

CONTACT CENTER AI IS INHERITING A DATA CRISIS

Here's how to manage it.



CONTACT CENTER AI IS INHERITING A DATA CRISIS HERE'S HOW TO MANAGE IT.

AI has quickly become a defining investment priority for contact center leaders. Organizations are deploying virtual agents, automated quality management (AQM), real-time agent assistance, conversation intelligence, predictive routing, and generative call summaries.

The potential is significant. Contact centers produce enormous volumes of customer intelligence every day, including direct signals about customer intent, product issues, agent performance, operational friction, and emerging business risks.

The greatest obstacle to realizing that potential is rarely the model itself. The more persistent threat is the condition of the data surrounding it.

AI projects often expose years of fragmented systems, inconsistent definitions, unclear ownership, and weak governance practices.

A model can process information at tremendous speed, but it cannot independently repair the operational history embedded in that information. When the underlying records are incomplete, contradictory, or poorly governed, AI scales those weaknesses across the contact center.

That creates a difficult reality for leaders. The success of the next generation of contact center technology will depend heavily on data management decisions that many organizations have postponed for years.

CONTACT CENTER DATA WAS **NEVER DESIGNED FOR AI**

Most contact center data environments were developed incrementally.

- Telephony platforms were implemented to route and record calls.
- CRM systems were added to manage customer records and cases.
- Workforce management (WFM) platforms handled forecasting and scheduling.
- Quality management systems evaluated agent performance.

- Chat, email, social media, and messaging platforms arrived as customer channel preferences expanded.

Each system was generally purchased to solve a specific operational need. *But few of them were designed to provide AI with a complete, reliable, and continuously updated view of the customer journey.*

This fragmentation creates immediate challenges when organizations attempt to connect data across platforms.

A single customer issue may produce a call recording, transcript, chatbot conversation, CRM case, agent note, authentication record, quality score, and billing adjustment.

These records may use different identifiers. They may have different timestamps, ownership rules, and retention schedules.

Some systems will record that the case was closed, while others may show that the customer contacted the company again the following day.

An AI application must somehow determine which records belong together, what happened during the interaction, and whether the customer's underlying problem was actually resolved.

That becomes difficult when the organization itself lacks a consistent answer.

- Terms such as "resolution," "escalation," "conversion," and "customer intent" often carry different meanings across teams and platforms.
- One department may classify an interaction as resolved when the agent closes the case. Another may require confirmation that the customer's issue was permanently corrected. A digital channel may use an entirely different taxonomy.

These differences may appear minor in a dashboard or monthly report. But once AI begins using them to recommend actions, evaluate agents, or automate customer conversations, they become operationally significant.

AI MAKES DATA PROBLEMS MORE EXPENSIVE

Poor data quality has always affected contact center performance. It contributes to inaccurate reporting, weak forecasting, and inconsistent customer experiences (CXs).

AI increases the scale and speed of the impacts:

- An AQM system may evaluate every customer interaction rather than a small sample of them.
- An agent-assist tool may produce recommendations during thousands of conversations each day.
- A virtual agent may communicate directly with customers without a human reviewing every response.

When the inputs are unreliable, the resulting error(s) can spread quickly.

Consider an agent-assist system recommending a retention offer. The system may have access to the current transcript, product documentation, and selected CRM fields. But it may lack a recent billing adjustment, an unresolved complaint, or updated eligibility rules. Consequently:

- The recommendation can still sound polished and highly specific, even though critical context is missing.
- The presentation of confidence can make the recommendation more persuasive than the underlying data warrants.

Generative summaries create a similar risk. A summary may omit a commitment made by the agent, misidentify the cause of the call, or incorrectly state that the issue was resolved.

If that summary becomes part of the official customer record, future agents and systems inherit the mistake. Over time, organizations may begin training, evaluating, or grounding AI information generated by earlier AI systems.

Consequently, one error becomes a source for another. Without clear labeling and lineage, machine-generated interpretations can gradually become



SUBSCRIPTIONS AVAILABLE @
ContactCenterPipeline.com

indistinguishable from original customer data.

This creates a self-reinforcing data problem. The organization may believe it is learning from customers when it is increasingly learning from its own (including flawed) automated outputs.

UNSTRUCTURED DATA CARRIES HIDDEN RISKS

The most valuable contact center data is often unstructured. Call recordings, transcripts, chat messages, emails, screen recordings, and free-form agent notes contain far more context than traditional disposition codes.

These sources can reveal customer sentiment, emerging product defects, recurring service failures, and the reasons customers abandon transactions. But they can also contain highly sensitive information.

Customers may disclose payment details, medical information, account credentials, financial hardship, or personal circumstances during a conversation.

Agents may repeat that information and enter it into notes or copy it into fields that were never built to hold sensitive data or meet the compliance requirements that govern it.

