capacity®

WHITEPAPER

The future of contact center AI is self-learning

AI that stops learning the day it ships quietly becomes a liability. The teams pulling ahead use AI that keeps improving.

www.capacity.com



Executive summary

Most contact center AI stops learning post-implementation. The teams pulling ahead use AI that learns from every interaction, so the operation improves instead of falling behind. Static AI rarely fails loudly. It quietly becomes less reliable, until you find out from a customer.

You bought an AI platform. The pitch was that it would resolve more on its own, keep customer experiences consistent, and take pressure off your team. What nobody mentioned in the demo is that the system was at its smartest on the day it launched. Left alone, it gets a little less accurate every week after that.

The world your AI was trained on keeps moving. Your policies, pricing, promotions, and products all change, and a static model doesn't change with them. Whether it works today matters less than whether it will still work nine months from now (and whether you'd even know if it wasn't).

That's the deeper problem: contact centers forget. In three ways, at once. AI doesn't automatically keep up with what changes around it. QA can't see enough of what happens to flag every problem. And when experienced people leave, they take institutional knowledge with them.

A learning loop gives the operation a memory. It reviews every interaction, surfaces what's changing, feeds improvements back into the system, and preserves what the operation learns. The pages ahead unpack each part of that problem, show what the loop looks like in practice today, and explain why human oversight is a feature worth paying for.



1 Static AI forgets what changed

A machine learning model is a snapshot. It captures the patterns in its training data at a particular moment, then holds that picture still. The world it describes doesn't.

Model Drift

The gradual decline in a model's accuracy as the live world diverges from the data it was trained on. It's a normal property of production systems. In a changing production environment, decay is the risk to manage, and it keeps happening unless someone actively works to prevent it.

As reality moves away from what the model learned, its answers become less reliable. This happens often enough to have a name.

Deploying AI is where the real work starts. The industry treats launch like the finish line, when launch is really just the moment the model is as accurate as it will ever be without help. The most direct evidence comes from a peer-reviewed study in Nature's Scientific Reports, by researchers at Harvard Medical School, MIT, and the Whitehead Institute, that tracked model accuracy over time across healthcare, finance, transportation, and weather. They found that **91% of models degraded as they aged**, even ones that were highly accurate the day they were trained.[1] IBM describes the same pattern in production systems: a model's accuracy can start slipping within days of going live, as real conditions drift from the data it learned on.[2]

None of this is surprising. It's what happens when you put a fixed model into a moving environment. The question for a contact center is: what, exactly, is moving?

2 Why it goes unnoticed in a contact center

Few environments are harder on a static model than a contact center. The environment moves fast, and quality review covers almost none of it.



The ground shifts weekly

Policies change. Prices change. Promotions launch and expire. Products get added, renamed, and retired. Every one of those changes is something the AI can now be wrong about, and a static system has no way of knowing that the correct answer it gave last month is wrong this month. It keeps serving the answer it learned.

QA doesn't review enough

In a traditional quality program, human reviewers read or listen to a sample of interactions and score them. The open secret of the industry is how small that sample is. A figure commonly cited across the QA field is that teams manage to review 1 to 2% of interactions, leaving the other 98% as an accepted blind spot.[3] It's a capacity problem. Manual review doesn't scale, so a sliver is all anyone can physically cover.

Put drift and sampling together and you get the trap. A stale answer doesn't announce itself. The AI delivers it to every customer at once, in the same confident tone it uses when it's right. And because 98% of interactions are never reviewed, the error keeps going quietly. It shows up later as churn, as a complaint, or as a compliance miss caught by someone outside the building. The answer may have been wrong hundreds of times before it ever landed on a report.

> **"We had a client whose previous voice agent kept quoting a promotion that had ended 3 weeks earlier. Nobody noticed until customers started calling in to complain. The answer was wrong hundreds of times before it ever showed up on a report."**
>
> **SAMMIE STEPHENS,** SVP OF CUSTOMER SUCCESS

3 Employee turnover makes it worse

Model drift is one way a contact center forgets. Employee turnover is the other. The AI forgets because the world changes around it. The organization forgets because its people leave, and they take what they knew with them.

Contact centers have always had a knowledge-retention problem, and attrition makes it worse. They lose agents faster than almost any other line of work. Every departure takes the workarounds, the exceptions, and the hard-won judgment out the door, and the person who replaces that agent starts over.



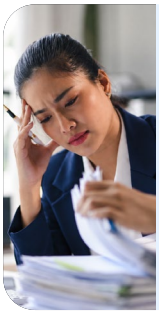
Attrition is brutal

Deloitte's 2024 Global Contact Center Survey put average agent attrition at **52% a year**, and much of it hits early, before a new hire is fully productive. McKinsey estimates it costs $10,000 to $20,000 to replace a single agent, counting recruiting, training, and lost productivity during ramp-up.[4]



Institutional knowledge leaves about as fast as new hires arrive

The tenured agent who knew the workaround, the exception, the right way to phrase the awkward thing: that knowledge was never written down, and now it's gone.



