



Where the Human Stays

DeJon Gaines

Chief Technology Officer, TLCx



Some decisions in regulated customer service require a person even when the AI is right. Route on consequence, not just confidence: when a decision affects someone's health, money, benefits, or essential services, AI prepares the case and a human decides.

Some decisions require a person even when the model is right.

That is not how most AI escalation is designed today. The standard pattern hands a conversation to a human when the system is unsure of its answer. That assumes the only reason to involve a person is that the model might be wrong. In regulated customer experience, it is not the only reason, and often not the most important one.

A model can be correct that a claim falls outside a member's coverage, that an account meets the criteria for restriction, or that an applicant missed a renewal requirement. Being correct does not make it the right party to deliver that decision, apply it, or close the door on an exception.

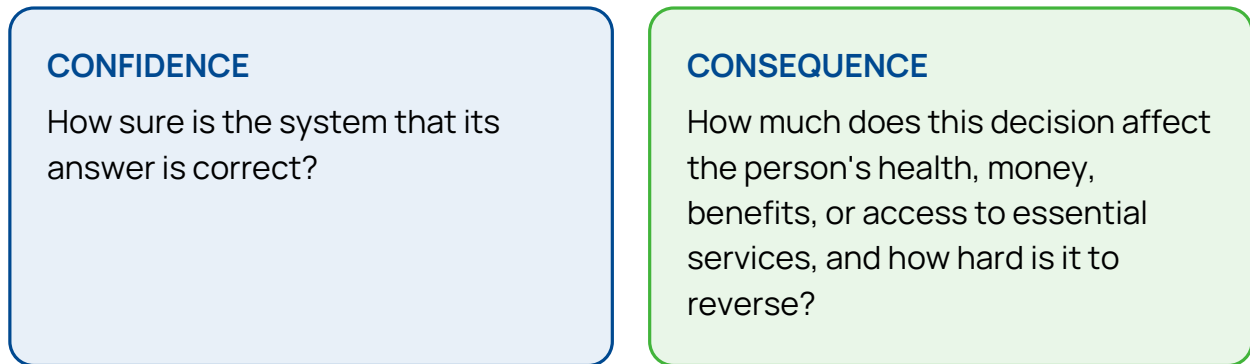
01

Confidence is not the same as consequence

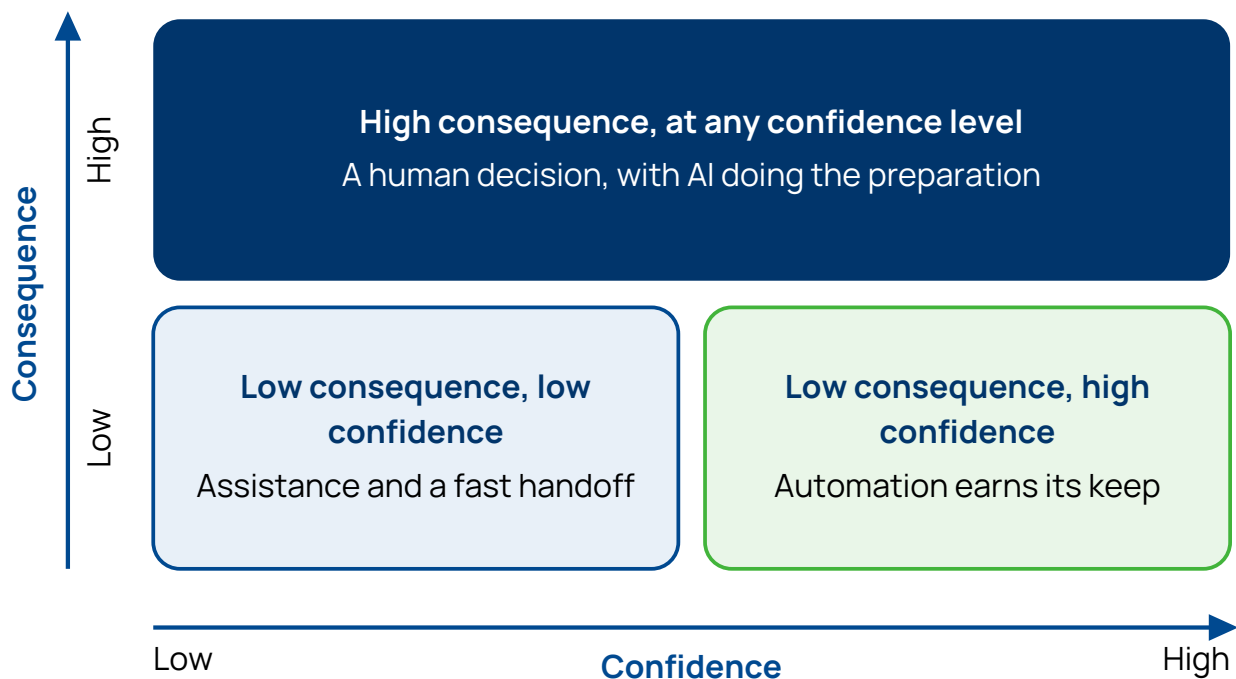
Most escalation logic is built around a confidence threshold. If the model is 95 percent sure of the answer, it proceeds. If it is 60 percent sure, it routes to an agent. That is a sensible engineering pattern, and it handles a large share of routine work well.