Many agents also receive little or no training on which details qualify as sensitive or how those details are supposed to be handled once captured: which means the exposure often begins well before the data ever reaches an AI system.

Once an interaction is transcribed, that information becomes searchable and easier to distribute. A detail buried in an audio recording can suddenly appear in a data lake, analytics platform, model prompt, generated summary, or third-party processing environment.

Many enterprise data governance programs were built around structured databases. Those environments typically have defined fields, access controls, and retention policies. A customer conversation is much less predictable. Sensitive information can appear at any point and in any format.



EVERY MAJOR CONTACT CENTER DATA DOMAIN SHOULD HAVE A CLEARLY ACCOUNTABLE OWNER.

Contact center leaders therefore need to understand exactly how unstructured data moves through the AI lifecycle.

- Which recordings are being transcribed?
- Where are the transcripts stored?
- Are sensitive details redacted before they reach the AI application, whether that is a generative summarization tool, an agent-assist model or an analytics engine, and at what stage in the pipeline does redaction happen?
- Are prompts and outputs retained by a technology provider?
- Can the data be used to improve a third-party model?
- How are recordings, transcripts, summaries, and embeddings deleted when the retention period expires?
- Which regulations apply to this data (for example, industry-specific rules or regional privacy laws). Can the organization demonstrate compliance with each one across every system the data touches?

An organization that cannot answer these questions has limited control over one of its most valuable and sensitive data assets.

CONSENT NEEDS TO FOLLOW THE INTERACTION

Consent is another area where legacy contact center practices collide with modern AI use cases.

A customer may hear a notice stating that a call will be recorded for quality or training purposes. But organizations should not assume that this language provides unlimited permission for future analysis, model training, or automated decision-making.

The challenge grows when data moves across platforms or external providers.

The recording system may document that a notice was played, but the transcript sent to an analytics environment may not include that consent record.

Here's the risk: once the interaction enters a broader data repository, the restrictions governing its use can become difficult to trace. That does not, however, transfer accountability.

The organization that originated the data generally remains responsible for how it is used downstream, including compliance, even after the data has passed through a third-party platform or model. Contractual terms with vendors should make that responsibility explicit rather than leaving it ambiguous.

Consent should travel with the data. Interaction records need metadata that identifies how the information was collected, what uses are permitted, which jurisdiction applies, how long the data can be retained, and whether any sensitive information is present.

These controls should remain attached as the data moves between systems, states, countries, and their provinces or states. Recording requirements, privacy rights, data residency obligations, and restrictions on automated processing can vary by jurisdiction.

Organizations need governance mechanisms that apply these requirements consistently rather than relying on employees to interpret them during individual interactions.

THE GOVERNANCE GAP IS AN OWNERSHIP GAP

Technology alone will not resolve these issues because many of the underlying problems involve accountability.

In numerous organizations, responsibility for contact center data is distributed across Operations, IT, Security, Legal, Compliance, Analytics, and CX teams. Each group owns part of the environment. Few own the complete lifecycle.

- The contact center may determine which fields agents complete.

- IT manages the platforms.
- Legal defines retention policies.
- Security manages access.
- Analytics teams transform the data.

An external provider may process the information through an AI model. But when an output is inaccurate or challenged, responsibility becomes unclear.

Contact center leaders need a formal role in data governance because they understand how the information is produced in practice.

They know why agents skip fields, when disposition codes are unreliable, and how transfers create duplicate or incomplete interaction records.

They also understand that a closed case may still represent an unresolved customer problem. That operational context is critical when determining whether data is suitable for AI.

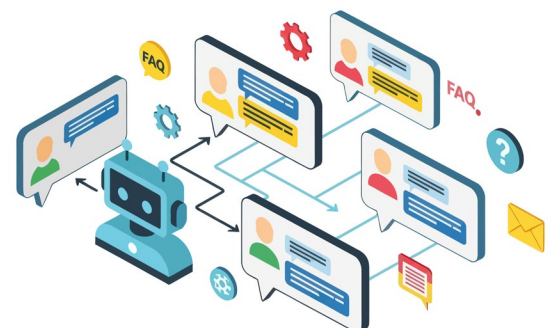
Every major contact center data domain should have a clearly accountable owner. That individual or team should define what the data means, how its quality is measured, who may use it, and how errors are corrected.

Organizations also need reliable data lineage. Leaders should be able to identify where information originated, how it was transformed, and which systems or models used it. When an AI recommendation is questioned, the organization should be able to reconstruct the relevant inputs and rules.

GOVERNANCE MUST REACH THE WORKFLOW

Many AI governance programs focus heavily on model approval. Teams review a model's accuracy, security, vendor terms, and technical architecture before deployment.

Those reviews are valuable, but model-level controls cannot compensate for weak operational data. Governance must extend through the entire workflow.



