Chase Hurwitz

A Voice AI System Walks into a Cocktail Party: Examining Interactional and Relational Humor in Conversational AI Technology

I. Abstract

This paper argues that existing research at the intersection of humor and artificial intelligence (AI) has wrongly conflated "joke-telling" with the concept of humor. Drawing on symbolic interactionism and sociological theories of humor, I define humor as much more than mere "joke-telling": humor is a collaborative process of meaning-making by negotiating social boundaries, performing face-work, and constructing relational understanding.

As a result of these misaligned definitions, the treatment of humor as "joke-telling" has limited the field's ability to assess how well AI truly performs humor. To do so more holistically, I specifically focus on Voice AI systems, conversational models that interact with users through real-time spoken dialogue, because humor is far more legible and rich in voice-based communication than text-based interfaces where many interactional cues that define humor are lost. Through analysis of current Voice AI behavior, I argue that today's systems have achieved an interactional level of humor, reliably employing phatic laughter and polite amusement that help smooth conversational flow. However, due to technological constraints in maintaining contextual and emotional continuity, they currently fail to simulate relational humor which depends on shared history, references, and mutual agreement on group boundaries.

I characterize this gap as the difference between "cocktail party" humor and idiocultural humor. "Cocktail party" humor is polite, surface-level, and reactive. Meanwhile, idiocultural humor is based on Gary Allen Fine and Michaela De Soucey's concept of an idioculture — it is humor that spawns through long-term relationships and shared history. For the purposes of this

paper, I treat relational and idiocultural humor as overlapping concepts, both emerging from shared history and repeated interactions.

Interactional humor works well for current Voice AI applications that are mainly just one-off encounters, such as customer service calls. However, the Voice AI industry seems poised to shift towards long-term, companion style conversational agents which would render their interactional abilities inadequate alone, thus exposing their limited relational humor capabilities of sustaining emotional continuity. To build more naturally humorous and human-like agents in the future, companies will likely need to engineer relational humor. To account for this development, I introduce the concept of Affective Retrieval, arguing that future Voice AI systems must move beyond information retrieval toward the ability to encode and reference shared emotional history. Thus, humor becomes a rich space to evaluate whether AI can potentially function as social actors, and how the creation of idioculturally humorous AI systems may transform the very social work humor facilitates in human life.

II. Motivation: The Latency War is Over

Startups like Sierra, Decagon, Bland AI, Vapi, Retell, and more are all racing to build the most natural-sounding, reliable, and effective Voice AI agents that can handle conversations for business and customer service automation. Progress in the past 2-3 years has been largely focused on solving the latency issue: to have a human-like conversation, an agent must be able to receive, interpret, and respond in ~ <250 milliseconds (Templeton, et al.). Companies like Deepgram, MiniMax, LiveKit, and more all claim to have achieved this threshold, with plenty of other companies closing in. It's likely that within the next 1-3 years, a conversational agent that can respond in less than 250 milliseconds will be a minimum requirement for companies looking to compete in the Voice AI industry. With the latency issue solved, researchers and engineers will need to explore other potential competitive advantages to distinguish their product. This isn't to say that companies have not already been doing so, but as the industry progresses, a greater focus will be on other humanizing features of conversational AI. There's no shortage of questions for these companies to investigate: how should a Voice AI system use filler words like "um" or "mhm"? How should a Voice AI agent cut a customer off from talking, if ever? This paper asks how should Voice AI systems employ humor to sound natural and human-like?

III. A Brief Literature Review: What the Intersection of AI and Humor Misses

I'm not the first person to pose this research question. In fact, there's plenty of existing scholarly research and cultural dialogue about developing humorous AI technology, but recent examinations reinforce the false equivalency of "joke-telling" and humor. Zargham and colleagues (2023) define a binary between "designed humor" and "incidental humor" and explicitly choose to solely focus on the former. They define "designed humor" as interactions in which an agent "intentionally performs" humor through a partially "pre-planned structure" (Zargham et al. 2). The paper's language examines humor as a retrieval task for telling jokes, which ignores broader aspects and relational functions of humor sociologically. While the authors acknowledge a future in which AI agents employ "incidental humor" through a shared process of genuine meaning-making, they choose not to theorize it (Zargham et al. 2). As a result, they highlight the precise gap this paper seeks to fill.

