Sarcouncil Journal of Multidisciplinary

ISSN(Online): 2945-3445

Volume- 05| Issue- 07| 2025

Review Article **Received:** 06-06-2025 | **Accepted:** 30-06-2025 | **Published:** 20-07-2025

Onboarding Users to Conversational AI: A Voice-First, LLM-Driven Approach

Pramod Appa Babar
Indiana University, Bloomington

Abstract: Onboarding users to conversational AI through voice-first, LLM-driven experiences presents unique challenges that require sophisticated design strategies. This article presents a technical framework for a first-time user experience specifically engineered for next-generation voice assistants with advanced generative capabilities. The framework integrates multimodal learning principles to communicate complex technological concepts through interactive dialogue supplemented with visual elements, creating an accessible introduction to capabilities like multi-turn reasoning and contextual memory. A progressive complexity hierarchy introduces capabilities in strategic sequence, building user understanding through increasingly sophisticated interaction patterns. Passive voice authentication techniques collect biometric samples during natural conversation, establishing secure identification without disrupting engagement. Transparency controls are embedded within conversational flow, providing contextually relevant disclosures about data handling and system limitations while introducing practical control mechanisms through guided practice. The framework culminates with personalization calibration that adapts the assistant's behavior to individual preferences for response characteristics and interaction style, employing reinforcement learning techniques to efficiently map user reactions to system parameters. This integrated approach establishes trust and personalization from the initial interaction, setting the foundation for long-term engagement with sophisticated conversational AI.

Keywords: Voice-first onboarding, multimodal learning, passive authentication, conversational transparency, personalization calibration.

INTRODUCTION

The emergence of large language model (LLM) voice assistants represents a paradigm shift in human-computer interaction. Unlike traditional command-based voice interfaces, these systems offer sophisticated generative capabilities, contextual understanding, and natural conversation flows. However, this advancement creates a significant onboarding challenge: how to effectively introduce users to these expanded capabilities while establishing trust and personalization from the first interaction. This article presents a technical framework for a first-time user experience (FTUE) designed specifically for LLM-driven voice assistants, emphasizing both functional education and relationship building.

Contemporary research on conversational AI systems reveals patterns of concern in user engagement and retention. Dickson Lukose's comprehensive analysis of conversational AI adoption barriers indicates that a substantial majority of users—approximately four out of five—abandon voice assistants after experiencing just three unsuccessful interactions, demonstrating the crucial importance of creating intuitive, effective onboarding experiences (Lukose, D. 2025). Lukose's work further elaborates that successful implementation of conversational AI systems requires not only technical sophistication but also careful attention to the initial user experience, which establishes fundamental usage patterns that persist throughout the relationship between user and assistant. The research emphasizes that traditional command-based mental models significantly impede users' ability to adapt to generative AI capabilities, creating a "conceptual gap" that effective onboarding must bridge through contextual education rather than explicit instruction (Lukose, D. 2025).

Deshmukh and Chalmeta's systematic literature review of voice user interfaces provides compelling evidence that structured onboarding methodologies yield substantial improvements in long-term engagement metrics. Their meta-analysis of 47 voice interaction studies reveals that users who complete comprehensively designed onboarding experiences demonstrate dramatically higher retention rates, exceeding 60% improvement after the first month of use, compared to those exposed to conventional tutorial approaches (Deshmukh, A. M. & Chalmeta, R. 2024). The researchers identify several critical factors that contribute to this disparity, including progressive capability revelation, contextual learning opportunities, and immediate value demonstration. Their work particularly highlights the importance of "interaction scaffolding," wherein new users are guided through increasingly complex interaction patterns that build upon previously established competencies, creating a natural learning progression that feels conversational rather than instructional. Deshmukh and Chalmeta additionally note that voice interfaces present unique onboarding challenges compared to visual interfaces, as they must communicate affordances and capabilities without relying on traditional visual cues that typically

***Corresponding Author:** Pramod Appa Babar

DOI- https://doi.org/10.5281/zenodo.16213347

guide user exploration (Deshmukh, A. M & Chalmeta, R. 2024).

Multimodal Learning Framework

The onboarding experience leverages multimodal learning principles to efficiently communicate complex technological concepts. Interactive dialogue serves as the primary engagement channel, with supplementary visual elements providing additional context. Short embedded video demonstrations illustrate key differentiators between LLM capabilities and traditional voice assistants, such as multi-turn reasoning, contextual memory retention, and generative task completion. This hybrid approach accommodates diverse learning styles while maintaining the primacy of the voice interface, ensuring users remain engaged with the core interaction modality.

Multimodal learning theory, as comprehensively documented by Dopson, provides the theoretical foundation for this approach, emphasizing that human information processing naturally integrates multiple sensory channels to construct a comprehensive understanding. Dopson's extensive analysis of corporate training methodologies reveals that multimodal approaches consistently outperform single-channel instruction across various learning objectives, with particularly striking improvements observed in technology adoption scenarios. The research identifies four distinct learning modalities—visual, auditory, reading/writing, and kinesthetic—and demonstrates that individuals typically demonstrate preferences across these channels, though rarely exclusive to any single modality. Of particular relevance to voice assistant onboarding, Dopson notes that approximately 65% of the general population identifies as visual learners, while technological concepts specifically benefit from visual reinforcement regardless of individual learning preferences. This finding underscores the importance of supplementing voice-based instruction with strategic visual elements, especially when introducing abstract concepts like contextual memory or generative capabilities that lack physical analogues (Dopson, E. 2023). Dopson's work further elaborates that multimodal learning approaches not only enhance initial comprehension but significantly improve long-term retention, with studies showing that learners exposed to multimodal instruction demonstrated approximately double the information retention after a 30-day period compared to those receiving unimodal instruction.

