

Redefining Agent Efficiency:

Unlocking Efficiency, Agility, and Growth

Research and Published by Cloudset (www.cloudset.net)



Table of Contents

Executive Summary	2
Introduction	3
Where Support Breaks Down: The Hidden Friction in Service Setups	5
The Role of AI: Amplifier or Band-Aid?	8
Why Agent Efficiency Needs a Reframe	12
What's Actually Holding Back Progress	15
Conclusion: Consistency Drives Efficiency for Effective Human-Centered Support	19

Executive Summary

As support teams grow, things start to break. Not all at once, and not in obvious ways. Response times stretch. Handoffs get murky. People aren't sure where to route complex issues, and onboarding starts to feel like folklore instead of process. Everyone's working hard, but somehow that hard work isn't adding up.

This paper looks at why. We talked to CX leaders and Zendesk specialists to understand what's really slowing teams down. What we found wasn't about motivation or individual performance. It was structural. Systems that don't scale. Processes that make sense in isolation but clash in practice. Tools that technically work but don't quite work together.

To explain the pattern, we introduce the "Figure of Eight" model - a way of seeing how efficiency and effectiveness feed into each other. When that loop is working, performance compounds. When it breaks, things stall, and leaders start chasing isolated fixes instead of addressing the system itself.

The key idea is simple: efficiency isn't something you bolt on later. It has to be designed from the start. Not as a one-time optimization, but as a way of thinking about how work flows, how decisions get made, and how people move through the system.

Get that right, and support teams can scale without losing speed or quality. Get it wrong, and no amount of dashboards or coaching sessions will fix the drag.

Introduction

Talk to any support leader, and the same signals show up: widening performance gaps, rising escalations, agent burnout, and customers repeating themselves.

The reflexive fix: Optimize the agent. Tighter scripts, more macros, a dose of AI, shorter handle times.

But what if the problem isn't the agent?

Agent efficiency is too misunderstood - flattened into a numbers game of speed and cost. That narrow view pushes it down the priority list, overshadowed by customer-facing metrics like satisfaction or resolution rate.

Yet when viewed through the right lens, efficiency isn't the enemy of quality - it's the foundation for consistency, clarity, and trust - at scale.

The Figure of Eight Model: Rethinking the Relationship

To articulate this dynamic, we introduce the "Figure of Eight" model, a looped framework that illustrates how efficiency (supply-side) and effectiveness (demand-side) are interdependent.

- **Efficiency (Supply-Side):**

Structured workflows, faster resolution times, reduced onboarding friction, and better access to knowledge to empower agents to work with greater clarity and consistency.

- **Effectiveness (Demand-Side):**

As agents spend less time on repetitive tasks and more time engaging meaningfully, outcomes like CSAT, First Contact Resolution (FCR), and customer loyalty naturally improve.

In many organizations, the loop breaks - because system-level inefficiencies get buried under day-to-day firefighting. The cost isn't just slower operations; it's lost potential.

By shifting focus from agent output to system design, support leaders can unlock scalable efficiency, restore flow, and elevate the experience for both agents and customers.

