ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

TRADUZIONE, INTERPRETAZIONE E INTERCULTURALITA'

Ciclo 37

Settore Concorsuale: 10/L1 - LINGUE, LETTERATURE E CULTURE INGLESE E ANGLO - AMERICANA

Settore Scientifico Disciplinare: L-LIN/12 - LINGUA E TRADUZIONE - LINGUA INGLESE

PRAGMATICALLY SPEAKING: HUMOR IN INTERACTIONS WITH AI

Presentata da: Jennifer Monroe

Coordinatore Dottorato

Chiara Elefante

Supervisore

Delia Carmela Chiaro

Esame finale anno 2026

Dedicated to

JDM who brought Alexa home,

MGM who invited her to stay, and

VP who talks to her like his pet.

Table of Contents

Acknowledgements

I am deeply grateful to the many people who supported me throughout this journey, particularly those who patiently endured my endless musings on how machines talk and what we say to them—even after they had heard it all before. To my friends, family, and colleagues who nodded along as I excitedly analyzed yet another aspect of human-AI conversation or dissected the linguistic nuances of Alexa speaking Italian using English phonetics: thank you for your patience, your encouragement, and for occasionally pretending my observations were as fascinating as I found them.

Special recognition goes to my supervisor, Delia Chiaro, whose vast experience in humor research surely proved essential in surviving me, her last official PhD student. That she chose to follow this project even after retirement speaks either to her extraordinary dedication or to my persuasive abilities—I suspect the former. Her guidance, insight, and willingness to engage with my ideas (repeatedly) have been invaluable, and I am simply honored that she was willing to work with me.

To everyone who contributed to this work through conversations, questions, laughter, and support: this PhD thesis exists because you were willing to listen, engage, and humor me—in every sense of the word.

Abstract

Humor in human-AI interactions is investigated through a metalinguistic-pragmatic lens, focusing on how users engage with AI voice assistants like Alexa and ChatGPT. Prior to LLM-based systems, humor in human-AI interaction research focused primarily on evaluating machine-generated humor, leaving gaps in understanding how humans initiate and respond to humor with conversational agents. A systematic literature review of six studies (published 2016-2024) examining voice assistants (Alexa, Siri, Google Assistant, Cortana) reveals that users employ humor to explore systems and test relationships rather than for entertainment. The most frequent category of humor is personality questions.

Addressing gaps in knowledge, this work presents a multimodal, multilingual dataset from 2016 to 2025 comprising three complementary data sets: (1) short-form videos from YouTube and TikTok featuring families interacting with AI voice assistants in English, Italian, and Spanish; (2) video recordings of interactions with Alexa in English and Italian; and (3) posts from Facebook, Instagram, and TikTok that explicitly signal humorous intent through linguistic and paralinguistic markers (English and Italian).

Three distinct patterns emerge from Study 1: English videos positioned elderly users as objects of amusement, with the person frequently responsible for communication failure (39%); Italian videos centered on language barriers rooted in dialect use, with fewer instances of human-attributed failure (23%); while Spanish videos celebrated successful connection with zero instances of human-attributed failure.

Study 2 reveals the fundamental asymmetry: people engage socially while AI generates language. Participants' metalinguistic commentary tests AI theory of mind, marks embodied/cultural knowledge boundaries, and constructs meaning from incoherent responses, revealing assumptions and common ground strategies with non-human interlocutors.

Study 3 analyzes 99 humorous social media posts targeting AI communication failures (comprehension, turn-taking) and successes (ChatGPT's excessive perfection). Italian posts uniquely imagine dialect-speaking AI, performing regional identity. Humor functions as a metacommunicative resource for negotiating boundaries with systems simulating social engagement.

Chapter 1 Introduction

This doctoral thesis examines how people communicate with artificial intelligence (AI), focusing on the humorous experiences that arise spontaneously as human social expectations meet

conversational AI. When talking machines began to appear in daily life, for example the cars in the 1980s that announced “lights are on” or “door is open” in place of the standard ding, the voices were easily recognized as pre-recorded (Martin, 2015) human voices repeating a unidirectional message. The spoken words were simple linguistic signs that the driver could accept or ignore and there was no pretense of shared meaning-making or conversation.

The expectations for interactive communication with technology changed dramatically with the public release of ChatGPT (Open AI) in November 2022. Unlike earlier systems based on pre-recorded or pre-programmed scripts, conversational AI like ChatGPT can understand, process, and respond to people by generating contextually appropriate replies that simulate human conversational patterns—complete with apparent comprehension of humor, nuance, and social cues. We can finally speak (or type) to our artificial assistants using natural dialogue; it is no longer necessary to learn specific commands or syntax to interact with these systems, as with earlier versions of voice-based assistants.