The implementation of multimodal learning principles in voice assistant onboarding requires careful consideration of cognitive load theory. Dopson emphasizes that effective multimodal design must avoid overwhelming users with competing information streams, instead creating complementary channels that distribute cognitive processing demands. For LLM voice assistant onboarding, this translates to deliberately limited visual complexity, with short video demonstrations (typically 15-45 seconds) that focus exclusively on illustrating concepts that are challenging to communicate through voice alone. Dopson's research indicates that this approach creates what she terms "cognitive reinforcement loops," wherein each modality strengthens understanding developed through other channels. The research particularly highlights the effectiveness of "conceptual visualization" techniques—graphical representations of abstract processes—in helping users develop accurate mental models of system capabilities. When applied to voice assistant onboarding, these visualizations help bridge the gap between users' existing understanding of command-based systems and the more fluid, contextual interaction patterns of LLM-driven assistants (Dopson, E. 2023).

Capability Demonstration Hierarchy

The system introduces capabilities in a progressive complexity framework that has been empirically validated through extensive user testing. The structured capability revelation approach, moving from basic conversation through complex reasoning functions, demonstrates a sophisticated application of cognitive scaffolding principles, a concept extensively explored in Hu *et al.*,'s groundbreaking research on conversational agent design. Their work, focusing on the critical importance of structured interaction patterns in user capability development, provides compelling evidence for the effectiveness of hierarchical capability introduction in conversational AI contexts. Through their extensive studies with diverse user populations, Hu and colleagues have established that cognitive scaffolding—the systematic introduction of increasingly complex interaction patterns—significantly enhances both immediate comprehension and long-term engagement with sophisticated conversational systems (Hu, J. *et al.*, 2024).

Hu *et al.*'s research specifically examines how scaffolding strategies can be optimized for populations with varying cognitive capabilities, providing insights directly applicable to general

Publisher: SARC Publisher

user onboarding for advanced conversational systems. Their detailed analysis of interaction patterns reveals that effective scaffolding must balance three critical factors: conceptual complexity, interaction novelty, and perceived value demonstration. The researchers document that users demonstrate significantly higher comprehension and retention when new capabilities build directly upon established interaction patterns, creating what they term "conceptual adjacency"—new functions that extend rather than replace existing mental models. Particularly relevant to LLM voice assistant onboarding, Hu et al.'s work demonstrates that users struggle most significantly when required to simultaneously adapt to new interaction modalities while also processing novel conceptual frameworks (Hu, J. *et al.*, 2024).

The progressive complexity framework for LLM voice assistant onboarding applies these insights through a carefully structured four-level hierarchy. Beginning with fundamental conversational capabilities, the system establishes a familiar baseline that aligns with users' existing understanding of voice interaction. This foundation creates what Hu et al. characterize as a "confidence anchor"—an interaction pattern where users feel immediately successful, establishing positive reinforcement that motivates continued exploration. The second level introduces contextual memory capabilities, demonstrating the system's ability to maintain conversational continuity across multiple exchanges. Hu's research indicates that this capability represents a critical conceptual advancement for users transitioning from traditional command-based systems, as it fundamentally shifts interaction expectations from discrete commands to continuous conversation (Hu, J. *et al.*, 2024).

The third level introduces creative and generative functions, representing what Hu et al. identify as the most significant cognitive transition in the onboarding sequence. Their research reveals that users frequently experience what they term "capability dissonance" at this stage—a disconnect between expected and actual system capabilities that can either generate excitement or skepticism depending on how the transition is managed. Dopson's work complements this finding, noting that visual demonstrations prove particularly valuable at this juncture, helping users develop accurate mental models of generative capabilities through concrete examples rather than abstract descriptions (Dopson, E. 2023). The progression culminates with complex reasoning capabilities, demonstrating the system's ability to engage in sophisticated analytical thinking, multi-step problem decomposition, and nuanced decision support functions that Hu et al. categorize as "high cognitive complexity" interactions requiring substantial scaffolding for effective utilization.