Interdependent Efficiency and Effectiveness Cycle

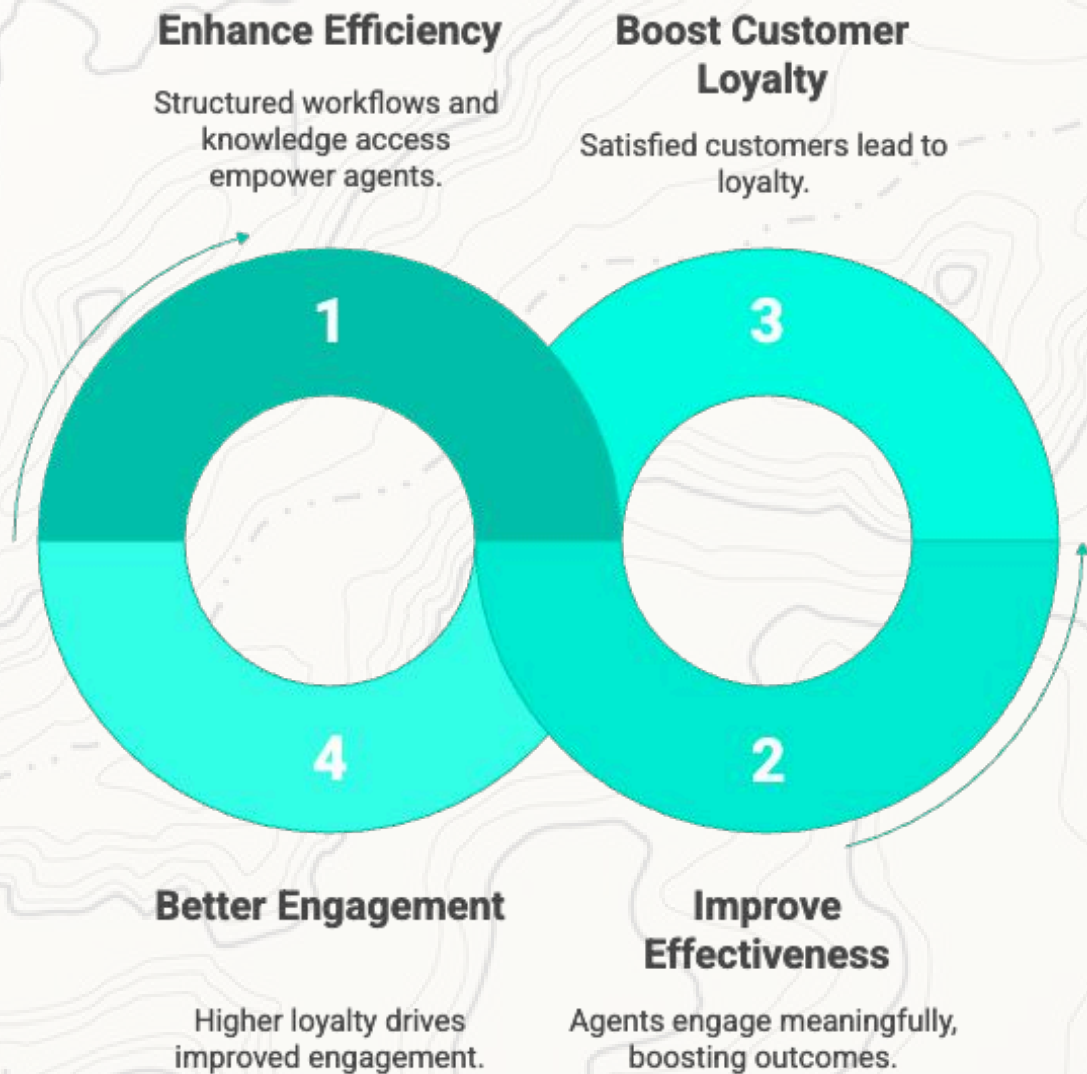
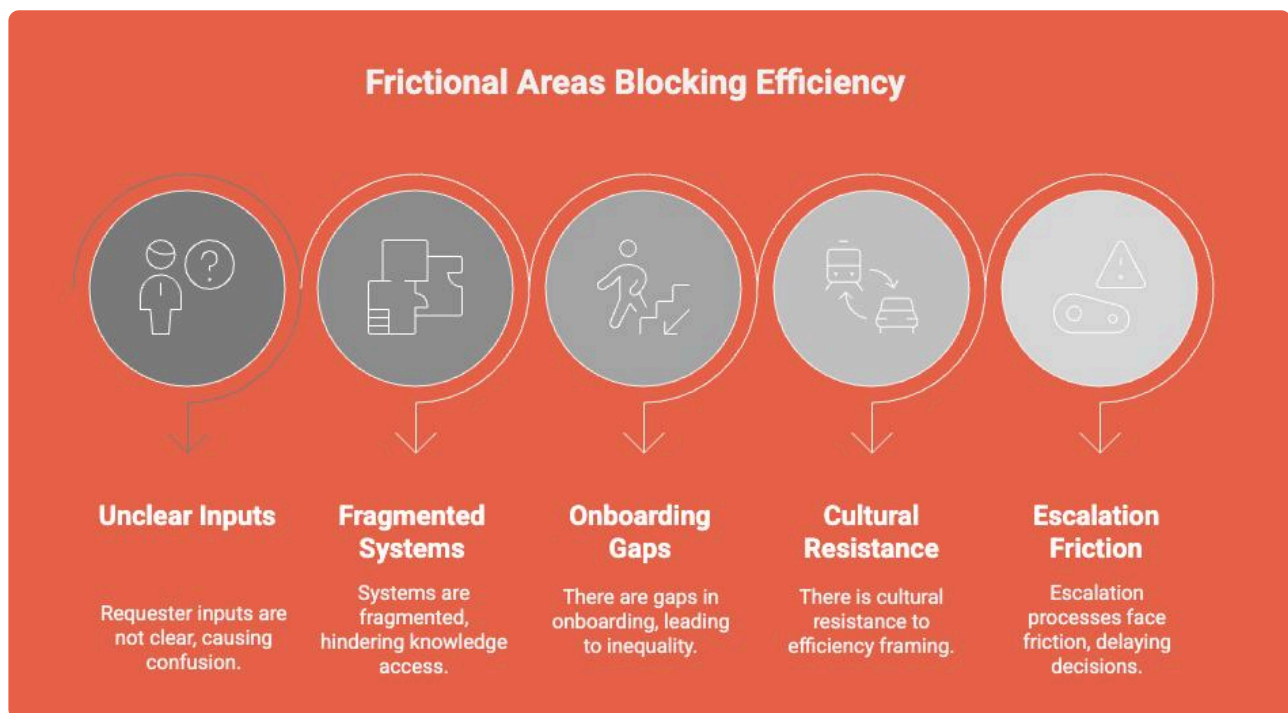