This represents a fundamental social change: the language that evolved to facilitate human interaction, create shared knowledge, and establish relationships now enables people to form seemingly similar bonds with artificial systems. In other words, the conversational nature of modern AI tools invites social engagement interwoven with task negotiation, thereby creating pragmatic tensions between our expectations for how conversation should unfold and the system’s actual capabilities (Luger and Sellen, 2016). People are engaging socially while AI is generating language, an inherent asymmetry that can disrupt the illusion of communication. When people encounter unexpected or irrelevant responses, silence, or unusual acoustic signals, the conversational flow is interrupted, evoking reactions that range from laughter and meta-commentary about the communication to confusion and resigned frustration. These moments of pragmatic breakdown expose the underlying pragmatic assumptions and cultural expectations we bring to conversation, creating fertile ground for irony, sarcasm, and teasing. As AI becomes increasingly embedded in daily activities, understanding how people navigate this asymmetry has implications not only for interface design but also for shaping user expectations, establishing appropriate interaction norms, and ensuring the effective social integration of conversational technologies.

Research that examines natural language interactions between humans and machines has emerged across multiple disciplines, adopting a wide range of analytical approaches. The results of these studies can be roughly separated into two complementary perspectives: 1) understanding how people engage with communication technologies and 2) designing systems that optimize user experience and satisfaction. Much of the work in the first perspective is guided by the “Computers Are Social Actors” or CASA paradigm (Nass et al., 1993; Reeves and Nass, 1996; Nass and Moon,

2000) which shows how people treat computers as they would treat another person, meaning the social scripts from human-human communication are also relevant to interactions with social technologies. Research continues to evolve in response to the increasing technological sophistication in society: Gambino et al. (2020) discuss how to extend the CASA framework, Heyselaar (2023) finds evidence that the CASA effects don't hold with older technology, Johnson and Gardner (2007) suggest that more experienced users are particularly likely to treat the computer as a social actor, Morkes et al. (1999) show that people's social responses towards computers fall on a continuum, Hall and Henningson (2008) find that a person's demographic characteristics play a role in the interaction, but according to Perkins Booker et al. (2024) these characteristics are not related to the amount of laughter in interactions, and Xu et al. (2022) adopt a novel methodology to tease apart the psychological mechanism underpinning people's social responses to technology. These scholarly insights on the evolving landscape of human interaction with technology provide a foundation for understanding how people engage with conversational technology and establish a framework for examining contemporary systems.

Conversational AI systems exist today because of developments in computational linguistics, particularly the advances made in processing and generating language-based interactions that are increasingly natural, human-like and contextually aware. Driven by industry and scholarly interest in creating more engaging interactions with AI-powered systems, decades of this research has concentrated on humor. Journalists often highlight the advanced humor capabilities of modern AI to emphasize that these tools are easy to use and technologically sophisticated. Headlines declare that AI can understand context and humor (Marr, 2023; Nover, 2024), possesses a "devastating sense of humor" (Manjoo, 2022), and now has "a better understanding of sarcasm" (Morrison, 2024). Additionally, the AI assistant Grok, is "designed to answer any question with a touch of wit and humor" (X, 2025). Beyond the media hype about AI's human-like abilities, scholarly research has approached computational humor as both a technical challenge and an opportunity to advance humor theory through empirical analysis (Ritchie, 2009, p. 79), by formalizing the key components of humor (Hemplemann, 2008, p.333) and specifically working to untangle humorous and non-humorous texts (Nijholt, 2006, p. 64).

Despite extensive research on AI's computational humor capabilities, we are still in the early stages of understanding the human side of these interactions. The development of LLM-based conversational AI systems with sophisticated linguistic capabilities and contextual awareness marks a significant advancement. These systems create new opportunities to investigate spontaneous humor in natural human-AI dialogue beyond what was possible with earlier systems. Studies based on previous-generation voice assistants have provided initial classifications of humorous utterances

directed toward AI assistants (Lopatovska, 2020; Lopatovska et al., 2020), taxonomies of playful requests (Shani et al., 2021), and document deliberately humorous interactions including irrational shopping queries (Shapira et al., 2023) and playful questions aimed at testing the systems or amusing family (Sciuto et al., 2018). Studies also confirm that humor is a desirable quality in conversational technology (Volkel et al., 2021) and serves as a coping mechanism when interactions fail to meet expectations (Martinovsky and Traum, 2006).

However, critical gaps remain as existing research examines intentional humor and playful testing of system limits, but not the spontaneous, unscripted moments when humor emerges organically from the interaction itself—the unplanned moments of amusement that arise when human expectations collide with AI limitations. Unlike programmed jokes or deliberate playfulness, spontaneous humor emerges from situational ironies, serendipitous misunderstandings, and the gap between human conversational norms and AI system constraints—precisely the moments where the asymmetry between social engagement and language generation becomes most visible. We know little about the linguistic strategies people deploy in humorous AI interactions, how communicative patterns emerge spontaneously, or what playful exchanges reveal about human-machine communication. Equally unexplored is how people collectively make sense of these experiences, particularly through social media, where the communicative challenges of human-AI interaction are transformed into shareable content, offering a window into broader cultural attitudes about conversing with machines.

