JBAI The Journal of Business and Artificial Intelligence
ISSN: 2995-5971 · Volume 3, Number 1

# The Third Era of Listening: From Suggestion Box to Survey to Conversation

Sonal Singhal

DOI: https://doi.org/10.66241/dvoly

**Sonal Singhal**
Sprinklr | Emory University

## Abstract

Enterprise listening has evolved through two distinct eras—each defined by a fundamental constraint on how organizations gathered signal from the people they served. The first was human and direct: suggestion boxes, town halls, one-on-one conversations. Rich in texture, impossible to scale. The second standardized the questions to enable aggregation: the Likert-scale survey. It scaled, but in doing so discarded the individual signal that made the first era valuable. We are now entering a third era. Agentic AI—systems capable of adaptive conversation at enterprise scale - holds the structural promise of resolving the trade-off that has constrained organizational listening for over a century, producing signal that is simultaneously rich and scalable. This commentary traces that evolutionary arc, examines why most current conversational AI deployments fall short of the structural promise, identifies the three failure points where enterprise deployments stall, and proposes a practical product architecture—Signal, Judgment, Action—for organizations and product leaders building genuine third era listening systems. The argument is not that surveys are obsolete. It is that they have reached the limit of what standardization can deliver, and that the organizations investing now in the architecture of the third era will operate on a qualitatively different informational foundation than those that do not.

## The First Era: When Listening Was Human

When Tokugawa Yoshimune, the eighth shogun of the Tokugawa shogunate in Japan, introduced the *meyasubako* - a box for petitions - to the Shogunate in 1721 (Lutz, R. C. 2022), the underlying challenge was identical to the one every large organization faces today: how do you hear from people whose daily experience you cannot directly observe? The instrument was primitive. The problem it addressed was not.

John Patterson institutionalized the employee suggestion system at the National Cash Register Company in the 1890s (Investment Masters Class, (2023)), and by the early twentieth century the model was standard practice across Western manufacturing. The logic was straightforward: workers knew things managers did not, and giving them a structured channel to communicate that knowledge produced both operational improvement and a measure of organizational trust (Nelson, 1975). Across factory floors, hospital wards, government offices—wherever large numbers of people did work that leadership could not fully see, some variant of the suggestion system appeared as a mechanism to close the gap.

What distinguished this era of listening was not its technological sophistication but the nature of the signal it produced. A suggestion, a complaint, a letter to management—these were human artifacts. They were authored by a specific individual describing a specific experience in their own words, unprompted by a question designer's assumptions about what was worth asking. When the signal was acted on, it could be traced back to its source. The organization could understand not just that something was wrong but where, for whom, and under what conditions.

The limitation was not quality but volume. A factory manager could read and respond to fifty suggestions a week. She could not read five thousand. As organizations grew through the late nineteenth and early twentieth centuries, the human model of listening became structurally unsustainable. The bottleneck was not willingness to listen, but the irreducible labor cost of processing unstructured individual signal at scale. Something had to give. What gave was the richness of the signal itself.

## The Second Era: Scale at the Cost of Depth

The standardized questionnaire as an organizational tool traces its roots to the late nineteenth-century social survey movement—seen in Charles Booth’s poverty mapping of London and the settlement house surveys of urban America. Its role in management accelerated in the mid-twentieth century as behavioral science entered mainstream organizational practice. Rensis Likert's 1932 paper introducing the numerical attitude scale that bears his name gave organizations a quantifiable instrument for measuring employee and customer sentiment that was simultaneously standardized, reproducible,

and aggregable (Likert, 1932). By the postwar period, the annual employee survey and the customer satisfaction questionnaire had become standard tools of organizational management across industries.

The core insight of the survey era was powerful: if you ask the same questions of everyone, you can compare responses across populations, track change over time, and identify statistical patterns invisible to any individual observer. The survey unlocked a form of organizational intelligence that the suggestion box never could—not the texture of individual experience, but the distribution of experience across a population. For most of the twentieth century, this represented genuine progress. Organizations could, for the first time, know with reasonable confidence whether engagement was higher in one business unit than another, whether satisfaction had improved or declined year over year, whether a particular initiative had moved the numbers.