Figure 1: The Figure of Eight Model

Where Support Breaks Down: The Hidden Friction in Service Setups

Most service leaders have a sense of where their team is struggling – slow responses, inconsistent resolutions, rising escalations – but the root cause lies beneath the surface. In our interviews with service leaders, five core friction points emerged again and again. These aren't individual failings, they're symptoms of gaps in process design, tooling, and systems thinking.

These are the hidden forces that hold back efficiency, erode agent confidence, and block great customer experiences.



Unclear Requester Inputs

- **Problem:** Many customer and employee requests arrive incomplete or ambiguous, forcing agents to spend valuable time clarifying the issue before resolution can begin.
- **Insight:** Leaders described frequent “back-and-forth” interactions, particularly in high-volume environments, where vague queries increase handle time and customer frustration.
- **Design Opportunity:** Structured intake, proactive education, and upstream classification can dramatically reduce this noise, and free agents to solve, not decipher.

Fragmented Systems & Knowledge Access

- **Problem:** Agents are forced to toggle between multiple systems, tools, and internal chats to resolve even simple tickets. This complexity increases onboarding time, decision fatigue, and inconsistency.
- **Insight:** Across interviews, leaders described the burden of “tribal knowledge,” with new agents leaning heavily on peer support due to poor documentation or scattered knowledge bases.
- **Design Opportunity:** Move from static help centers to embedded knowledge, surfaced contextually and in-flow nudges. Your best agents shouldn't also be your knowledge base. Bake guidance into the flow so the system supports everyone equally.

Onboarding Gaps & Experience Inequality

- **Problem:** New agents face a long ramp-up period (2–6 months), struggling to internalize systems and develop the judgment needed for complex cases. In some orgs, top performers become bottlenecks, shouldering complex tickets and supporting others.
- **Insight:** Several leaders described a “confidence cliff” at the three-month mark, where agents stagnate without structured coaching or exposure to real-world variation.
- **Design Opportunity:** Shift from onboarding as an event to enablement as a continuous system:
 - Use scenario-based simulations early, and revisit them as agents advance.
 - Trigger coaching prompts based on ticket types or QA flags.
 - Create tailored playbooks for specific workflows and update them as the system evolves.

Confidence scales when agents are supported in real time, not just day one.

Cultural Resistance to Efficiency

- **Problem:** Efficiency is still perceived by some teams and leaders as synonymous with pressure or performance micromanagement.
- **Insight:** While most interviewees acknowledged efficiency as important, they also revealed hesitations - worried it might conflict with customer-centricity or agent wellbeing if pursued without context.
- **Design Opportunity:** For efficiency to become strategic, it must be reframed not as a demand for more output, but as a design principle that enables better outcomes for customers and agents alike.
Reframe efficiency as friction reduction - less about output, more about removing blockers to great work.

Escalation Friction & Decision Latency

- **Problem:** In complex environments, unclear escalation paths delay resolution and force agents to make repeated context switches. This not only inflates handle time but erodes confidence and trust in internal processes.
- **Insight:** In highly regulated or multi-product organizations, agents were frequently described as “navigating ambiguity” with limited procedural clarity or case routing support.
- **Design Opportunity:** Escalation protocols must be reengineered to reduce friction and provide agents with confidence and autonomy.
Clarity and automation here can significantly reduce rework, repeat contacts, and supervisor load. Build routing logic around how work gets resolved, not org charts.