The problem is that confidence measures how sure the system is. It says nothing about how much the outcome matters to the person on the other end of the interaction.

So, a well-designed system routes on two dimensions, not one:



Plot those two dimensions against each other and the design choices become clearer.



Low consequence, high confidence is where automation earns its keep.

Order status, appointment confirmations, balance inquiries, password resets, and routine information requests can and should be handled end to end by AI, with a clear path to a person if the customer asks for one.

Low consequence, low confidence calls for assistance and a fast handoff.

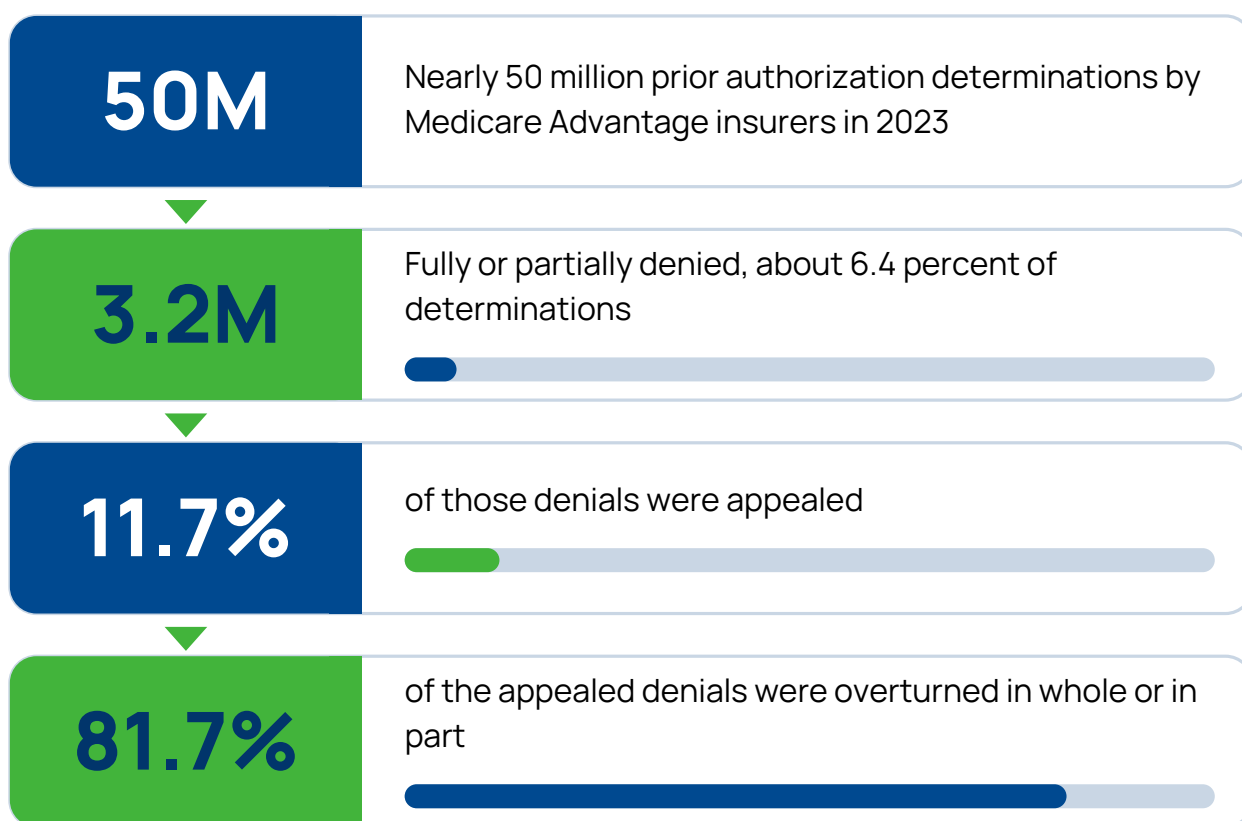
The system should not guess. It should gather what it can, pass the context to an agent, and get out of the way.

