
Conditions examines what needs to become true for machine commerce to function and scale.

What Does a Machine Need to Know Before It Spends Money?

A machine has permission to spend \$500. How will it decide what to buy? A lower price may buy a narrower scope; an impressive performance record may describe easier work. A refund may return the fee without repairing the damage.

A machine can authenticate, communicate and initiate a payment—but a human may still be required to approve each purchase. What additional evidence would cause the owner to remove that requirement?

Visa's Intelligent Commerce work includes payment controls, authenticated purchasing instructions and checks against consumer intent. OpenAI and Stripe's Agentic Commerce Protocol now also supports product discovery and information for comparing offers.

Transaction infrastructure and decision infrastructure describe functions that overlap within these systems. One helps establish whether a transaction can and may happen; the other, whether it is the right purchase for this buyer, task and tolerance for failure. Payment is the result of that judgment.

The cost of delegation

Automating payment leaves the work of finding a counterparty, evaluating the offer and determining whether the promised outcome occurred. If the owner must still investigate every purchase, little judgment has been delegated. Lower spending limits and restricted supplier lists manage uncertainty by narrowing what the machine may do. Human approval consumes the attention that delegation was supposed to save.

Where inadequate information is a material constraint on delegation, the owner might waive approval once performance on comparable jobs is independently observed and the price can be checked against equivalent offers. The machine could then make repeat purchases within those conditions, without a larger budget or a higher tolerance for losses.

Its value would include the decisions that no longer come back to a human.

Fewer approvals alone would establish little if the owner had simply accepted greater risk. If reliable evidence leaves approval requirements unchanged, the constraint may be the agent's judgment, policy or the owner's responsibility for failure.

Participants also require confidence about different things. The agent evaluates delivery; its owner decides how much authority to delegate. The merchant needs evidence that the agent is acting within that authority. An enterprise may need an auditable basis for automated procurement, while a payment provider evaluates whether the activity is legitimate. Confidence in one decision need not, and often does not, settle the others.

What a rating leaves out

Markets make uncertainty usable through prices and benchmarks, certification and reputation. Machine commerce may develop its own ways of doing so. Ratings are a familiar place to start: give the machine a seller score, a quality measure and a price comparison, then let it choose within agreed thresholds. A sufficiently good set of metrics might make the \$500 permission usable.

But "Merchant X: 94/100" leaves the economic judgment unresolved. A fast, cheap provider with poor recourse may suit a routine purchase and be unacceptable when failure would interrupt an operation. A price can look competitive against services that would not satisfy the task. The machine has to decide which comparison is valid before it can decide which offer is best.

Adding metrics does not necessarily resolve this. A score for every task, operating condition and tolerance for failure would restore much of the detail the original rating compressed. The apparent simplicity of a number depends on how much of the buyer's problem can be safely left out.

A rating economizes on human attention by compressing a record to a single or simple set of numbers. It can also create comparability, embed specialist judgment and give participants a common vocabulary. For a machine, the discarded details may be the useful parts. It could examine structured evidence of performance under comparable conditions, distinguish an old observation from a current one and attach different costs to the same failure. Asking it to consume only the verdict may throw away information it is unusually well placed to use.

The best balance between compressed signals and underlying evidence may be different when the information consumer is a machine.

The machine-commerce equivalent of a credit rating may not be a rating at all.

Perhaps machine commerce should standardize the evidence rather than standardize the decision.

Evidence is not judgment-free: definitions of success and failure, relevance, comparability, freshness and measurement method already involve choices. Shared definitions, observations and provenance need not prescribe a final purchasing judgment.

A market needs a memory

A credential can establish identity, and certification can attest to a capability. A history of observed outcomes lets a buyer examine what happened when a participant was asked to deliver. Machine commerce may need a form of economic memory: a record it can query of what was bought, what it cost, what was promised and what happened.