Efficiency is a System Design Outcome

Across all five constraints, a single insight becomes clear: efficiency doesn't fail at the agent level, it fails at the system level.

To improve agent performance, leaders must shift from focusing on individual productivity to redesigning the conditions around the agent. Clearer customer inputs, unified knowledge access, effective onboarding, cultural reframing, and smarter escalation flows form the foundation of scalable, sustainable efficiency.

The Role of AI: Amplifier or Band-Aid?

Service leaders widely see AI as a key lever for efficiency - but its effectiveness hinges on thoughtful deployment. When integrated with context and care, AI can streamline repetitive work and surface the right information at the right time. Without trust, oversight, and situational awareness, though, it becomes a brittle layer - failing precisely when reliability matters most.

Leader Perspectives: Augment, Don't Fully Automate

Support leaders broadly favor AI-assisted agents over end-to-end automation, emphasizing a balance between efficiency and human connection.

Participants confirmed two dominant pillars of AI usage:

- **Contact Deflection:** AI is used for ticket triage and after-hours support, particularly for high-volume, low-complexity queries.
- **Agent Support:** AI copilots assist agents during live interactions by surfacing content, suggesting replies, or guiding resolution steps.

One participant shared that they currently use AI for after-hours support and ticket deflection, and are exploring Zendesk's copilot-style tools to reduce handling time. Another participant noted AI's potential to assist with **70–80%** of common cases, but emphasized maintaining a human touch for emotionally nuanced or complex queries.

Pillars of AI Usage

Agent Support
AI assists agents with information and resolution suggestions.



Contact Deflection

AI handles simple queries, especially during off-hours.

“

“See AI’s potential, but prefer AI-assisted human agents over full automation at this stage.”

- Intan Vermeer, Head of Customer Service, Digitec Galaxus

At the same time, not all leaders are convinced yet:

“We’re not that impressed with agent assistance yet due to our complexity. Chatbots have their place because they’re convenient for customers.”

- Sébastien Faure, Head of Customer Service & Support, Devialet

Where AI Supports Efficiency

When applied intentionally, AI can drive measurable gains across the support operation:

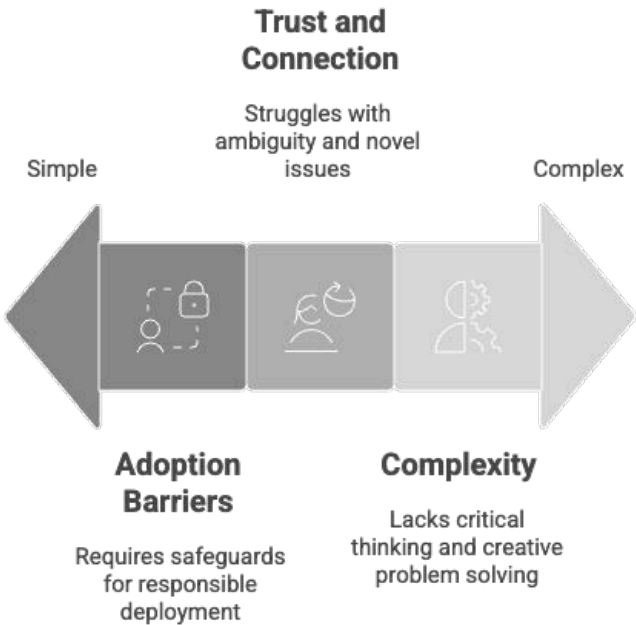
- **Knowledge Access:** Copilot-style tools ingest help center content and recommend answers in real time, reducing reliance on informal backchannels - useful for general text based guidance.
- **Ticket Deflection:** AI chatbots handle routine queries during off-hours or peak loads, minimizing backlog and freeing agents for higher-value work.
- **Resolution Speed:** AI suggests macros and content in context, reducing Average Handle Time (AHT) and helping teams meet benchmarks (e.g., ~6 minutes) without increasing pressure.

Research from ServiceNow (2024) [10] reports **20% productivity gains** by using AI agents to automate internal service requests, demonstrating how assistive AI scales without compromising quality.