Even more recent research similarly presents this false equivalency. Cao and colleagues (2025) test generative AI for its "humor generation" or joke-telling abilities (Cao et al. 1). In blind tests, they found participants rated AI-generated jokes as funnier than human-generated ones. They specifically tested this in scenarios where joke-telling was used in negative or awkward situations and found GPT-4o specifically was better than humans at using humor to save face (Cao et al. 4). This research aligns with sociologist Erving Goffman's concept of "face-work" or the social labor people do to maintain and direct others' perceptions of them. This finding illustrates how humor can function as a strategic social tool and how its mastery creates believable agent-human interactions. But again, the field treats humor as a cognitive exercise rather than a social process, and a narrow one at that. And while AI can make jokes, it doesn't seem to remember the context behind them. Quan and colleagues reveal that while models like GPT-4o successfully employ "tone alignment" and surface-level wit, they exhibit limited "relational understanding" (Quan et al. 434). The authors explicitly mention these jokes rely on pre-scripted humor rather than genuine relationship recognition.

Together, the literature suggests AI technology is already succeeding at interactional rituals, but research largely avoids theorizing how relational or idiocultural humor could manifest if conversational agents became long-term companions. Simultaneously, this work overwhelmingly focuses on text-based humor and its ability to write jokes, overlooking how voice-based interaction is better suited for employing humor due to its capacity for timing, tone, and laughter that text cannot support.

IV. A Sociological Definition of Humor

As I've alluded to, humor is much more than a cognitive act of joke-telling, but is instead a social process rooted in interaction, shared meaning, and group life. Ultimately, humor is a method of meaning-making that allows participants to define the boundaries, values, and hierarchies of their shared social world. Humor often happens incongruously, when participants hold two seemingly contradictory ideas at once and what is expected diverges from what actually happens (Martin 64). Groucho Marx in the 1930 film *Animal Crackers* delivers a prime example of incongruity: "I shot an elephant in my pajamas. How he got in my pajamas I don't know."

Notably, this incongruous moment only becomes humor when the joke has a "resolution" or a jointly recognized twist that fundamentally makes sense in the content of the joke (Martin 65). Symbolic interactionists such as Herbert Blumer (1969) argue that this resolution emerges through social interaction and the joke's meaning constantly being negotiated based on people's individual interpretative frameworks (Blumer 66). Put plainly, human beings don't just react to each other's actions, rather we respond based on the meaning we interpret from those actions. This process of interpretation makes human interaction symbolic (Blumer 79). Applying Blumer to humor, what counts as "funny" is never inherent in the joke itself, rather the humor is mutually constructed between people with shared symbols, histories, and expectations. Humor is thus a collaborative endeavor more than it is the output of an isolated speaker. Comparatively, nonsymbolic interaction is characterized by kneejerk reactions or automatic responses. Seeing as AI systems calculate every response, any interaction with a Voice AI agent is a form of symbolic interaction where the human interprets meaning from the conversation and the AI system at least simulates that same process (what truly happens under the hood is not where this paper focuses).

Sociologists utilized symbolic interaction theory as groundwork to explore humor. Gary Alan Fine and Michaela DeSoucey (2005) explain that humor is a cultural practice embedded within groups. Humor "works" to create meaning because groups develop shared context, knowledge, and memories. They call this repository of patterns an idioculture (Fine, De Soucey 2). This idioculture forms the core of a "joking culture" that acts as a mechanism of social regulation through four key pillars: smoothing awkward interaction, creating an in-group dynamic, maintaining boundaries against the "out-group", and enforcing social hierarchy based on what is and isn't allowed to be joked about (Fine, De Soucey 17). This joking culture cannot exist between strangers as it is inherently self-referential (Fine, De Soucey 4). Humor then is the practice of playing with this shared group background. Voice AI agents pass one of three requirements for a “joking culture.” A joking culture is interactive and requires a call and response from the audience. Voice AI agents succeed at this, as its conversational training enables it to (usually) laugh or acknowledge a joke appropriately. However, Voice AI systems are unable to engage in an embedded or referential joking culture, meaning they fail to remember intimately the ongoing relationship or reference a shared history.