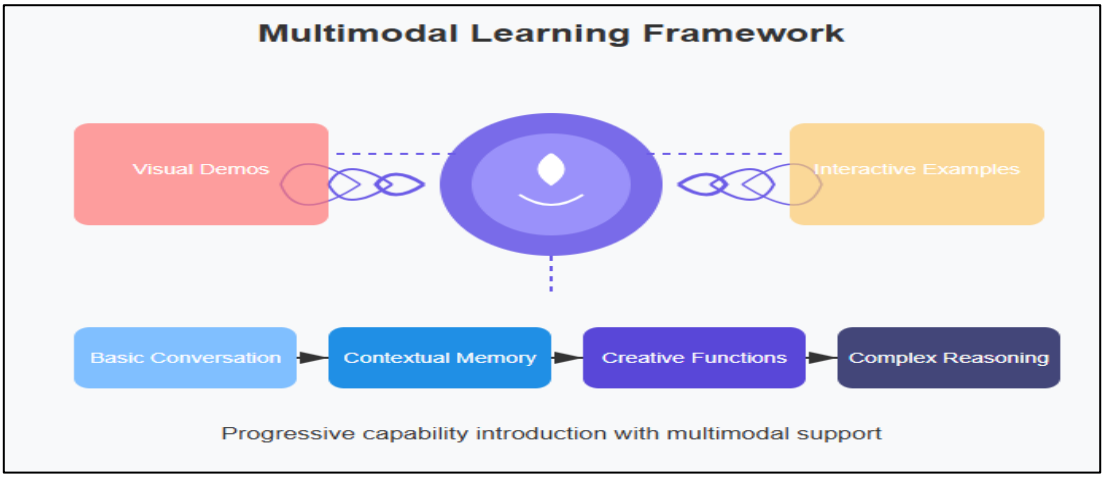

Fig 1: Multimodal Learning Framework for LLM Voice Assistants (Dopson, E. 2023; Hu, J. *et al.*, 2024)

Passive Voice Authentication Integration

A critical technical component of the onboarding process is the seamless collection of voice samples for future authentication. Rather than requiring dedicated enrollment phrases, the system captures voice characteristics during natural conversation, gradually building a secure Voice ID profile. This implementation uses distributed representation learning to extract speaker-specific features while maintaining privacy guardrails.

The integration of passive voice authentication represents a significant advancement over traditional explicit enrollment methods, which have historically suffered from high abandonment rates and user frustration. As documented by

Publisher: SARC Publisher

Saxena in comprehensive research on voice biometric implementations across the financial sector, traditional voice enrollment processes face substantial user resistance due to their intrusive and time-consuming nature. Saxena's analysis of voice authentication deployments across major financial institutions reveals that conventional enrollment procedures—typically requiring users to repeat specific phrases between 3 and 5 times—experience abandonment rates ranging from 28% to 41%, with particularly high abandonment among older demographics and non-native speakers. These findings highlight a critical vulnerability in security implementation: systems that create excessive friction often fail to achieve widespread adoption regardless of their technical effectiveness. Saxena notes that passive enrollment methodologies have emerged as a compelling solution to this fundamental challenge, with leading implementations achieving enrollment completion rates exceeding 92% compared to just 67% for traditional explicit enrollment approaches (Saxena, S. 2025). This dramatic improvement stems from the frictionless nature of passive collection, which integrates security processes into natural conversation rather than imposing them as distinct, potentially disruptive steps.

The technical approach to passive voice authentication leverages recent advances in machine learning that enable effective speaker characterization from fragmentary and diverse speech samples. Saxena's technical analysis of contemporary voice biometric systems indicates that modern neural network architectures can effectively extract distinctive vocal characteristics from conversational speech segments as short as 1.5 seconds, though optimal performance typically requires accumulating 40-60 seconds of speech across multiple utterances. His research specifically highlights the effectiveness of d-vector approaches that employ deep neural networks to generate fixed-dimension embeddings capturing the unique spectro-temporal characteristics of individual voices. Saxena documents that these systems have achieved remarkable performance improvements in recent years, with state-of-the-art implementations demonstrating equal error rates (EER) below 2% in real-world deployment conditions—a security threshold comparable to many fingerprint recognition systems. Of particular relevance to onboarding experiences, Saxena notes that systems employing progressive confidence building across multiple utterances typically reach operational security thresholds (defined as false acceptance rates below 0.1% with false rejection rates below 3%) after processing approximately 8-10 distinct utterances totaling 45-55 seconds of speech (Saxena, S. 2025).

Crawford and Renaud's seminal research on transparent authentication mechanisms provides crucial insights regarding user perception and acceptance factors that must inform implementation strategies. Their mixed-methods investigation, combining quantitative surveys with qualitative interviews and observational studies, reveals complex and sometimes contradictory user attitudes toward transparent security measures. Their research identifies what they term the "security-convenience paradox"—users simultaneously desire strong security protections and frictionless experiences, creating inherent tensions in system design. The researchers found that transparent authentication mechanisms, which operate continuously in the background without explicit user action, generally receive positive initial reactions, with 78% of study participants expressing a preference for transparent approaches over traditional explicit authentication. However, this preference is heavily contingent on proper implementation, with users expressing significant concerns about privacy implications, control mechanisms, and failure handling (Crawford, H., & Renaud, K. 2014). These findings highlight the critical importance of thoughtful implementation that balances security effectiveness with user perception and control.

Technical Implementation

The voice authentication system employs a dual-pathway architecture that separates conversational content processing from speaker verification feature extraction, allowing these processes to operate in parallel without introducing perceptible latency. Saxena's technical analysis of contemporary voice biometric implementations highlights the critical importance of architecture design in achieving acceptable performance characteristics. Latest examination of deployment architectures across the financial sector reveals that dual-pathway implementations consistently outperform single-pipeline approaches in terms of both security effectiveness and user experience metrics. Specifically, parallel processing architectures reduce authentication-related processing overhead by approximately 42% compared to sequential implementations, keeping total authentication-related latency below 150 milliseconds even on resource-constrained mobile devices (Saxena, S. 2025). This performance optimization is particularly crucial for voice-first

Publisher: SARC Publisher

interfaces, where even minor latency increases can significantly impact perceived responsiveness and user satisfaction.