A contact center should know which data enters an application, how frequently it is updated, what happens when information conflicts, and which sources have priority.

Leaders should define when a human must verify an output, how suspected errors are reported, and what happens when model performance declines.

These controls need to be embedded directly into systems and processes. For example:

- An AI tool should be prevented from displaying an offer to an ineligible customer.
- Sensitive information should be redacted before it reaches the AI application processing it, whether that application generates summaries, powers agent-assist recommendations, or feeds an analytics model.
- A generated summary should be labeled as machine-produced.
- High-impact recommendations should require human validation.

Data outside its approved retention period should be deleted across recordings, transcripts, derived datasets, and model indexes.

A policy document cannot enforce these protections on its own. Technical controls and operational workflows must carry the governance burden.

START WITH THE DECISION, THEN EXAMINE THE DATA

Contact centers do not need to repair every historical data issue before implementing AI. They do need to understand which data problems could compromise a particular application. The most effective starting point is the decision the AI system will influence.

Leaders should define what the system is expected to do, who will rely on the output, and what happens when the output is wrong. They can then identify the minimum data required to support each use case responsibly.

The required standard should reflect the consequences of failure.

- A low-risk application that categorizes general call topics or tags interactions by subject for reporting purposes, may tolerate occasional errors.

That is because a mistake mainly affects internal analytics rather than the customer directly.

- A high-risk application, one that determines customer eligibility, provides financial or medical information, or recommends actions for vulnerable customers, requires stronger data validation, monitoring, and human oversight.

That is because an error there can directly affect a customer's finances, access to service, or wellbeing.

Organizations should also examine whether the data represents the full customer population.

- Historical interaction records may underrepresent certain languages, channels, demographic groups, or types of customer problems.
- A model trained on that history may perform well for common interactions while producing weaker results for customers whose needs appear less frequently in the data.

Continuous monitoring is essential because data environments change.

- Products are updated, policies shift, channels expand, and customer behavior evolves.
- A model that performed well during testing can degrade as its inputs move away from the conditions under which it was evaluated.

DATA DISCIPLINE WILL SEPARATE AI LEADERS FROM EXPERIMENTERS

Contact center AI will continue to advance:

- Virtual agents will handle more complex requests.
- Human agents will receive richer contextual support.

- Leaders will gain access to insights that previously remained buried across millions of customer conversations.

The organizations that benefit most will approach data as operational infrastructure.

- They will establish consistent definitions, preserve consent, track lineage, identify machine-generated content, and assign clear ownership.
- They will connect governance controls to actual workflows rather than relying solely on policies and review committees.

THE COMPANIES THAT ADDRESS THESE ISSUES NOW WILL BE POSITIONED TO USE AI WITH GREATER CONFIDENCE...

AI is forcing contact centers to confront data weaknesses that have accumulated through years of platform expansion and fragmented ownership. That pressure can be productive. It gives leaders an opportunity to create a more reliable foundation for customer service, analytics, and automation.

The companies that address these issues now will be positioned to use AI with greater confidence and at greater scale. But those that continue building on fragmented and poorly governed data will find that every new AI capability introduces another layer of operational risk. 



Frank Palermo is the Chief Operating Officer of NewRocket, where he helps guide the company's growth strategy and strengthens its position as a leading advisor in digital workflows, AI, and enterprise transformation. He brings decades of experience building and scaling technology and consulting organizations.

PREMIUM CONTENT + RESPECTED CONTRIBUTORS FOR CONTACT CENTER PROFESSIONALS

ENABLING A NEW GENERATION OF CX AND EX PROFESSIONALS to create successful customer management strategies, develop cutting-edge technologies, refine the skills necessary to advance their career, and build a culture that advances the contact center within the organization—that's what we do.

Since 2009, **Contact Center Pipeline** has leveraged the insight of today's notable CX and EX thought-leaders, along with our advisory board, expert magazine authors, blog contributors, and industry insiders; keeping our audience ahead of the trends transforming the contact center and customer sales, service, and support industries, improving outcomes, and the way companies engage with their customers.

Contact Center Pipeline Magazine is published monthly digitally and in print. For more information and to subscribe, visit our website.



INSIGHT AND INSPIRATION FOR CONTACT CENTER PROFESSIONALS

© COPYRIGHT 2026, CONTACT CENTER PIPELINE, INC. ALL RIGHTS RESERVED.

REPRODUCTION IN WHOLE OR IN PART WITHOUT WRITTEN PERMISSION FROM THE PUBLISHER IS PROHIBITED.
THE VIEWS EXPRESSED HEREIN ARE THOSE OF THE AUTHORS AND/OR SPONSORS AND DO NOT NECESSARILY REFLECT THE
OPINION OF THE OWNERSHIP OR MANAGEMENT OF CONTACT CENTER PIPELINE, INC. OR PIPELINE PUBLISHING GROUP, INC.