But the survey solved the scale problem by accepting a trade-off whose cost has grown steadily as organizational complexity increased. To make signal aggregable, it had to be standardized - the same questions, in the same order, using the same response scales, regardless of individual context. The result was an instrument well-calibrated for measuring what the survey designer already knew to ask about, yet systematically blind to what they did not. An issue that had not yet been codified into a survey question could not appear in the data. An experience that did not map neatly onto a five-point scale was either distorted into one or lost entirely. The specific texture of organizational life—the workflow creating friction, the precise moment when a customer's loyalty broke, the specific manager whose behavior was driving attrition—was averaged away in the aggregation that made the data useful at scale.

The digital era accelerated the deployment of surveys without resolving their structural limitations. Platforms emerged that could distribute questionnaires to millions of recipients simultaneously and generate dashboards that made the output appear authoritative. Response rates declined—steadily and significantly, with many enterprise programs now operating at completion rates well below 20% for key employee and customer populations (Gallup, 2023). More troubling than the volume decline is its composition. Research consistently demonstrates that survey completion is correlated with extreme sentiment, recent negative experience, and higher organizational engagement—meaning the organizations relying most heavily on survey data are drawing conclusions from a self-selected sample that systematically misrepresents the broader population they are trying to understand (Tourangeau et al., 2013). The instrument that was designed to make signal representative is producing less representative signal the more sophisticated it becomes.

By the second decade of the twenty-first century, the survey paradigm was yielding diminishing returns. The instruments were more sophisticated, the platforms more

capable, the dashboards more elegant—yet the organizational intelligence they generated was increasingly disconnected from the lived experience they claimed to measure. The trade-off that the second era accepted in order to scale was becoming the dominant fact of the data it produced.

## The Third Era: Conversation at Scale

Agentic AI systems capable of genuinely adaptive conversation—not scripted chatbot flows but systems that probe based on emerging context, follow the thread of an individual's experience, and extract structured meaning from open-ended expression - represent the first structural opportunity in the history of organizational listening to escape the richness-versus-scale trade-off entirely. Conversation, unlike the survey, does not require the designer to know in advance what questions are worth asking. It does not force individual experience into a standardized format. It can pursue the unexpected, respond to the specific, and adapt its line of inquiry based on what is actually being communicated—not what was anticipated. And with AI as the conversational agent, it can do this simultaneously across thousands of interactions, at the scale the suggestion box never achieved and the survey only achieved by giving up richness.

This is the structural promise. The gap between that promise and what most enterprise deployments of conversational AI are currently delivering is large, and understanding that gap is the prerequisite for closing it.

Most systems marketed today as conversational listening are surveys with a chat interface. Across deployments, the pattern is consistent: pre-authored questions presented in sequence, free-text responses funneled into sentiment scoring engines, and outputs that are structurally identical to what a Likert-scale instrument would have generated. A common implementation—scripted decision trees paired with NLP-based sentiment classification—illustrates the limits of this approach: the branching logic is still designer-defined, the topics still pre-constrained, and the "score" still the destination. The interaction feels more natural than clicking radio buttons, and free-text responses contain marginally more texture than numerical ratings—but if the underlying logic is designer-driven, sequential, and topic-constrained, the system has not entered the third era. It has digitally redecorated the second.

Gartner's projection that 33% of enterprise software will include agentic AI by 2028, up from less than 1% today (Gartner, 2025), has been widely cited as evidence that the third era is arriving. What is less cited is Deloitte's parallel finding that only 11% of organizations currently have any agentic AI in production (Deloitte, 2025), and Gartner's own warning that over 40% of agentic AI projects may be abandoned by 2027 due to unclear business value (Gartner, 2025). These are not signs of a technology failing—they

are signs of a classification problem. Organizations are investing in systems they believe to be genuinely agentic, discovering that the systems behave like sophisticated surveys, and attributing the gap to the technology rather than to the architectural choices that created it.

The distinction that separates a third-era system from a second-era system with a conversational interface is whether the system genuinely adapts. An adaptive conversational agent follows the individual: if someone mentions a specific manager, it asks about that manager; if someone's account reveals a pattern the designer did not anticipate, it pursues that pattern; if someone signals distress, it responds to the person before returning to the organizational question. The output resembles what a skilled human interviewer would gather in a one-on-one conversation—specific, contextual, responsive to the individual—rather than what any questionnaire could produce. And critically, unlike the skilled human interviewer of the first era, the AI agent can conduct thousands of these conversations simultaneously without degradation in quality.

## Where Enterprise Deployments Stall

Three failure points account for the majority of the gap between the structural promise of conversational AI and the operational reality of most enterprise deployments. Each corresponds to an architectural layer that organizations are consistently underbuilding.