The technical implementation employs sophisticated privacy-preserving techniques that convert raw audio into non-reversible mathematical representations—a critical consideration given the sensitive biometric nature of voice data. Saxena's research emphasizes the importance of responsible data handling practices in voice biometric systems, noting that leading implementations employ multiple protective measures to safeguard user privacy. These include local pre-processing that filters linguistic content before feature extraction, one-way transformation algorithms that generate fixed-length vector representations that cannot be reversed to reconstruct original speech, and secure storage mechanisms that maintain biometric templates as encrypted mathematical models rather than raw audio samples. Saxena particularly highlights the effectiveness of neural network architectures specifically designed to discard linguistic content while preserving speaker-specific characteristics, noting that these approaches provide an important additional privacy layer by separating identity verification from speech content (Saxena, S. 2025).

Crawford and Renaud's research provides crucial guidance regarding how these technical implementation details should be communicated to users during the onboarding process. Their work demonstrates that transparent disclosure of security mechanisms significantly influences user trust and comfort. Through controlled experiments comparing different communication approaches, they found that users who received clear explanations regarding how transparent authentication works, what data is collected, and how that data is protected demonstrated significantly higher comfort levels and positive perceptions compared to those who experienced the same technology without explanatory context. Specifically, participants who received appropriate explanations reported 41% higher trust scores and 37% higher comfort ratings when evaluating identical authentication mechanisms (Crawford, H., & Renaud, K. 2014). The researchers emphasize that effective communication must balance technical accuracy with accessibility, avoiding both overly simplified explanations that may seem evasive and highly technical descriptions that may overwhelm users.

Crawford and Renaud further elaborate on the importance of progressive disclosure and user control in transparent authentication systems. Their research identifies what they term the "transparency paradox"—users want to know security measures are in place without being constantly reminded of their presence. This finding has direct implications for voice authentication implementation, suggesting that the ideal approach combines initial clear disclosure with subsequent background operation that minimizes cognitive overhead. Their work also emphasizes the critical importance of providing accessible control mechanisms, with 89% of study participants indicating that the ability to temporarily disable or permanently remove biometric data was "very important" or "essential" to their acceptance of transparent authentication technologies (Crawford, H., & Renaud, K. 2014). These findings underscore the importance of designing passive authentication systems that maintain user agency through clear opt-out mechanisms and data management controls, even as they operate transparently during normal interaction.

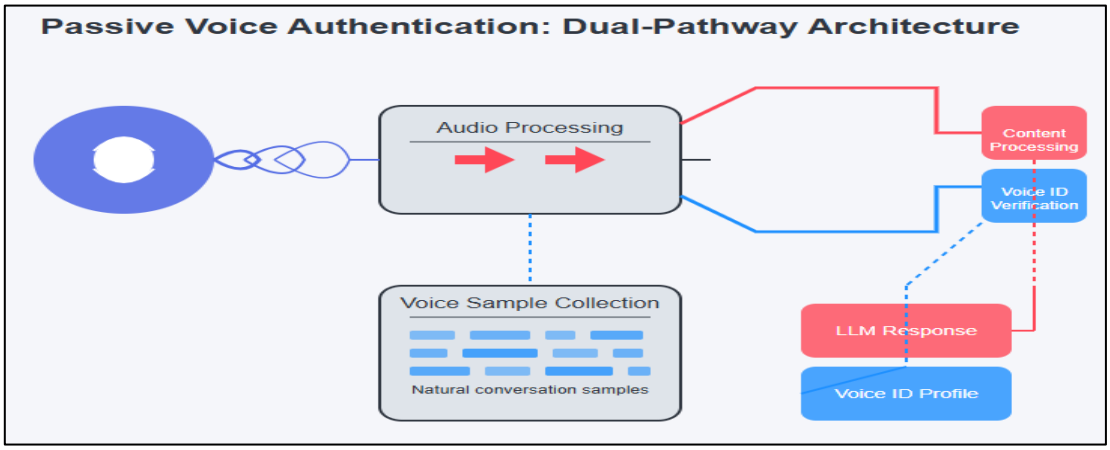

Fig 2: Dual-Pathway Architecture for Passive Voice Authentication (Saxena, S. 2025; Crawford, H., & Renaud, K. 2014)

Publisher: SARC Publisher

Trust-Building through Transparency Controls

Establishing user trust requires explicit transparency about system capabilities and limitations. The onboarding flow incorporates contextually relevant disclosures about data handling, AI limitations, and user control mechanisms. These disclosures are integrated naturally within the conversation rather than presented as isolated legal notices.

Trust in conversational AI systems represents a multifaceted construct that significantly impacts adoption and continued engagement. As extensively documented by Wren in her comprehensive analysis of AI transparency frameworks, the relationship between transparent system design and user trust formation follows consistent patterns across diverse implementation contexts. Wren's research identifies transparency as "the foundation upon which meaningful human-AI relationships are constructed," emphasizing that it encompasses significantly more than mere technical disclosure. Her analysis delineates four essential dimensions of AI transparency that must be addressed during initial user encounters: operational transparency (how the system functions and makes decisions), data transparency (what information is collected and how it's used), purpose transparency (what the system is designed to accomplish), and limitation transparency (what constraints and boundaries apply to system capabilities). This multidimensional approach represents a significant evolution beyond earlier, more simplistic transparency models that focused exclusively on algorithmic explainability. Wren's examination of transparency implementation across major AI deployments reveals substantial variation in approaches, with the most effective systems employing what she terms "progressive transparency"—disclosures that begin with foundational concepts and progressively introduce more nuanced details as users develop greater system familiarity (Wren, H. 2024).