The people who replace them don't know where anything lives, and finding it is a real tax on the day

The people who replace them don't know where anything lives, and finding it is a real tax on the day. McKinsey's analysis of interaction work found that the average worker spends **nearly 20% of the workweek** just looking for internal information or tracking down a colleague who can help.[5]

Here's the point that changes how you look at the whole problem. In an operation with this much churn, the people can't be the durable memory. They leave too often. The system has to become part of the durable memory of the operation. A static system can't do that: it doesn't capture what departing agents figured out, and it doesn't pass that knowledge to the new ones.

A learning loop gives that knowledge somewhere to stay. When the system captures the resolutions that actually worked and feeds them back in, the next new hire inherits the answer instead of rediscovering it. **An operation that forgets stays flat. One that keeps what it learns gets better every quarter.**

4 What “self-learning” actually means

“Self-learning” means different things to different vendors. Here’s how we define it at Capacity, and what it does and doesn’t do today.

It’s a visible, governed cycle with 4 moves:

- **Measure** every interaction, all the way to 100%.
- **Surface** what’s failing, what’s drifting, and what customers are newly asking for.
- **Feed** those findings back into the knowledge base and the guidance your agents get.
- **Repeat**, so the operation gets a little more right each cycle.

Here’s the important distinction: a learning loop worth buying keeps a human inside it. Insights get reviewed before they change what the AI tells a customer. The system doesn’t quietly rewrite itself overnight and hope for the best.

That human step is what makes AI safe to scale. Today’s models are good and getting better fast, purpose-built tools already outperform general ones, and each model generation improves on the last. But none of them hit zero. Even the leading purpose-built AI tools, grounded in trusted content, still get a meaningful share of answers wrong in independent testing.[6] That’s not a reason to hold back on automation. It’s the reason to automate inside a loop.

The math is simple: at the contact-center scale, one wrong answer isn’t one wrong answer. It’s the same answer repeated across thousands of interactions before anyone notices. A governed loop is what catches those cases, corrects them, and feeds the fix back, so the system gets more reliable the more it runs, instead of quietly drifting. That’s the difference between AI you can trust in production and AI you’re hoping works.

This is where a subject-matter expert earns their place: deciding what the system learns and what it doesn’t, keeping answers consistent, and protecting the trust that’s slow to build and quick to lose. Oversight frameworks agree. The U.S. government’s NIST AI Risk Management Framework calls for defined human oversight, and for human intervention where an AI system can’t catch its own errors.[7]

The stakes here are real. MIT’s 2025 report The GenAI Divide: State of AI in Business found that **about 95% of enterprise generative-AI pilots delivered no measurable return**, and its authors tie the failures to tools that don’t retain feedback or adapt to the specific business they were dropped into.[8] A loop is what keeps a pilot from stalling out at that point.

> **“The pattern we see at Capacity surprises people. The teams automating the most are the ones who built the tightest feedback loop around their AI. Oversight earns you the confidence to automate more.”**
>
> **JULIA DALTON** | SVP OF PRODUCT, CAPACITY

5 What powers the loop

Capacity is building toward a platform where every interaction, human and virtual, feeds back into the system, so agents and answers continuously improve based on real conversational insights. Here's what that looks like in practice, and where a person fits in.

Capacity's learning loop runs on AI Agents and our Post-Interaction Suite. It works in three steps: measure everything, surface what matters, and feed the learning back.

Measure everything

Automated QA scores 100% of interactions, human and virtual, across voice, chat, and tickets, against the scorecard you define. That's full coverage, where a traditional program typically reviews 1 to 2%. AI CSAT predicts customer satisfaction without depending on survey responses. You stop guessing what the other 98% looked like.

