In the evolving world of customer support, conversational AI has become a key player. From chatbots on websites to voice agents in call centers, companies like **Suprmind** and **Air Canada** utilize these technologies to enhance customer experience. But a pressing question emerges: between voice agents and chatbots, which one hallucinates more, and why?

Before diving in, let's clarify what "hallucination" means in this context. Often misused as a catch-all for AI errors, hallucination specifically refers to the generation of inaccurate or fabricated responses that appear plausible, but lack grounding in factual data. Understanding the nuances — and real sources of truth — is critical for improving **voice vs text agent accuracy**.

Seven Failure Points in Voice Agents

Voice agents face unique challenges compared to text chatbots, which contribute to higher error rates and hallucinations. Based on insights from deployments by companies like Suprmind and Air Canada, here are seven common failure points:

1. **Speech-to-Text (STT) Errors**: Ambiguities and inaccuracies in transcribed speech lead to downstream misunderstandings.
2. **Entity Misrecognition**: Poor capture of key customer data such as phone numbers, reservation codes, or account IDs.
3. **Context Tracking**: Difficulty maintaining the conversation's true intent over several dialog turns.
4. **Knowledge Base Gaps**: Outdated or incomplete information causing AI to guess responses.
5. **Latency and Real-Time Constraints**: Limits on processing times lead to oversimplifications or truncated answers.
6. **Text-to-Speech (TTS) Synthesis Artifacts**: Mispronunciations or unnatural pacing confuse customers and obscure meaning.
7. **Dialog Management Errors**: Inadequate handling of interruptions, clarifications, or confirmation prompts.

Each failure point cascades into greater hallucination risk—especially if no robust source of truth or verification is embedded in the workflow.

Retrieval-Augmented Generation (RAG): Benefits and Limits

Many modern systems integrate **RAG (retrieval-augmented generation)** to ground responses in a knowledge base rather than relying solely on pretrained language models. This has shown promise in reducing hallucinations by referencing live data. However, RAG has its boundaries:

- **Knowledge Base Hygiene**: The quality and freshness of the indexed documents directly impact output accuracy. Stale or incorrect information induces hallucinations.
- **Retrieval Precision**: Retrievers may fetch irrelevant or ambiguous passages, amplifying error.
- **Compositional Challenges**: Language models might still generate fabricated conclusions by synthesizing retrieved snippets incorrectly.
- **Latency Constraints**: Especially critical in voice agents, slower RAG retrieval and integration can degrade real-time interactions.

Organizations must prioritize rigorous knowledge base maintenance and employ robust retrievers to minimize these failure modes.

Live Tools as Source of Truth for Customer-Specific Facts

One concrete strategy to tackle hallucinations is integrating live verification tools and external APIs as ground truth sources—especially for sensitive unsupported claim rate customer-specific data. For example:

- **Reservation Lookup APIs**: Air Canada uses live access to reservation databases to confirm itinerary details rather than relying solely on AI model memory.
- **Account Verification Systems**: Ensures that the voice agent or chatbot cross-checks account numbers or payment info with real-time backend checks.
- **Call Transcript and Audio Archives**: Examine real call snippets—such as "B three one seven two"—to build entity confirmation patterns and evaluation datasets.

These live tools serve as the true "source of truth" and reduce dependency on model-based inference alone, minimizing hallucination risks substantially.

High-Precision Entity Confirmation and Readback

Another best practice critical in voice systems is **high-precision entity confirmation and readback**. Whether it's a tracking number, membership ID, or flight code, correctly decoding and confirming these entities is vital:

- *Multi-Modal Confirmation:* Confirming entities both audibly and visually (where applicable) reduces mishearings.
- *Explicit Readbacks:* Voice agents read back entities to customers to ensure accuracy, e.g., "Is that B three one seven two?"
- *Phonetic Spelling Techniques:* Using phonetic alphabets or segmented spelling to reduce ambiguity in speech recognition.
- *Confidence Thresholding:* Leveraging STT confidence scores to trigger repeat or clarification steps only when uncertainty exceeds thresholds.

Suprmind's implementation of these techniques in their tau-Voice benchmark has demonstrated measurable improvements in **voice vs text agent accuracy**, particularly in real deployments subject to noisy environments and diverse accents.

Comparing Hallucination Tendencies: Voice Agent vs Chatbot

Key question: do voice agents hallucinate more than text chatbots?

Aspect Voice Agent Text Chatbot Source Input Speech-to-Text (error-prone) Direct text input (lower error rate) Entity Capture Challenging, needs high-precision confirmation Easier with typed data; less ambiguity Knowledge Grounding Often relies on RAG plus live backends Commonly RAG with large knowledge DBs or APIs Model Latency Constraints Stricter real-time limits may sacrifice accuracy More flexible, can batch or wait longer Hallucination Impact Higher risk due to compounded errors and modality challenges Lower, but still susceptible to knowledge base errors

Put simply, voice agents face additional layers of complexity and noise that can amplify hallucination risk. However, rigorous implementation practices—supported by tools like **speech-to-text and text-to-speech pipelines**, **live**

data integration, and **precise entity confirmation**—can bring this risk closer to parity with text-based systems.

Insights from OpenAI and tau-Voice Benchmarking

OpenAI, a key player in foundational and applied conversational AI, recognizes these challenges and actively incorporates retrieval and grounding methods in its APIs. Their models when deployed with RAG help reduce hallucination but depend heavily on how knowledge bases are managed.

The **tau-Voice benchmark**, promoted by organizations like Suprmind, offers empirical measurement of model failures in real-time telephony scenarios, capturing the nuanced differences in **realtime model failures** between voice and text agents. Insights from tau-Voice highlight that:

- STT transcription errors contribute to over 30% of unavoidable failures.
- Entity readback protocols reduce miscommunication rates by 40% or more.
- Clean and frequently updated knowledge bases cut hallucination events by half.

This data-driven approach is essential to move beyond buzzwords and deliver measurable accuracy improvements in both voice and text conversational agents.

Conclusion: The Real Source of Truth Matters Most

Hallucinations in customer support AI systems are not randomly occurring glitches but often stem from identifiable failures in input capture, knowledge grounding, and verifier integration. While voice agents currently exhibit more hallucinations than chatbots due to layered pipeline complexities, applying robust engineering controls—such as live source of truth tools, exacting entity confirmation, and stringent knowledge base hygiene—helps close the gap.

As companies like Suprmind and Air Canada continue to innovate, combining advanced RAG methods with real-time telephony-aware evaluation via tau-Voice benchmarks offers a practical path forward. Rather than simply labeling errors as hallucinations, what truly matters is understanding their root cause and optimizing system components accordingly.

To wrap up, always ask: *"What is the source of truth for that sentence?"*—and rely on live, verified data wherever possible. That mindset is the real guardrail against hallucination in any conversational AI, voice or text.