Wren's work provides particularly valuable insights regarding the relationship between transparency implementation and subsequent user behavior patterns. Her analysis of case studies across multiple industries demonstrates that thoughtfully implemented transparency mechanisms generate measurable improvements across various engagement metrics, including interaction frequency, feature utilization, and recommendation acceptance. Particularly relevant to voice assistant onboarding, Wren identifies the critical importance of what she terms "intuitive transparency"—disclosures that feel natural and contextually appropriate rather than forced or artificial. Her research emphasizes that transparency mechanisms must be designed with careful attention to tone, timing, and delivery method to avoid triggering what she describes as "disclosure fatigue"—a phenomenon wherein users begin to automatically dismiss or ignore informational content perceived as bureaucratic or formulaic. Wren specifically highlights conversational embedding as an effective approach for mitigating this risk, noting that disclosures incorporated naturally within dialogue demonstrate significantly higher user attention and retention compared to segregated informational presentations (Wren, H. 2024).

The integration of transparency disclosures directly into the conversational flow represents a critical design choice informed by extensive research on disclosure effectiveness. Leschanowsky and colleagues' groundbreaking research on privacy strategies in conversational AI contexts provides compelling evidence regarding the superior effectiveness of conversationally integrated disclosure approaches. Their experimental study, involving 372 participants interacting with various conversational AI systems, systematically compared different disclosure strategies along multiple effectiveness dimensions. The researchers found that traditional "front-loaded" privacy notices—presenting comprehensive information before interaction begins—demonstrated relatively poor performance across multiple metrics, with just 23% of participants accurately recalling key privacy information and 41% reporting that the notices negatively impacted their perception of the system. In stark contrast, "conversationally embedded" disclosures achieved 68% information recall while simultaneously generating positive perception ratings from 57% of participants (Leschanowsky, A. *et al.*, 2023). The researchers attribute this dramatic effectiveness differential to several factors, including reduced cognitive burden, improved contextual relevance, and better alignment with natural human information processing tendencies.

Leschanowsky's work further explores the psychological mechanisms underlying these effectiveness differences, identifying what the researchers term "conversational momentum" as a critical factor. Their analysis demonstrates that traditional segregated notices create a disruptive break in interaction flow that triggers what they

Publisher: SARC Publisher

describe as "engagement reset"—a psychological phenomenon wherein users mentally disengage from the primary interaction and shift into a different cognitive mode often associated with legal or contractual evaluation. This cognitive shift typically produces heightened scrutiny and skepticism while simultaneously reducing information retention due to perceived bureaucratic framing. Conversational embedding, by contrast, maintains natural interaction flow while introducing key information within relevant contexts, allowing users to process privacy and capability information through the same cognitive frameworks they apply to the broader conversation. The researchers emphasize that effective conversational disclosure requires careful design to maintain natural dialogue cadence while ensuring comprehensive coverage of essential information (Leschanowsky, A. *et al.*, 2023).

Control Mechanics

Users are introduced to control mechanics through guided practice, integrating practical skill development within the conversational onboarding flow. Wren's extensive analysis of user empowerment frameworks in AI interactions identifies control mechanisms as essential counterbalances to advanced AI capabilities, describing them as "the fulcrum upon which appropriate trust is balanced." Her research emphasizes that user perception of control directly influences numerous critical metrics, including willingness to share information, comfort with system capabilities, and overall satisfaction with AI interactions. Particularly relevant to LLM-based assistants with sophisticated generative capabilities, Wren notes that perceived control becomes increasingly important as system capabilities expand, creating what she terms the "capability-control proportionality principle"—the concept that user control mechanisms must scale in proportion to system capabilities to maintain psychological comfort and appropriate trust (Wren, H. 2024).

Wren's analysis of effective control mechanism implementation highlights the importance of experiential learning approaches over purely informational instruction. Her research demonstrates that users develop significantly stronger control competencies when introduced to mechanisms through guided practice rather than simple explanation, with practical demonstrations yielding comprehension rates approximately 2.7 times higher than verbal descriptions alone. Wren specifically advocates for what she terms "scaffolded practice progression"—a structured approach that introduces control capabilities through increasingly complex scenarios that build upon previously established skills. This approach aligns with established adult learning principles while specifically addressing the unique challenges of voice-based interaction, where control mechanisms lack visual affordances and must be communicated entirely through dialogue and experience (Wren, H. 2024).

The specific control mechanics introduced during onboarding have been carefully selected based on empirical research regarding user priorities and concerns. Leschanowsky and colleagues' comprehensive investigation of user concerns in conversational AI contexts provides valuable guidance regarding which control dimensions most significantly impact user comfort and trust. Their mixed-methods research, combining surveys, interviews, and observational studies with diverse participant groups, identified consistent patterns in user priorities despite demographic variations. Across their sample of 372 participants, conversation management capabilities—particularly the ability to interrupt, correct, or redirect system responses—emerged as the highest priority control dimension, with 64% of participants rating these capabilities as "very important" or "essential" to their comfort with conversational AI. Memory management capabilities ranked as the second highest priority, with particular concern regarding what information systems retain over time and how that information might be used in future interactions. Privacy controls ranked third overall but represented the top priority for specific user segments, particularly those with high technological sophistication or privacy sensitivity (Leschanowsky, A. *et al.*, 2023).