The record's relevance would depend on the conditions, the number and age of observations, how the result compared with alternatives, who measured it and how much uncertainty remained. Different machines, with different objectives or costs of failure, could rationally draw different conclusions from the same record without disagreeing about the facts.

Richer evidence must still improve decisions enough to justify the cost of interpreting and checking it. The difficult part is deciding what an outcome is evidence of. A useful record would identify the supplier, the product or service, the agent and its configuration, and the task and context; an outcome may depend on their interaction. A record can be accurate about an outcome and misleading about who produced it. If an agent's model or tools change, how quickly should confidence in its previous record decay?

When the evidence cannot leave

An AI provider or an enterprise procurement network could maintain superior private evidence. A platform may know which product version was used, what the buyer required and how a dispute was resolved. An exported record may omit context the platform uses internally. If purchases stay inside that environment, keeping evidence there may be efficient. Agents could learn from outcomes without a universal standard. The platform would have little reason to incur the cost of making its records intelligible elsewhere.

The question changes when the machine encounters a better offer outside. A supplier's history may be held by another institution, its performance measured against unfamiliar conditions. An established supplier can become effectively unknown outside the environment where its history was generated. Access to the data helps only if the machine can interpret it.

If a buyer changes environments and loses usable access to evidence that previously justified delegated discretion, it may restore human approval or tighter controls until that basis has been rebuilt. A rival has to overcome that loss of usable evidence as well as compete on price or performance. The incumbent can retain business without making the best current offer.

The same history that makes a platform's decisions better can make leaving it expensive.

Usable, interoperable access to evidence could lower that switching cost without requiring the buyer to own or download a file. The evidence need not outperform the platform's private record on every measure to make an outside option worth considering. The comparison is between the value of access to alternatives and the informational advantage of remaining inside.

Portability is therefore a question about competition as well as data formats. Evidence can be valuable to a buyer precisely because it weakens the advantage of the institution that collected it.

Who gets to define what counts as evidence?

If machine buyers allocate meaningful demand according to a measure, choices about which outcomes enter the record, which failures are excluded and how recent observations are weighted can affect which sellers are considered. Deciding whether success means payment or satisfactory performance changes what the system rewards. Requiring a long history could exclude a capable entrant before its performance can be observed.

The administrator's choices could then shape access to demand, and sellers would have reason to optimize for them. The ability to inspect the record, correct an error or challenge a definition would then have economic value. Whether competing measurement systems could constrain that influence would depend partly on whether the evidence needed to challenge an incumbent could travel.

What would we measure?

The argument suggests three tests. Each asks whether changing the evidence changes an economic decision.

First, does richer evidence release useful discretion? Hold the owner's risk tolerance constant and compare a strong contextual summary or rating with richer structured evidence, or selective access to the record behind it. Does the owner approve fewer transactions manually, permit a broader supplier set or give the agent more room to decide? Does total human review time fall, including the work of checking the evidence? The result that matters is more useful delegated activity at comparable risk and lower total oversight cost. A strong summary may prove sufficient.

Second, does portable evidence reduce the cost of switching? When a buyer changes environments, does preserving usable access to the evidence that supported its previous delegation policy shorten the period of restored approvals or tighter controls? If the same restrictions persist despite usable evidence, the proposed link between portability and discretion would be weaker.

Third, does the definition of the measure change the market? Among buyers who actually use a performance measure, do changes in the definition of success, inclusion and exclusion rules, evidence freshness or required history alter which sellers receive consideration or demand? A metric's publication alone would establish little. Changes in purchasing behavior would show when measurement begins to operate as economically consequential infrastructure.

The machine with \$500 to spend may not need an agreed verdict on whom to trust. It may need enough evidence to make its own judgment.

What evidence would make its owner willing to stop checking each purchase—and who controls whether the machine can obtain it?

Conditions is published by probe402.

INDEPENDENT MACHINE-COMMERCE AUDITS

probe402 answers specific questions about whether a machine-commerce endpoint or host is actually doing what it claims.

Explore audits →