The first is the design instinct. Organizations that have spent decades optimizing survey instruments bring the survey mental model to conversational AI design. They author conversation flows the way they author questionnaires - topic by topic, question by question, with conditional branch logic replacing adaptive intelligence. The result is a system that feels conversational but thinks in survey logic. Building genuinely adaptive systems requires a different starting discipline: defining what the organization needs to learn rather than what it plans to ask, designing probing logic that pursues organizational goals through individual dialogue, and building quality measures around signal sufficiency rather than question completion. This is not a technology problem. It is a product design problem, and it manifests before a single line of code is written.

The second failure point is the interpretation gap. A survey produces a database of numerical responses immediately queryable by any analyst. A genuine conversation produces unstructured text that must be interpreted before it can inform a decision. Extracting structured, actionable insights from thousands of simultaneous open-ended conversations is now technically possible - but operationally complex. Reliable theme categorization, sentiment analysis, and mapping to the organizational constructs decision-makers actually use all degrade in predictable ways across multiple languages, cultural contexts, and domain-specific vocabularies. Many organizations have discovered this only

after deployment, when their conversational system was producing richer data than their survey platform and less decision-ready output.

***Practical idea**: Interpretation quality doesn't announce itself when it fails - it quietly surfaces as low confidence in the outputs. A baseline QA protocol helps - sample 2–5% of interpreted conversations weekly for human review, stratified by language and region rather than randomly drawn, since interpretation error concentrates in non-English and lower-resource language contexts. Calibration sessions between the AI's theme taxonomy and the organization's internal language - what "escalation" means in a contact center versus a product team, for instance - should be scheduled at deployment and revisited quarterly. The operational lift is modest; the alternative is discovering the gap when a decision has already been made on degraded signal.*

Finally, the third and most consequential failure point is the action gap. The survey era built the bridge to action - routing individual responses to follow-up queues, triggering cases, updating CRM records - but never completed the crossing: systemic patterns remained unrouted, and organizational response lagged too far behind the signal to matter. A listening system that can close the loop on a single dissatisfied customer but cannot route a systemic pattern to the person with authority to fix its cause has optimized for responsiveness theater. Completing that layer - not just at the individual ticket level but at the level of systemic organizational response - is the product design challenge that defines whether a conversational AI system is genuinely third-era or merely a richer survey with a chat interface.

## The Signal – Judgment - Action Architecture

For product leaders building third-era listening systems, three architectural layers determine whether a conversational AI deployment transcends the survey paradigm or merely imitates its interface.

**Signal.** Third-era signal design begins with the organizational question, not the conversation script. What does the organization need to know? What decisions will this intelligence inform? What does a high-quality signal look like in this context, and how will the system recognize when it has gathered one? Building backward from these questions - defining the probing logic, the extraction model, the sufficiency thresholds - is a fundamentally different discipline from writing survey items. It requires close collaboration between product design, data science, and the organizational leaders who will act on the output, and it requires that collaboration before the system is built. Organizations that skip this step and begin with the conversational interface discover, after deployment, that they have a system capable of generating interesting conversations and incapable of generating actionable intelligence.

**Judgment.** In a survey system, all judgment is human and downstream - analysts interpret aggregated data after collection is complete. In a genuinely agentic system, judgment is embedded in the system itself: the agent decides when to probe deeper and when it has gathered sufficient signal, when to escalate a conversation to a human responder and when to proceed autonomously, when a pattern in the data warrants immediate action and when it represents normal variance. Designing this layer means specifying in advance, with precision, which decisions the system is authorized to make without human review and which require it. Organizations that treat this boundary as a default rather than a design decision discover it through failure: systems that over-escalate destroy the efficiency case for automation; systems that under-escalate create liability and erode trust. The boundary between machine judgment and human judgment is a product design decision, and the organizations that make it deliberately - and instrument their systems to audit it over time - are the ones whose agentic deployments survive contact with production.

**Action.** This is the layer that closes the loop the suggestion box opened and the survey never completed. In the first era, feedback could in principle be traced to response - the manager read the note, changed the policy, told the worker what changed. The survey era severed that connection: signal went in, scores came out, and the path from individual feedback to organizational action became invisible to the person who had provided the feedback and often to the organization itself. Third-era systems, built with the action layer as a first-order architectural requirement, restore that connection at scale: automatically routing urgent individual signals to the right human responder within a defined time window, triggering process reviews when patterns across conversations reach defined thresholds, and - where governance structures permit - initiating responses autonomously for categories of signal where the appropriate response is well-defined and the cost of delay is high. This is where the organizational value of the transition from survey to conversation actually lives, and it is where most current enterprise deployments stall.