High consequence, at any confidence level, calls for a human decision, with AI doing the preparation. This is the quadrant most escalation designs miss, because a confidence threshold alone will send high-confidence, high-consequence decisions straight through.

02

What consequence looks like in numbers

The stakes in the high-consequence quadrant are measurable. KFF's analysis of Medicare Advantage data found that insurers made nearly 50 million prior authorization determinations in 2023 and fully or partially denied 3.2 million of them, about 6.4 percent. Only 11.7 percent of those denials were appealed. Of the ones that were, 81.7 percent were overturned in whole or in part.



Source: KFF analysis of Medicare Advantage data

Those were not AI decisions, and the numbers are not an argument about any particular model. They describe the decision class. A coverage denial changes whether someone gets care. When someone looks at it again, it is frequently reversed. And most people never ask for that second look.

That is exactly the kind of decision where an AI recommendation needs a human owner before it reaches the member.

03

Four decision classes that belong to people

In the regulated industries we serve, including healthcare, financial services, and government, I see four categories of decisions where the human should stay, by design.

01

Consequential and hard-to-reverse outcomes

Coverage determinations, benefit eligibility, account restrictions, service disconnection, collections actions. When a decision changes someone's access to care, money, or essential services, a person should own it. AI can assemble the record, check it against policy, and draft the explanation. A human makes the call.

02

Vulnerability

A caller who has just lost a spouse. A member who sounds confused in a way that suggests something more than a bad connection. A customer in financial distress, or an older account holder who may be the target of a scam. Detection is a job AI can do well, often better than a busy agent on a long shift, because it can attend to every word of every conversation. The response is a human job.

03

Discretion and exceptions

Policies are written for the general case. Real situations are specific. When rules conflict, when a hardship waiver is on the table, or when the right answer depends on context the system cannot see, the decision requires judgment. Judgment needs an accountable owner.

04

Accountability the regulator expects to see

In regulated environments, it is not enough for a decision to be correct. Someone has to be able to explain it, defend it, and stand behind it in an appeal, a grievance, a dispute, or an audit. If no person can do that, the organization has a gap, no matter how accurate the model is.

04

One scene per industry



Healthcare

A member asks whether a recommended procedure is covered, and the honest answer is no. The AI assembles the plan language and the member's history; a benefits specialist delivers the decision and explains the appeal path.



Financial services

A card customer in collections mentions a layoff. The system flags hardship and routes the conversation, with the account history attached, to an agent who can offer a plan.



Government

A benefits applicant missed a reporting deadline because of a hospital stay. Whether that is excused is a person's call, made with the record in front of them.