Where AI Falls Short

...
Today’s AI tools come with real limitations, and many leaders expressed a measured optimism shaped by those constraints.

Short Falls with AI on the Complexity Spectrum



- **Complexity:** AI struggles with cases that require critical thinking, multi-system troubleshooting, or creative problem solving. One participant noted new products lack an adequate knowledge base for AI to pull from, reducing its reliability.
- **Trust and Connection:** Over-automation risks alienating customers, especially when it replaces human empathy with scripted logic. It also struggles with ambiguity, edge cases, and novel issues that require judgment, critical thinking, or empathy.
- **Adoption Barriers:** In regulated industries like healthcare, concerns about data privacy, explainability, and compliance can stall implementation. Teams require safeguards like audit trails, fallback options, and transparency to deploy AI responsibly.

The Bottom Line on AI in Support

Interviews and external research point to a consistent insight: AI is most effective as an amplifier, not a replacement. It speeds up routine work and boosts consistency, but trust remains a barrier. To succeed, AI must be implemented with care - supporting agents, not sidelining them - especially in complex or sensitive interactions where human judgment is essential.

Why Agent Efficiency Needs a Reframe

Agent efficiency is misunderstood - too easily conflated with burnout or corner-cutting, and seen as at odds with customer-centricity. But that mindset is changing. Forward-thinking support leaders are redefining efficiency as reducing friction, not increasing pressure.

Instead of speed for speed's sake, efficiency is being redefined as thoughtful enablement: the ability to deliver great service with less friction.

“Efficiency means delivering a ‘wow’ experience to customers while being as efficient as possible.”

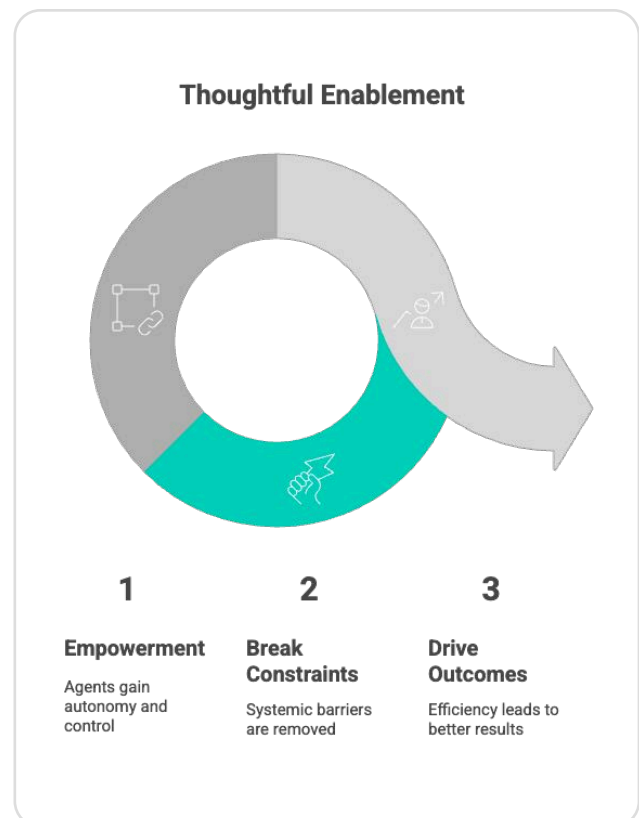
- Raymond Culp, Sr. Director of Customer Care, Heartland

From Pressure to Empowerment

Leaders across interviews described a shared shift in mindset:

- From rigid time-based KPIs to holistic service outcomes
- From working harder to working smarter - with better tools and workflows
- From a cost lever to a design principle for better customer and agent experiences

Efficiency now includes system design, cultural mindset, and access to the right knowledge, what one leader called "the golden zone of not under- or over-servicing customers."



“

“Efficiency, to us, means agents understand the width and depth of the industry and can use that knowledge to solve problems fast.”

- Director of Support Operations

Systematic Constraints break the Loop

Organizational challenges - unclear inputs, scattered systems, sluggish escalations - do more than slow things down. They break the feedback loop where efficiency and effectiveness should reinforce each other. When agents are stuck navigating broken processes, both speed and service quality suffer.

Efficiency Drives Outcomes

When designed well, efficiency doesn't compete with customer satisfaction - it strengthens it. Research shows it can cut support costs by up to 33% while boosting loyalty and retention.