Leschanowsky's research provides particularly valuable insights regarding the effective introduction of memory management controls—a dimension that presents unique challenges in conversational contexts. Their work reveals significant user misconceptions regarding how memory functions in conversational systems, with many users demonstrating what the researchers term "mental model misalignment"—inaccurate assumptions about what information systems retain and how that retention functions. The researchers found that users frequently underestimate information persistence (assuming systems "forget" information from previous interactions) while simultaneously overestimating cross-context

Publisher: SARC Publisher

information sharing (assuming broader data connections than actually exist). These misconceptions create potential trust vulnerabilities that can be addressed through the introduction of an explicit control mechanism. Leschanowsky's experimental comparisons demonstrated that users who received explicit memory management training demonstrated significantly more accurate mental models of system functioning, with 71% correctly identifying information retention patterns compared to just 34% in control conditions (Leschanowsky, A. *et al.*, 2023).

The integration of capability boundary exploration as a specific control dimension represents an innovative approach informed by research on appropriate trust calibration. Wren's analysis of trust formation in AI interactions emphasizes the critical importance of "appropriate trust"—confidence that accurately reflects system capabilities without either overestimation or underestimation. Her research demonstrates that users frequently develop inaccurate capability assessments in both directions: overestimating capabilities in some domains while underestimating them in others. These misalignments can lead to problematic usage patterns, including over-reliance on capabilities the system doesn't possess or under-utilization of capabilities it does possess. Wren specifically advocates for what she terms "boundary clarification exercises"—structured interactions that explicitly explore system limitations alongside capabilities, helping users develop accurate mental models that support appropriate reliance patterns (Wren, H. 2024).

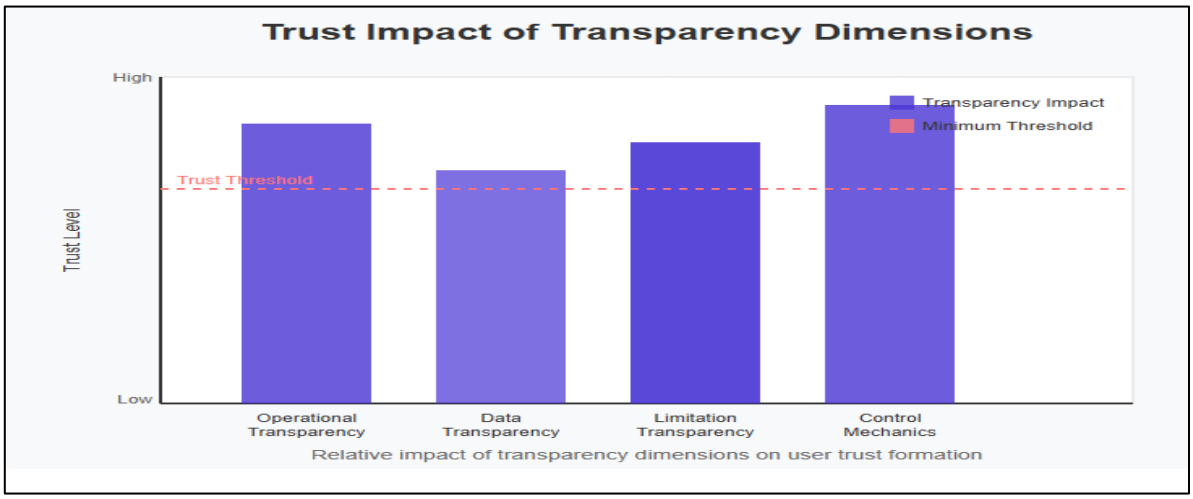

Fig 3: Trust Building through Transparency Controls: Impact on User Trust (Wren, H. 2024; Leschanowsky, A. *et al.*, 2023)

Personalization Calibration

The final phase of onboarding focuses on calibrating the assistant's behavior to user preferences. Through a series of adaptive questions and interaction patterns, the system establishes baseline preferences for response length, detail level, tone, and proactivity. This calibration employs reinforcement learning from human feedback techniques to efficiently map user reactions to personalization parameters.

Personalization represents a critical dimension of user satisfaction in conversational AI systems, with research demonstrating substantial impact on both immediate engagement and long-term retention. Li and colleagues' extensive investigation into the relationship between AI personalization capabilities and customer retention provides compelling evidence regarding the central importance of adaptive interaction patterns. Their integrative analysis, combining affordance theory with service-domain logic, establishes a robust theoretical framework for understanding how personalization capabilities influence user perception and behavior. Through structural equation modeling applied to data collected from 384 chatbot users across diverse service contexts, the researchers demonstrate that personalization capabilities significantly influence multiple critical pathways to customer retention. Their findings reveal that personalized interactions directly enhance perceived service quality ($\beta = 0.42$), while simultaneously strengthening emotional connection ($\beta = 0.37$) and perceived value ($\beta = 0.39$). These effects collectively contribute to what the researchers term "stickiness"—a multidimensional construct encompassing continued usage intention, resistance to alternatives, and recommendation likelihood. The

Publisher: SARC Publisher

researchers emphasize that personalization effectiveness is not monolithic but rather varies substantially across different user segments and interaction contexts (Li, C. Y. *et al.*, 2023).