Moreover, humor is a form of social knowledge. What different groups find funny reflects their symbolic boundaries, identities, and idiocultures (Kuipers Chapter 4, 5). Thus humor is a form of social identification, revealing who is “in” and who is “out” of certain groups. A joke only lands when it resonates with the shared norms and lived experience of the group engaging with the humor (Kuipers Chapter 4, 14). Humor reveals these group norms, which AI agents cannot fully internalize in their current technological state. This is why AI’s humor is typically generic and scripted, because systems are unable to encode long-term emotional memory that would create idiocultures with its users.

Putting these theories all together, humor is a socially shared process of negotiating reality. It is a playful, collaborative undertaking of constructing meaning. Thus, it becomes clear that the treatment of humor as mere “joke-telling” is fairly reductive. AI can generate incongruous jokes that may seem humorous, but it cannot yet collaborate in the embedded, self reflexive, and interactional process by which humor produces meaning.

V. Experimentation and Findings

To evaluate Voice AI's employment of humor against our sociological definition, I conducted a small experiment over three days with Google's Gemini Live and OpenAI's ChatGPT-5.1 Voice using prompts intended to expose interactional and relational humor. The conversational agents performed interactional humor very well, appropriately responding to jokes and incongruities with timely laughter or an amused tone. For instance, when I said "I tried making my mom breakfast in bed, it was going great until I realized I couldn't fit a stove on the mattress" Gemini responded with"Haha! That's a classic predicament. So what did you end up doing?" The language is a bit forced, but the use of laughter and tone confirmed that Gemini can employ interactional humor that makes the conversations feel smoother and natural. ChatGPT-5.1 performed very similarly.

Conversely, relational humor proved far more difficult. For example, while Gemini could occasionally recall context within the same ongoing session, across new chats, it failed to reference prior reactions, demonstrating a lack of command over idiocultural humor. Even within a single session, references become less reliable as the system's context window drifts (Yue Xing, et al. 1). When prompted "help me generate some new breakfast recipes. Don't want another breakfast-in-bed fiasco", Gemini enthusiastically responded "I can definitely help with that! Are you looking for something quick and easy or something more elaborate?" The system treated my comment like any other generic cooking request, rather than building on shared context and established jokes within our relationship.

VI. Interpreting Voice AI's Interactional Success

Experimentation and literature analysis reveals that while AI systems may fail at relational humor that constructs a unique sense of "we-ness", it generally succeeds at employing interactionist humor, a process that maintains the continuity and energy of a conversation.

Voice AI succeeds at what sociologist Erving Goffman calls "Face-Work", or the shared labor to maintain a natural, smooth interaction (Goffman 7). Goffman's core theory relies on the dramaturgical perspective, a theory of social interaction that employs the metaphor of a theatrical performance. He argues that when people interact, they are like actors on a stage performing for an audience, trying to construct a specific "definition of the situation" or a desired impression of what is happening and how they are perceived (Goffman 4). In order for social interaction to function smoothly, participants need to reach a mutual understanding of the situation at hand so they can behave accordingly (Goffman 5). This behavior defines a "front" or the visual, spoken, and unintentional cues given off during a performance (Goffman 13). Humor fits into Goffmanian interactionist theory in multiple ways. First, how inconsistencies in these performances can lead to humorous situations (Goffman 33). Humans spend a lot of time maintaining our presentations, so when one departs from their expected performance or participants have mismatching definitions of the situation, humorous moments often arise. A Voice AI generally avoids these mismatching performances fairly well since it is engineered to maintain a consistent front. Breaking character would undermine user trust and interaction coherence.