By automating routine tasks, agents are freed to focus on what matters most: resolving complex issues, earning trust, and delivering standout experiences. Efficiency, when reframed as an enabler, becomes a catalyst for better outcomes across the board - for customers, agents, and the business.

What's Actually Holding Back Progress

Everyone already knows the basics: unify your knowledge base, improve onboarding, streamline workflows, train agents better, use AI to free up agents for more complex resolutions.

But if those fixes were enough on their own, this paper wouldn't exist.

The challenge isn't knowing what to do, it's deciding what you're willing to change.

Through conversations with CX leaders, five structural tensions emerged - persistent blockers that stall progress. These aren't process flaws. They're deeper questions of ownership, accountability, and system design.

1. Are you Leveraging the Top 10% for the Good of the Middle 80%?

Too many support strategies chase what top agents do best. That's a dead end. You don't scale by idolizing outliers - you scale by eliminating variation. Consistency beats brilliance when you're building a system that works every day, for everyone.

The real unlock is designing for consistency.

- Reduce cognitive overhead so average agents can resolve complex tickets without heroics.
- Let top performers mentor the system, not just junior agents - encode their judgment into workflows and tools.

"I particularly resonate with the idea of leveraging the top 10% for the good of the middle 80%. I cheated and only hired top 10% staff, but at scale that simply isn't an option. It's important to recognize that you can neither hold everyone to the standard of the very best, nor can you settle for mediocrity. What feels like a catch-22 is actually resolved by letting top performers mentor the system."

- Cory LaPlante, LaPlante Management

2. Have You Replaced “Soft Skills” with “Hard Systems”?

Empathy, tone, and judgment are still central to good service, but they can’t be your only defense against bad processes.

If agents need emotional intelligence to navigate broken workflows, they’ll burn out.

- Build systems that reduce emotional labor, not just tinker with an AI assistant in place corrections.
- Use AI to handle the easy repetitive, so agents have bandwidth for the human moments that matter.

The goal isn’t to replace soft skills, it’s to stop relying on them to patch bad design.

3. Are You Still Delegating Process Ownership to Service Teams?

Most service inefficiencies start upstream - baked into policies, SLAs, or tools shaped by product, legal, or ops. But it’s the service teams who end up patching the fallout. They’re left solving problems they didn’t create, with systems they didn’t design.

This isn’t just inefficient. It’s unsustainable. It’s not fair.

- **Make efficiency a cross-functional goal**, not just a CX KPI.
- **Shift accountability from people to systems** - design flows that prevent rework, not just handle it.

Systemic efficiency only happens when other teams stop treating service like the last mile, and start seeing it as a core part of delivery.

4. Do Your Escalation Paths Reflect How Work Really Happens?

Most escalation paths lack clear criteria - resulting in escalations that are too early, too late, or inconsistently applied. Combined with poor routing and weak communication, this creates delay, confusion, and unnecessary rework.

The friction isn’t in the agent. It’s in the flow.

- **Define clearly escalation criteria** - where does confidence break down?
- **Redesign escalations around clarity** - are you leveraging ticket data?

“

**“AI is helpful when it’s assistive,
but it won’t replace humans.”**

- Interviewee, Director of Support Operations

5. Are You Using AI to Cut Costs at the Peril of Trust?

AI can scale support - but only when built on a solid knowledge foundation. Without it, AI delivers shallow or wrong answers at speed, creating more problems than it solves.

- **Use AI to give agents leverage, not limitations** - codified knowledge needs to be accurate, otherwise it fuels hallucinations, complicates guardrails, and undermines trust.
- **Pair every automation with accountability** - fallback paths, explainability, and human handoff.

This Isn't About Optimization, It's About Ownership

None of this is easy. It won't come from tips or surface-level tweaks. Real change means rethinking how support is designed, how success is measured, and who owns the friction.