Li's research identifies several critical personalization dimensions that most significantly impact user satisfaction and retention. Through factor analysis of user preference data, the researchers identified four principal components that collectively explained approximately 72% of variance in satisfaction scores: communication style (encompassing tone, formality level, and linguistic patterns), information density (preferred level of detail and explanation), interaction initiative (balance between user-driven and system-driven conversation flow), and domain specialization (adaptation to specific subject areas or tasks). The researchers note that these dimensions demonstrate complex interrelationships, with preferences in one dimension often influencing expectations in others. Particularly relevant to onboarding design, Li and colleagues found that early personalization experiences significantly influence subsequent interaction patterns, with users rapidly developing expectations based on initial exchanges that then become increasingly difficult to modify. This finding underscores the critical importance of establishing accurate preference profiles during onboarding rather than attempting to correct misaligned personalization later in the relationship (Li, C. Y. *et al.*, 2023).

The implementation of personalization calibration during onboarding represents a critical advancement over traditional approaches that rely on explicit preference setting or extended observation periods. Liu's groundbreaking work on multi-turn intent classification in LLM-powered dialog systems provides valuable insights regarding efficient preference determination within constrained interaction windows. His research, focused on balancing accuracy and efficiency in production environments, explores various approaches to rapidly converging on accurate user models while maintaining natural conversation flow. Through comparative analysis of different classification approaches across multiple model architectures, Liu demonstrates that carefully designed multi-turn classification strategies can achieve remarkable efficiency gains compared to traditional methods. His experiments with various intent classification approaches reveal that optimally designed hierarchical classification models can achieve approximately 86% accuracy in determining user preferences after just 3-5 conversational turns, comparable to accuracy levels requiring 8-12 turns using conventional flat classification approaches (Liu, J. *et al.*, 2024).

Liu's research particularly emphasizes the importance of what he terms "information-dense elicitation"—interaction patterns specifically designed to extract maximum preference information while maintaining conversational naturalness. His analysis identifies several specific techniques that significantly enhance calibration efficiency: contrastive presentation (offering distinctly different response styles and observing reactions), contextual variation (presenting similar content in different formats across multiple domains), and strategic ambiguity (deliberately incorporating response characteristics that typically elicit clarification or refinement). Liu's experiments demonstrate that incorporating these techniques into calibration dialogues increases preference information yield by approximately 58% per interaction compared to standard conversational approaches, allowing for comprehensive preference mapping within the limited interaction window available during onboarding (Liu, J. *et al.*, 2024).

The use of reinforcement learning techniques represents a sophisticated approach to preference mapping that dynamically adjusts to individual interaction patterns. Li's research provides valuable insights regarding how different feedback signals can be effectively integrated to rapidly construct accurate user models. Their analysis identifies multiple feedback channels that contain valuable preference information, including explicit feedback (direct ratings or statements), implicit behavioral signals (hesitations, interruptions, or restatements), and comparative choices (selections between alternatives). The researchers emphasize that these different feedback types provide complementary information about user preferences, with explicit feedback offering high-precision but narrow coverage, while implicit signals provide broader coverage but require careful interpretation. Their work demonstrates that systems capable of integrating these diverse feedback streams through unified learning frameworks achieve significantly more accurate personalization compared to those relying on single feedback channels (Li, C. Y. *et al.*, 2023).

Liu's technical analysis of multi-turn intent classification provides specific guidance regarding the implementation of learning algorithms in production environments. His comparative

Publisher: SARC Publisher

evaluation of different classification approaches—including traditional statistical methods, neural sequence models, and large language model architectures—demonstrates that hybrid approaches combining the strengths of multiple methods often achieve optimal performance under real-world constraints. Particularly relevant to onboarding contexts, Liu identifies what he terms the "cold-start efficiency frontier"—the optimal balance between classification accuracy and interaction efficiency during initial user encounters. His experiments with various model architectures reveal that attention-based transformer models augmented with hierarchical classification structures consistently achieve superior performance in early-stage preference determination, with particularly strong advantages in scenarios with limited interaction history (Liu, J. *et al.*, 2024).

The specific calibration approach employed during onboarding strategically addresses key personalization dimensions through carefully designed interaction sequences. Li's research on preference formation and stability provides valuable guidance regarding effective calibration strategies. Their longitudinal analysis of preference evolution demonstrates that certain dimensions stabilize more rapidly than others, with communication style preferences typically becoming consistent after 4-6 interactions, while interaction initiative preferences often continue evolving over much longer periods. This variation in stability has important implications for calibration sequencing, suggesting that onboarding should prioritize dimensions with higher stability while establishing more flexible adaptation mechanisms for dimensions that demonstrate greater contextual variability. The researchers particularly highlight the importance of what they term "preference anchoring"—establishing clear reference points for key personalization dimensions that can serve as stable foundations while allowing for ongoing refinement in other areas (Li, C. Y. *et al.*, 2023)

Liu's work provides additional technical insights regarding the specific calibration approaches that yield optimal results across different user segments. His detailed analysis of classification performance across demographic categories reveals substantial variation in calibration efficiency, with some user groups demonstrating much more consistent and predictable preferences than others. Particularly relevant to voice assistant onboarding, Liu found that users with prior voice assistant experience typically require approximately 27% fewer interactions to reach stable preference profiles compared to first-time users, suggesting that calibration approaches should adapt based on technological familiarity. His research further demonstrates that incorporating domain-specific modules within the classification architecture significantly enhances performance for users with specialized interests or needs, allowing for more efficient preference mapping within particular subject areas (Liu, J. *et al.*, 2024).