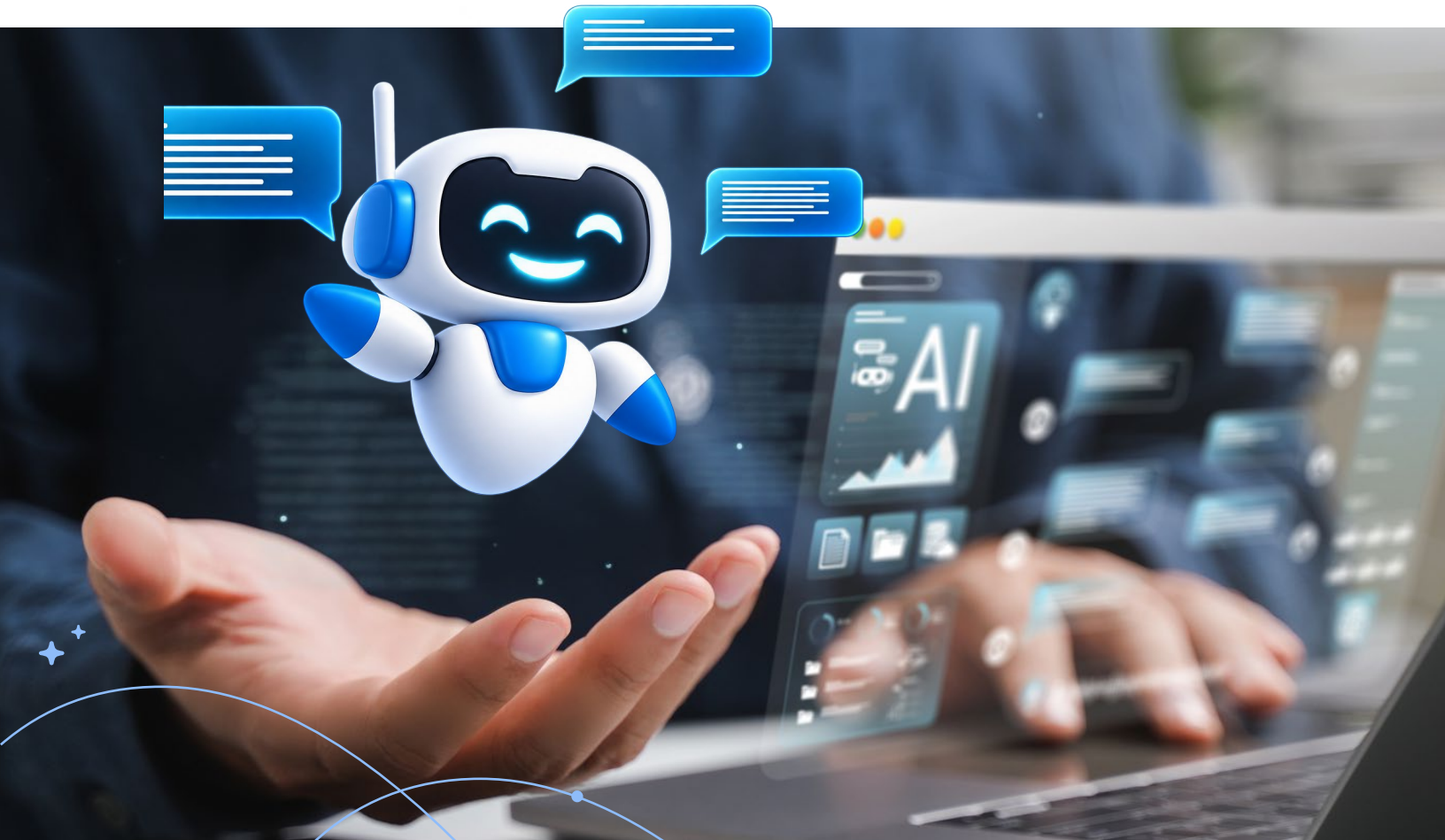
Surface what matters

From that measurement, the platform identifies the patterns worth acting on: recurring issues, compliance risks, training needs, and new automation opportunities. You get a ranked shortlist of what to fix instead of a dashboard you have to interrogate.

Feed it back

Those insights feed back into the platform. Human agents improve through real-time guidance informed by performance data, and virtual agents consistently improve with each interaction, improving deflection and accuracy.

Today a person stays on the decisions that matter, reviewing the important changes before they reach a customer, and Capacity is building toward a loop that optimizes both virtual and human agents automatically, creating ongoing improvement with less manual effort.





THRASIO

100% Coverage

$260K Saved/year

97% CSAT score

Thrasio, a consumer-goods company resolving 70,000+ support tickets a month, could manually audit only about 3% of interactions. By their own estimate, **reviewing 100% by hand would have taken 528 analysts.**

With Capacity scoring every interaction instead, they:

- Reached 100% coverage
- Saved about $260K a year in QA efficiency
- Gave agents data-backed coaching instead of subjective spot-checks

On the customer side, predictive CSAT scoring flags tickets likely to disappoint, and a human reviews those before they go out, a step their VP of Global Services calls “checking your homework.”

That extra layer of review helped raise their CSAT from 89% to 97%.

As the platform matures, more of its products will operate inside the learning loop, and more of the optimization will happen automatically. As that automation proves itself, the need for hands-on SME oversight lightens over time. That’s where contact center AI is heading, and where we’re investing.

