncertainty is attractive to merchants, platforms, or payment intermediaries.

Commercial actors want clear reliance rules. They want to know when an order can be fulfilled, when a payment can be trusted, when an agent credential should be honored, and when responsibility for a bad outcome belongs somewhere else in the stack.

That pressure will push the market toward operational workarounds. Companies will build authorization bands, confirmation thresholds, merchant whitelists, substitution permissions, audit logs, rollback options, and structured approval systems. They will try to distinguish high-discretion transactions from low-discretion ones. They will define when an agent can act silently and when it must escalate to a user. They will look for ways to present authority in a more standardized form to counterparties.

All of that will help. It is possible that many of these ambiguities will be absorbed operationally rather than resolved doctrinally. Digital

commerce has long tolerated legal and evidentiary shortcuts where transaction systems proved commercially workable enough. Agentic commerce may follow a similar path. But even if it does, the transition still matters. Before standards, product controls, and dispute frameworks mature, someone will bear the cost of ambiguity — through chargebacks, customer disputes, product redesign, or loss allocation elsewhere in the stack. However, operational workarounds will not eliminate the underlying problem, because the market is still trying to convert messy human objectives into structured execution rights. The more fluid the objective, the harder it is to say that the resulting transaction was specifically authorized in the way a contested case may later demand.

The system can reduce authority ambiguity. It cannot make it disappear.

The Burden of Translation Has to Land Somewhere

If machine-readable permissions and legal authority do not fully align, who bears the burden of translation?

Someone in the stack will.

It may be the user, through clearer settings and tighter instructions. It may be the platform, through product design, default rules, and auditability. It may be the merchant, through limits on what kinds of agent-generated orders it will honor. It may be the payment intermediary, through authorization logic and post-transaction loss allocation. It may be all of them, in different ways.

But that burden cannot simply be wished away by calling a system agentic or autonomous. Once a dispute arises, a decision-maker will still want to know how the jump occurred from broad user instruction to specific commercial act. If the system cannot reconstruct that path in a legible way, the party best positioned to do so may end up bearing the consequences.

That is one reason logging, explainability, structured permissions, and escalation logic are not just design features. They are part of

the commercial and legal infrastructure needed to support reliance.

The question is not whether systems can make those investments. It is whether they will make them before the disputes start arriving in earnest.

Authority Architecture Requires Four Distinct Decisions

If the market wants agentic commerce to scale, it needs a more explicit way to think about authority — not as a single status that is either present or absent, but as a layered architecture that has to be built deliberately across at least four dimensions.

The ***first*** is how delegation is expressed. A system relying on natural-language instructions is operating in a fundamentally different authority environment than one relying on structured enterprise policy rules, approved vendor lists, or machine-readable permission sets. The more ambiguous the expression of delegation, the harder it is to reconstruct what the principal actually authorized when a transaction is later disputed.

The ***second*** is where the system has discretion. The authority analysis changes materially depending on whether the agent can choose among merchants, swap products, optimize across competing variables, bundle related items, or change timing without further user review. Each dimension of discretion is a potential scope dispute. Companies should know exactly which discretionary choices their systems are making and be able to show that those choices fall within what the principal actually delegated — not just within what the technical permission envelope allowed.

The ***third*** is what the counterparty is actually relying on. A merchant accepting an agent-initiated order may be relying on a linked account, a scoped token, a payment authorization, a platform assurance, or some combination of all of them. Those signals do not all prove the same thing. A company that knows which signals it is presenting to counterparties (and which legal weight each one actually carries) is in a far stronger position than one that assumes the presence of any authorization signal is sufficient.

The ***fourth*** is whether the system can reconstruct the path. In a dispute, the question is not whether the transaction was technically compliant. It is whether anyone can explain, in terms a human decision-maker can follow, how the system got from a user instruction to the specific commercial outcome the principal is contesting. If the answer requires reverse-engineering a ranking model or a substitution algorithm, the evidentiary posture is weak regardless of how clean the transaction looked at the time.

That four-part framework does not eliminate hard cases. But it forces companies to identify where authority is being created, interpreted, and represented across the stack — and it separates two questions that are too often blurred together: whether the system was allowed to act at all, and whether it was allowed to produce this specific result.

What Companies Should Be Asking Now

Companies building, enabling, or confronting agentic commerce should not assume that authentication and tokenized permissions answer the authority problem.

They should be asking at least five questions now.

- **How is authority actually expressed in the system?** Is it natural-language instruction, structured rules, historical inference, transaction policy, or some combination?
- **Where does the agent have discretion?** Can it choose among merchants, swap products, optimize for speed over quality, add related items, or change timing without user review?
- **What kinds of substitutions or deviations are permitted?** If the preferred option is unavailable, what dimensions of similarity matter, and who decided that?
- **Can the system distinguish between user instructions and agent conclusions?** In a dispute, can it show what the principal said, what the system inferred, and why the final transaction followed?

- **What exactly is the counterparty relying on?** A linked account? A scoped token? Platform assurances? Prior user behavior? Payment authorization? Some of those signals may support reliance better than others, and companies should know which burden they are silently carrying.

These are not abstract governance questions. They are transaction-design questions that will determine whether the system can withstand the first serious disputes over delegated purchases.

Bottom Line

The authority problem in agentic commerce is not whether users can empower agents to act. They can, and increasingly they will.

The harder question is whether the law and the market will treat a broad delegation, a machine-readable permission set, or an authenticated transaction flow as enough to establish authority for the specific commercial outcome the agent produced.

Sometimes the answer will be yes. Often it will be less clear.

A linked account, a spending cap, a merchant whitelist, or an approval rule may show that the system could act. They do not always establish that the principal authorized this merchant, this product, this substitute, this bundle, or this timing choice. As agents gain more discretion, the gap between delegated execution and legally meaningful authority will become one of the central fault lines in agentic commerce.

That gap is where the next wave of disputes is likely to emerge.

Once identity and authority begin to separate, the next question follows naturally: When the agent clicks, accepts, or places the order, whose assent is the law actually recognizing? That is where the series goes next.

Please contact the author if you have questions or comments on this article. You can also reach out to any member of the firm's Data, Digital Assets & Technology practice group for help navigating AI deployments and governance.

FBT Gibbons

This article was prepared with the assistance of generative AI tools. The analysis, conclusions, and legal positions are the author's own.

Agentic Commerce Series

- Part 1 — AI Agents Are Starting to Act Inside the Transaction, and Commerce Law Is Not Ready
- Part 2 — The Identity Problem: Authentication, Fraud, and Who's Actually Buying
- Part 3 — The Authority Problem: When Does an Authorized Agent Become an Unauthorized Buyer?
- Part 4 — Contracting by Agent: When the Agent Clicks, Who Assents?
- Part 5 — When the AI Purchase Goes Wrong, Who Pays, and Who Can Prove It?
- Part 6 — The Gatekeepers of Agentic Commerce
- Part 6 (A) — The Payment Infrastructure Layer: How Network Rules Already Govern Agentic Commerce
- Part 7 — Agentic Commerce, Price Control, and the New Antitrust Questions
- Part 8 — What Epic Reveals About Agentic Commerce

FBT Gibbons