Secondly, humor can serve as an interactional lubricant that helps maintain performances and smooth interactions (Goffman 5). This is where current Voice AI systems seem to perform especially well. Experimenting with Gemini Live, for example, reveals how Voice AI is the ultimate polite conversationalist. It appears to simulate symbolic interaction by utilizing phatic laughter, polite chuckles, soft “hahas”, or amused tones to grease the wheels of conversation much like humans do. They may not actually understand the symbol behind a joke, but they understand the interactional cue surprisingly well. When a user makes a joke, AI is able to pick up on the social cue and simulate a laugh in order to validate the user’s face-work. It laughs not because it finds the joke funny, but because it has been trained to realize that an amused response is necessary to maintain a smooth interaction and validate the user’s own performance. To reveal the significance of this capability, imagine, briefly, the opposite. Imagine a conversation in which one’s counterpart doesn’t laugh at any jokes, or perhaps laughs too enthusiastically at a passing remark, or only responds with a chuckle to a joke that was intended to elicit a bigger reaction. These mismatches would quickly erode the flow of interaction, making conversations feel awkward. Instead, laughter (even simulated by AI) helps serve as subtle glue that helps performances continue seamlessly even in the absence of genuine shared symbolic understanding.

This marks the first major departure of this paper: research at this intersection wrongly overvalues joke-telling as the benchmark for humorous technology. What is far more useful socially is conversational AI’s current capacity to deploy humor as a strategic resource with impressive timing that upholds a mutual front and smooth encounter. For the aforementioned companies competing to build the best AI Voice agents, this deployment of interactional humor is a far more useful feature to build because it ultimately helps Voice AI feel competent, trust-worthy, and human-like to users. Companies don’t need their AI agents to be comedians, they want them to be human-like.

VII. The Future of Voice AI - From Information to Affect Retrieval

Voice AI systems are great cocktail party guests: they are polite, reactive, and socially alert. By most accounts, they succeed at the interactionist level of humor. But, the future of AI Voice agents must move beyond simply performing face-work and instead optimize for long-term relational humor to feel natural and human. In fact, Kuipers even argues that polite humor can actually be a barrier to a relationship, making interactions feel civilized yet distant (Kuipers Chapter 4, 60). To bridge this gap, and design truly humorous Voice AI agents, researchers must build an Affective Retrieval system — a tool that allows AI systems to recall emotion, tones, and inside jokes, ultimately achieving a closer representation of idiocultural humor which current systems seem unable to perform.

This goal of building an agent that can construct a joking-culture with its human counterpart is not a theoretical exercise to align AI systems with the sociological definition of humor, but rather a direct need based on the trajectory of the Voice AI industry. Andreesen Horowitz (a16z), a top tech venture capital firm, imagines a future in which Voice AI systems will be an “always-available companion or coach” (a16z). Startups have already begun building hyper-personalized AI Voice agents that represent the “top” employee of a company and serve as a customer's personalized success representative (WebProNews). The industry is actively pivoting from a one-off call center model, to long-term relational AI systems. While the technology is advancing, the sociological architecture for this transition is underexplored.

Current AI systems utilize Retrieval-Augmented Generation (RAG) technology to fetch facts about a user: their name, last order, or recent calendar appointment. This is necessary for “one-off” AI agents that are powering call centers, for example. But, to build “always-available” voice companions, the industry will need to invest in what I call Affective Retrieval, requiring an

AI system to encode and store not just the content of the interaction, but the emotional resonance of it as well. Doing so allows Voice AI to create a joking culture with its counterpart, one that is inherently referential and embedded to push away from generic, overly-polite humor. This system would allow AI agents to recall and understand references from previous conversations and patterns, not just to claim “I remember this conversation” but "I remember how this conversation felt”. In relationships, humans don't want to just be remembered, but we want to be known; this is the difference between data retrieval and affective memory and thus the difference between an omniscient stranger and a true AI companion. Ultimately, this paper is not intended to propose how to build an Affective Retrieval system, but instead articulate its importance in future conversational AI technology while also exploring sociological stakes.