**A note on governance:** Routing at this level of automation is not a purely technical decision. Who receives an escalated signal, under what conditions, and with what visibility to the individual who provided the feedback are questions that touch privacy regulation, labor agreements, and organizational trust. The organizations that will get this right are those that treat consent frameworks, escalation criteria, and audit trails as design inputs - not compliance checkboxes applied after the architecture is already set.

## Conclusion

Three eras. One persistent problem. Organizations have always needed to understand the experience of the people they serve - employees, customers, the communities within which they operate - and have always lacked a mechanism simultaneously rich enough to be

meaningful and scalable enough to be useful. The suggestion box was rich but could not scale. The survey scaled but discarded richness. Agentic conversational AI offers, for the first time, a credible path to both – not only because the technology is impressive, but because its structural properties directly address the constraint that has defined each of the two previous eras.

Whether that promise is realized in any given enterprise deployment depends on architectural choices that most current implementations are not making. A conversational interface layered over survey logic is not a third-era system. An AI that produces rich unstructured signal without a judgment layer to interpret it and an action layer to close the loop has improved the input without improving the outcome. The organizations that understand this distinction and invest in the full architecture - signal design, judgment design, action design - rather than the interface alone will be operating on a qualitatively different informational foundation than their competitors over the next few years.

The suggestion box had a good run. The survey had a longer one. What replaces them will be better - but only if the organizations building it do so with clear eyes about what the third era actually requires.

## Disclosure Statement

No confidential or customer-identifiable information is included. All opinions, errors, and omissions are the author’s own.

## References

Deloitte. (2025). Agentic AI strategy. Deloitte Insights: Tech Trends 2026. https://www.deloitte.com/us/en/insights/topics/technology-management/tech-trends/2026/agentic-ai-strategy.html

Gallup. (2023). State of the global workplace report. Gallup Press.

Gartner. (June 25, 2025). Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027. Gartner Press Release. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22(140), 1–55. https://psycnet.apa.org/record/1933-01885-001

Lutz, R. C.. (2022). Tokugawa Yoshimune. https://www.ebsco.com/research-starters/history/tokugawa-yoshimune

Investment Masters Class. (2023). Learning from National Cash Register's John H Patterson. https://mastersinvest.com/newblog/2022/10/20/learning-from-national-cash-registers-ncr

Ajay Bandi, Bhavani Kongari, Roshini Naguru, Sahitya Pasnoor, and Sri Vidya Vilipala. (2025). The Rise of Agentic AI: A Review of Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges. *Future Internet, 17(9), 404*. https://doi.org/10.3390/fi17090404

Nelson, D. (1975). Managers and workers: Origins of the new factory system in the United States, 1880–1920. University of Wisconsin Press. https://archive.org/details/managersworkerso0000nels

Tourangeau, R., Edwards, B., Johnson, T. P., Wolter, K. M., & Bates, N. (Eds.). (2013). Hard-to-survey populations. Cambridge University Press. https://assets.cambridge.org/97811070/31357/frontmatter/9781107031357_frontmatter.pdf

## About the Author

### Sonal Singhal

Director of Product Management at Sprinklr | Emory University

The author is Director of Product Management at Sprinklr, where she leads an AI-native product line including conversational AI agents for enterprise CX workflows. Previously at Medallia, she led Experience Management products analyzing billions of daily data signals across enterprise employee and customer listening programs serving Fortune 10 companies. The views expressed are the author's own and do not represent the positions of Sprinklr, Medallia, or their affiliates. No proprietary customer data or confidential product information has been disclosed.

LinkedIn

## Copyright Notice

## About the Journal

The Journal of Business and Artificial Intelligence (ISSN: 2995-5971) is the leading publication at the nexus of artificial intelligence (AI) and business practices. Our primary goal is to serve as a premier forum for the dissemination of practical, case-study-based insights into how AI can be effectively applied to various business problems. The journal focuses on a wide array of topics, including product development, market research, discovery, sales & marketing, compliance, and manufacturing & supply chain. By providing in-depth analyses and showcasing innovative applications of AI, we seek to guide businesses in harnessing AI's potential to optimize their operations and strategies.