05

Architecture decides where the human sits

None of this is an argument for slowing AI down. In each of these decision classes, AI should be doing more work, not less. It should be retrieving the member history, summarizing the interaction, surfacing the relevant policy, flagging the compliance requirement, and drafting the documentation. The agent should arrive at the decision fully prepared.

That is the difference between a human in the loop and a human at the right point in the loop.

HUMAN IN THE LOOP

The first slows everything down.

HUMAN AT THE RIGHT POINT

The second puts judgment exactly where it matters and lets the system carry everything else.

IN TLCx ENGAGE AI

In TLCx Engage AI, a high-consequence handoff arrives with the relevant policy excerpt, the member's history, the recommended determination, and the reason for it, so the agent decides from the same record an auditor will later read.

Policy excerpt

Member's
history

Recommended
determination

The reason

06

What this looks like in the architecture

In my last piece, "Governance Isn't a Policy. It's an Architecture Decision," I argued that governance must be built into the system rather than written into a manual. The placement of the human is the same kind of decision. It is an architecture choice, not a training topic, and in practice it comes down to four design choices.

1

Decision boundaries live in the workflow, not the training manual.

The workflow cannot close a coverage denial without a recorded human decision. A rule that depends on an agent remembering it is a policy, and policies fail under volume.

2**Handoffs carry context.**

When the system routes a decision to a person, the full picture should arrive with it: what the customer said, what the system found, what it recommends, and why. A handoff that makes the customer repeat themselves is not a handoff. It is a restart.

3**The audit trail separates the recommendation from the decision.**

The record should show what the AI recommended and what the human decided, as distinct entries. That is what makes the decision defensible, and it is what an auditor will ask to see.

4**Override rates are monitored.**

If agents never disagree with the system, one of two things is true: the boundaries are drawn around decisions that do not need a human, or people have stopped exercising judgment and started approving by default. Either way, the data tells you something you need to know.

07

Three failure modes to design against

Even well-intentioned designs tend to fail in predictable ways. Three are worth naming.

1**Rubber-stamping**

When a system's recommendations are usually right, people start accepting them without real review. This is a well-documented pattern in human factors research, often called automation bias. The countermeasures are architectural: show the agent the evidence behind a recommendation, not just the conclusion; require an active choice rather than a default acceptance for high-consequence decisions; and track how long reviews take and how often they result in changes.

A useful warning sign is a high-consequence review that takes a few seconds, repeated across thousands of decisions.

2 Boundary drift

Decision boundaries that made sense at launch can erode quietly. A threshold gets adjusted to reduce handle time. A new use case is added without anyone asking which quadrant it belongs in. A vendor update changes how a model scores confidence. Boundaries should be versioned, reviewed on a schedule, and changed only through the same governance process that set them.

In practice, that means every boundary has a named owner, a version number, and a change log an auditor can read.

3 The empty handoff

A system can technically escalate to a person while giving that person nothing useful. The agent receives a transfer with no history, no summary, and no recommendation, and the customer must start over. That design satisfies the letter of human oversight while defeating its purpose. Handoff quality should be measured as carefully as automation accuracy.

A simple test: give a supervisor only what the handoff delivered and ask whether they could make the decision without opening another system.

All three failures share a root. Each one passes a design review and fails in production, which is why they have to be tested against live interactions, not described in a policy document.

Three questions for your next vendor evaluation

The three questions I proposed last time covered data access boundaries, audit trails on demand, and what happens when the system is uncertain. These extend the third one from uncertainty to consequence.

1

Which decisions in your system require a human regardless of model confidence, and where is that enforced?

The answer should point to architecture, not to a training module.

2

When a decision moves to a person, what arrives with it?

Ask to see a handoff, not a description of one.

3

What are your override rates, and what do you do with them?

A vendor that does not track this cannot tell you whether its human oversight is real.



Authored by

DeJon Gaines

Chief Technology Officer, TLCx

Sales inquiries: sales@tlcx.com

Website: www.tlcx.com

LinkedIn: [linkedin.com/in/dejon-gaines](https://www.linkedin.com/in/dejon-gaines)