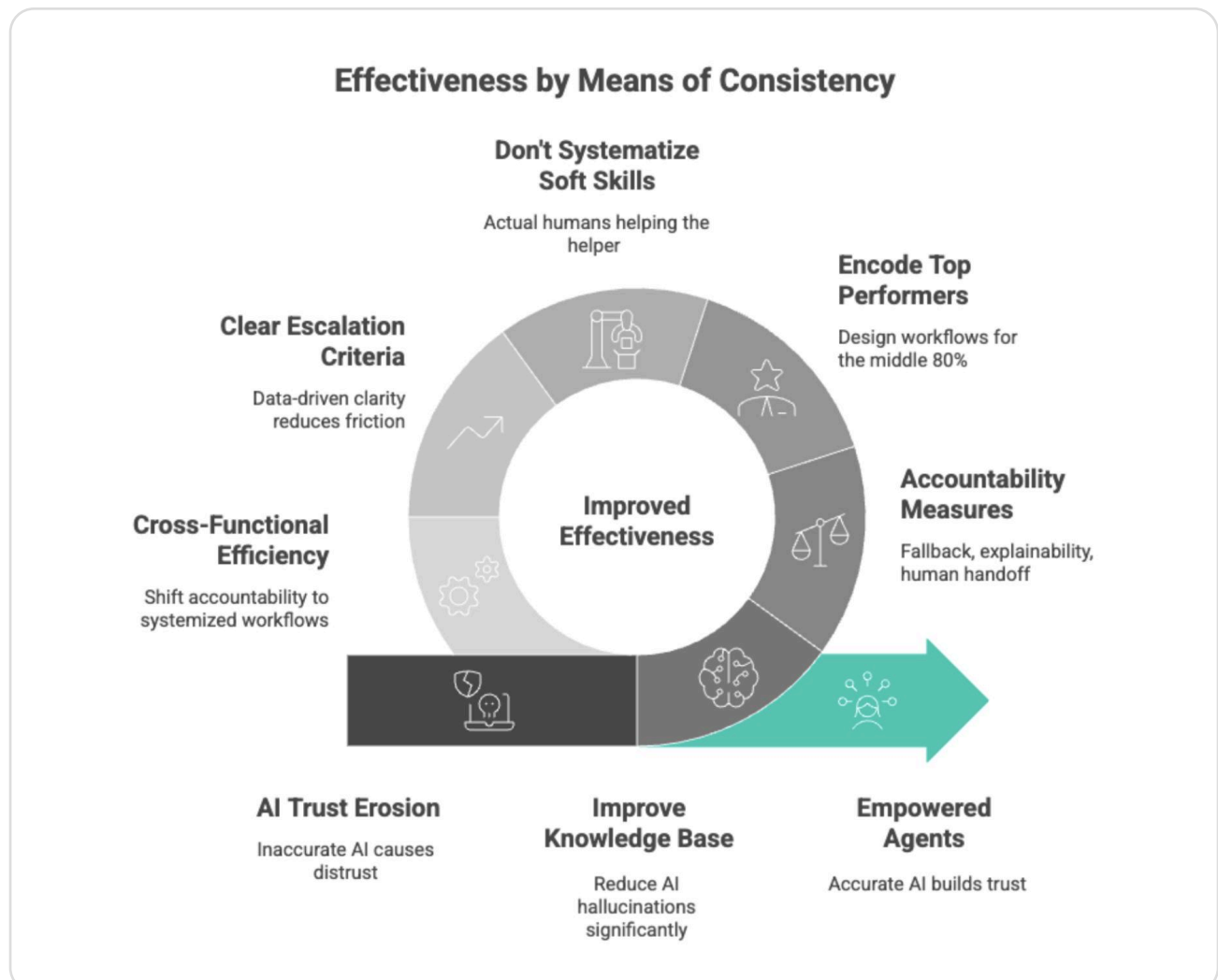
But leaders who take that on don't just get faster resolutions or lower costs - they build a team that scales with confidence and a customer experience that actually earns trust.