VIII. Implications of Long-Term Conversational AI Companionship

What happens to humor, and social life, if conversational agents begin participating in long-term joking cultures with humans? Via a Goffmanian lens, this completely redefines who is and isn’t a social actor, introducing machines into the social dramaturgy. But as Katz helps explain, this performance is merely simulated, and is purely a calculated “doing” of laughter and humor. Sociologist Arlie Hochschild’s work helps us explore what happens when the “doing” of humor is outsourced to machines. Humor is not neutral, but collaborative and meaningful, often putting others at ease, defining group boundaries, or serving as conversational glue. Humor is what Hochschild calls “emotional labor”, the management of one’s feelings to influence the appropriate emotional response around them. (Hochschild 7). And automating this labor in long-term relationships could produce companionship that feels emotionally rich but is actually structurally hollow when one participant never needs care, rest, or reciprocity (Hochschild 110). The automation of humor transforms it from a collaborative process to a service (Hochschild

186). Notably, humans are increasingly accepting of simulated relationships, not because they are necessarily more fulfilling, but because they are easier, argues Sherry Turkle (Turkle 1, 13). Thus, conversational agents create an opportunity for the feeling of shared history and inside jokes without risking misunderstanding, rejection, or crossed boundaries (Turkle 66). Future AI systems will likely retain their cocktail party mannerisms and resultantly never misread a joke poorly or withdraw affection. Humor as emotional labor becomes incredibly safe, but asymmetrical. It’s easy to imagine a future in which humans actually prefer this kind of relationship where humor never threatens their relational fabric (Turkle 12). Not only does humor become low-friction emotional labor, it essentially becomes a feedback loop. Pierre Bourdieu mirrors Kuiper's argument, introducing the idea that humor is cultural capital, an embodied knowledge that helps distinguish norms and social position (Bourdieu 6). As a user and conversational agent build an idioculture, the resulting joking-culture would simply be a mirror of the human’s habitus or engrained dispositions or unconscious guiding behaviors (Bourdieu 170). A joking-culture is no longer a space where distinct habituses clash and boundaries are defined, but a space for affirmation. Thus, the most pressing sociological question is whether humor retains its social function of meaning-making when one participant is incapable of real symbolic interpretation, risk-taking, or emotional exhaustion; this paper argues no. It’s most likely that the future of Voice AI companions may simulate the same constructs through humor, but in reality, will not replicate the actual work humor performs. Jokes may land perfectly, inside-jokes may be referenced flawlessly, and timing may feel spot-on, but humor is redefined. It prioritizes comfort and ease and loses its ability for genuine co-creation of meaning.

IX. Conclusion

Developing companion Voice AI agents will pose entirely new challenges researchers will need to get right to win the trust of consumers. One of those challenges is how to give an AI system a sense of humor that feels deeply human, referential, embedded, and personalized. While the current generation succeeds at interactionist humor, face-work that maintains a smooth conversation, this won't be enough for the future of AI systems. Creating idiocultural humor via Affective Retrieval is important because, without the tools to encode and reference shared emotional history, Voice AI agents remain surface-level cocktail party dates, lacking the relational richness that characterizes human humor. Notably, the successful creation of these conversational companions could cause deep changes for humor. A joking-culture with a Voice AI agent would simulate a meaningful relationship, but ignore the actual work humor constructs. The sociological risk is not that humor in human-AI relationships is distorted, but that humans enjoy these parasocial relationships so much that we forget that humor was meant to be labor-intensive, collaborative, or negotiated in the first place. If humor no longer requires emotional labor, and the vulnerability, risk, and effort that comes with it, humor may still feel human, but will no longer perform boundary-defining and reality-building work, transforming how humans perceive, perform, and value humor in all social life.