6 What you get when AI compounds

Picture 2 operations side by side. One runs AI that decays a little each week and gets corrected, if at all, by whatever the 1 to 2% sample happens to catch.

The other runs a loop. It measures everything, surfaces what's slipping, and feeds the fix back in. In month 1 the gap is invisible. By month 9 it's the whole story. Here's what the loop buys the second operation, and why each one follows from the loop itself.



Better answers, fewer repeats

Because the loop scores 100% of interactions, drift gets caught and corrected before it turns into churn or a compliance finding, instead of hiding in the 98% nobody reviewed.



New agents start further ahead

Because the loop keeps the resolution that worked, a new hire inherits it instead of grinding through months of ramp to rediscover it. A large study of 5,000+ support agents by economists at Stanford and MIT, published in the Quarterly Journal of Economics, found that AI guidance built from top performers' behavior raised productivity 14% on average and 34% for the newest, least-experienced agents, effectively moving them down the experience curve faster.[9] That matters when replacing an agent runs $10,000 to $20,000.



The economics can move in your favor

McKinsey estimates that resolving a customer-support issue costs $7 to $20 with a human versus a fraction of a cent when AI handles it. As the loop safely improves how much the AI can resolve on its own, more volume shifts to the cheaper path, which changes the cost base of the whole operation.[10]



It improves key KPIs

The improvements can show up in the KPIs you already run the operation on: shorter average handle time, higher first-contact resolution, and less hold time spent hunting for an answer.

An operation that forgets pays for the same mistakes twice. An operation with a memory compounds what it learns.

That's the difference between deploying AI and operating it. The teams pulling ahead aren't asking whether their system was smart when it launched. They're running systems that are smarter this month than last, because every interaction gives the operation something to learn.

That's what self-learning AI means here: AI that learns continuously from the operation, improves deliberately, and stays accountable to a person for what it changes. The loop is how a contact center stops forgetting.



Final words

Keep learning, or fall behind.

Every week your AI runs without a learning loop, it drifts further from reality, quietly serving stale answers until a customer catches what your QA never saw. Capacity closes the loop, so your AI gets smarter every month instead of quietly falling behind.

Contact our experts today to secure your place at the forefront of the industry.

CONTACT US

sales@capacity.com

www.capacity.com

capacity®

Sources and Notes

Selected references supporting the claims in this paper.

1. Vela, Daniel, et al. “Temporal Quality Degradation in AI Models.” Scientific Reports, vol. 12, no. 11654, 2022. Nature, https://www.nature.com/articles/s41598-022-15245-z.

2. IBM. “What Is Model Drift?” IBM Think, https://www.ibm.com/think/topics/model-drift.

3. The 1–2% manual-review figure is widely cited across the contact center QA field. See, e.g., RingCentral. “AI quality management: Why your contact center can’t rely only on manual QA anymore” RingCentral, https://www.ringcentral.com/us/en/blog/ai-quality-management-why-your-contact-center-cant-rely-only-on-manual-qa-anymore/.

4. Deloitte Digital. Balancing Priorities in a New Era of Contact Center Transformation (2024 Global Contact Center Survey). Deloitte Development LLC, May 2024, https://www.deloittedigital.com/us/en/capabilities/contact-center-agent-experience.html. (Average agent attrition of 52% in 2023.) On replacement cost, see McKinsey & Company. “Boosting Contact-Center Performance through Employee Engagement.” McKinsey & Company, https://www.mckinsey.com/capabilities/operations/our-insights/boosting-contact-center-performance-through-employee-engagement. ($10,000–$20,000 to replace one agent.)

5. McKinsey Global Institute. The Social Economy: Unlocking Value and Productivity through Social Technologies. McKinsey & Company, July 2012, https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy.

6. Magesh, Varun, et al. “Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools.” Journal of Empirical Legal Studies, 2025. Stanford RegLab, https://reglab.stanford.edu/publications/hallucination-free-assessing-the-reliability-of-leading-ai-legal-research-tools/.

7. Tabassi, Elham. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1, National Institute of Standards and Technology, U.S. Department of Commerce, Jan. 2023, https://doi.org/10.6028/NIST.AI.100-1.

8. MIT Project NANDA. The GenAI Divide: State of AI in Business 2025. Massachusetts Institute of Technology, 2025, https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf.

9. Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. “Generative AI at Work.” The Quarterly Journal of Economics, vol. 140, no. 2, 2025, pp. 889–942. NBER Working Paper 31161, https://www.nber.org/papers/w31161.

10. Hämäläinen, Lari, and Steffen Fuchs. “The Decision Dividend: How AI Creates Economic Value.” McKinsey & Company, 26 Aug. 2026, https://www.mckinsey.com/industries/industrials/our-insights/the-decision-dividend-how-ai-creates-economic-value.