Conclusion: Consistency Drives Efficiency for Effective Human-Centered Support

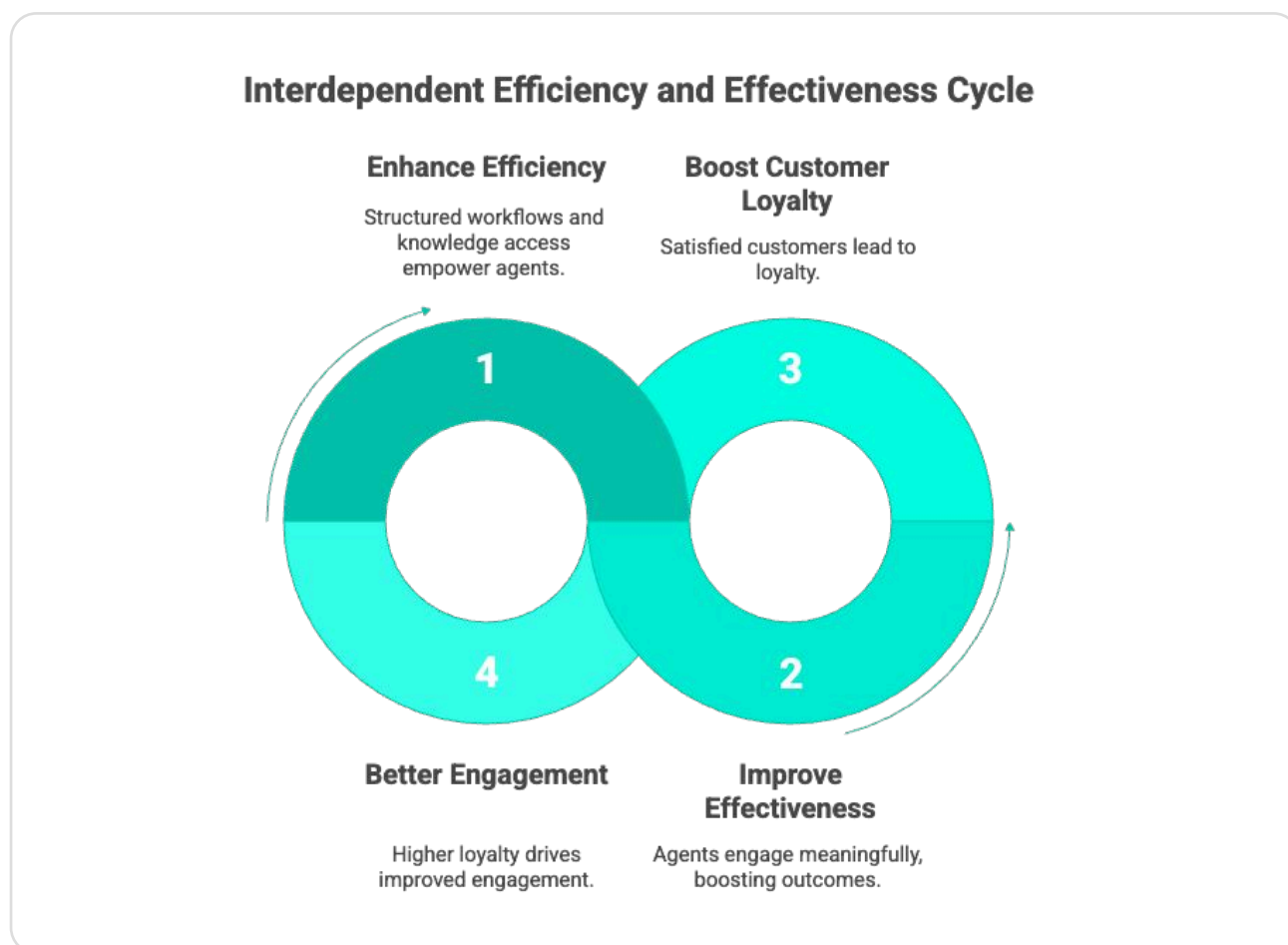
Agent efficiency isn't a trade-off with customer experience - it's what makes it possible at scale.

When efficiency is defined as the deliberate removal of friction - not just speed for its own sake - agents gain the bandwidth to do what only humans can: listen, think, and resolve with care.

This paper makes one thing clear: efficiency isn't about pushing agents harder - it's about building smarter systems around them. The real performance gaps don't come from lack of effort, but from fragmented tools, broken processes, and outdated assumptions. Fix those, and agents thrive. And when agents thrive, customers notice.



The Figure of Eight model captures a core truth at the heart of scalable support: efficiency drives effectiveness, and effectiveness reinforces efficiency in return. It's not a one-way street - it's a system loop. Break it, and performance stalls. Close it, and support becomes a growth engine.



Despite widespread agreement, most teams still hesitate. Efficiency is acknowledged as important - but treated as optional.

So the question isn't what to do. That's already clear.

The real question is: What's stopping you from closing the loop?



Research and published by Cloudset (www.cloudset.net)

Interviews conducted by Graham Robson

Research by Saumya Seth

Premise & Editing by Graham Robson & Joan Westenberg

Thank you to all the interviewees and contributors for their inputs and perspectives.