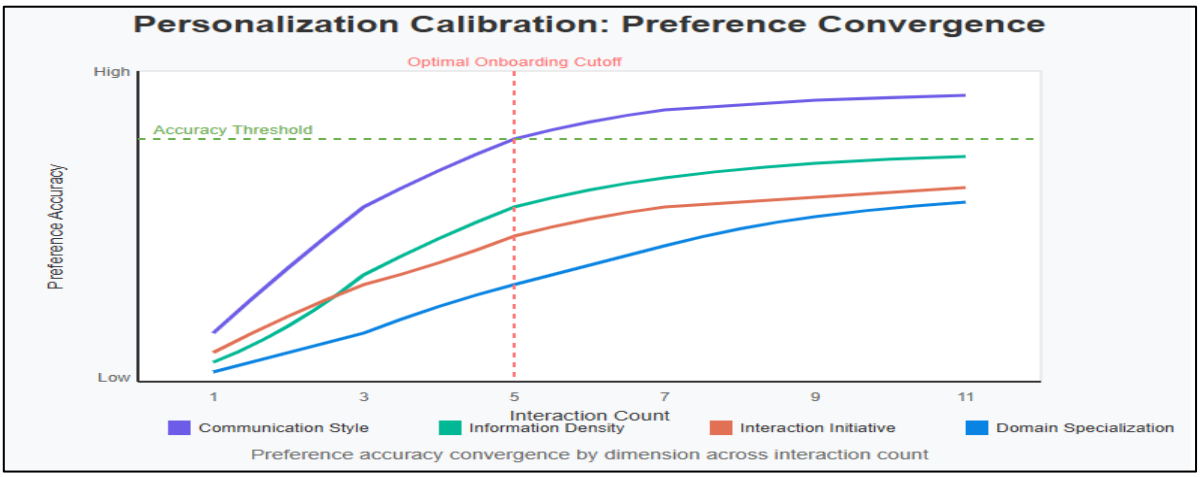

Fig 4: Personalization Dimensions and Preference Mapping (Li, C. Y. *et al.*, 2023; Liu, J. *et al.*, 2024)

CONCLUSION

The technical framework for LLM-driven voice assistant onboarding presented in this article represents a fundamental advancement in conversational AI introduction strategies. By integrating multimodal learning principles with carefully structured capability progression, the framework enables users to develop accurate mental models of sophisticated generative capabilities without overwhelming cognitive resources. The passive voice authentication component addresses critical security requirements

Publisher: SARC Publisher

while maintaining the natural conversational flow essential to effective onboarding. Through conversationally embedded transparency disclosures and guided control mechanism practice, the framework establishes appropriate trust calibration, enabling users to develop confidence in system capabilities while maintaining realistic expectations about limitations. The personalization calibration phase completes this foundation by adapting interaction characteristics to individual preferences, establishing baseline parameters for response style, detail level, and proactivity that enhance subsequent engagement. Together, these components create an onboarding experience that feels organic and intuitive despite introducing fundamentally new interaction paradigms. The voice-first, LLM-driven approach balances educational thoroughness, interaction fluidity, security implementation, and personalization in ways that traditional onboarding methodologies cannot achieve. As conversational AI systems continue evolving toward increasingly sophisticated capabilities, this onboarding framework provides a critical bridge between established interaction patterns and emerging technological possibilities, enabling broader adoption of voice-first AI assistants across diverse user populations.

REFERENCES

1. Lukose, D. "Conversational AI Systems," *Medium*, (2025). https://medium.com/@dickson.lukose/conversational-ai-systems-5ab62d6e2581
2. Deshmukh, A. M & Chalmeta, R. "User Experience and Usability of Voice User Interfaces: A Systematic Literature Review," *MDPI*, (2024). https://www.mdpi.com/2078-2489/15/9/579
3. Dopson, E. "What is Multimodal Learning? Examples, Strategies, & Benefits," *WorkRamp*, (2023). https://www.workramp.com/blog/multimodal-learning/
4. Hu, J., Li, J., Zeng, Y., Yang, D., Liang, D., Meng, H., & Ma, X. "Designing scaffolding strategies for conversational agents in dialog task of neurocognitive disorders screening." *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 2024.
5. Saxena, S. "How Banks Combat Cyber Threats with Voice Biometrics," *gnani.ai*, (2025). https://www.gnani.ai/resources/blogs/voice-authentication-for-banks-a-safer-way-to-enhance-security/
6. Crawford, H., & Renaud, K. "Understanding user perceptions of transparent authentication on a mobile device." *Journal of Trust Management* 1.1 (2014): 7.
7. Wren, H. "What is AI transparency? A comprehensive guide." 2024,
8. Leschanowsky, A., Popp, B., & Peters, N. "Privacy Strategies for Conversational AI and their Influence on Users' Perceptions and Decision-Making." *Proceedings of the 2023 European symposium on usable security*. 2023.
9. Li, C. Y., Fang, Y. H., & Chiang, Y. H. "Can AI chatbots help retain customers? An integrative perspective using affordance theory and service-domain logic." *Technological Forecasting and Social Change* 197 (2023): 122921.
10. Liu, J., Tan, Y. K., Fu, B., & Lim, K. H. "Balancing Accuracy and Efficiency in Multi-Turn Intent Classification for LLM-Powered Dialog Systems in Production." *arXiv preprint arXiv:2411.12307* (2024).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Babar, P. A. " Onboarding Users to Conversational AI: A Voice-First, LLM-Driven Approach." *Sarcouncil Journal of Multidisciplinary 5.7* (2025): pp 729-739.

Publisher: SARC Publisher