This study addresses these gaps by examining the human side of spoken AI interactions—specifically, the linguistic and pragmatic strategies people employ when humor emerges spontaneously.

Research questions:

1. What forms of linguistic playfulness emerge spontaneously during spoken interactions with AI assistants?
    - 1.1 How are these linguistically cued?
    - 1.2 Who/what is targeted in humorous moments, as evidenced by verbal commentary and participant laughter?
    - 1.3 Do patterns differ across English, Italian, and Spanish speakers?
2. What aspects of human-AI communication are targeted in humorous social media posts in Italian and English?

To address these research questions, this work employs three complementary studies, each analyzing a distinct dataset of human-AI interactions.

Study 1 addresses RQ 1.2 and contributes to RQ 1.3 through thematic analysis of 117 videos of spontaneous interactions with AI assistants from YouTube and TikTok (2016-2023) in English, Italian, and Spanish. This study examines who participates in laughter and when it occurs sequentially in the interaction to reveal who or what becomes the target of humor—whether these moments reflect shared amusement at the interaction itself or mockery of the user's failure.

Study 2 addresses RQ 1.1 and RQ 1.3 through pragmatic analysis of 40 video-recorded interactions with Alexa in English and Italian (2023-2024). Drawing on Bateson's (1972; 1987) concept of meta-communication signaling play and Chafe's (2007) concept of "non-seriousness," this study examines the language people use in spontaneous playful utterances directed at AI, with particular attention to meta-comments that reflect on AI's communicative abilities or the ease/difficulty of the interaction itself.

Study 3 addresses RQ 2 through metalinguistic analysis of 161 social media posts from Facebook, Instagram, and YouTube (2023-2025) that humorously depict human-AI communication. This analysis examines what aspects of AI interaction become sites of humorous commentary, revealing broader cultural attitudes about conversing with machines and how the human-AI relationship is evolving.

Together, these three studies capture humor both within live interactions (Studies 1 and 2) and in reflective social discourse about those interactions (Study 3), providing a multi-faceted view of how humor functions in human-AI communication.

This work contributes to the literature on humor, human-computer interaction and human-machine communication, by foregrounding the human perspective and documenting the specific linguistic and pragmatic strategies people use when interacting with AI. Through this multi-layered approach, the results demonstrate how humor serves as both a communicative resource and a lens for social interpretation of AI technology, advancing our understanding of language use with artificial communication partners.

Key terms and operational definitions

Given the rapid pace of change in AI technologies, establishing clear operational definitions is essential for meaningful discussion. Acknowledging the wide variety of architectures and underlying technology in AI systems, this thesis addresses technology in general terms, specifying device characteristics only when directly relevant to the analysis. This broad treatment reflects my analytical focus on the interactional and experiential aspects of human-AI communication. I also use human-related terminology (e.g., AI 'understands' or 'hears') as convenient descriptors of

system behavior, with the understanding that such language refers to computational processes rather than implying consciousness or sentience. Questions of whether AI possesses genuine understanding or whether its behavior in theory of mind tasks is comparable to humans remain beyond the scope of this work.

Theory of Mind (ToM) refers to the cognitive capacity to attribute mental states—beliefs, intentions, desires, emotions, and knowledge—to oneself and others, and to understand that others may hold perspectives different from one's own (Premack & Woodruff, 1978). This ability to infer what others think, feel, or know is a valuable skill in social interactions, allowing people to align with others by anticipating their behavior (Lenaerts et al., 2024) or make plans that lead others to change their course of action (Ho et al., 2022). Just as people pay attention to what other people think, they also attempt to predict the mental states of AI "minds," expecting conversational technology to comprehend shared context, remember past interactions, and demonstrate consistency in knowledge. For this research, ToM is relevant when examining how users linguistically signal playful intent (implicitly assuming AI recognizes such signals) and when analyzing meta-comments about AI's communicative abilities (which often implicitly attribute mental capacities to the system).

Conversational technology refers broadly to any system capable of sustaining dialogue with users, regardless of its underlying sophistication or modality. **Conversational AI** is a more specific term within this category, emphasizing systems designed to engage in natural language interactions using artificial intelligence, including natural language processing, machine learning, neural networks, and large language models (LLMs), to process and generate human-like responses. In very basic terms, **LLMs** are AI systems that have been trained on an extensive amount of text data so they can understand and generate human language. They learn the statistical patterns and relationships between words, which allows them to predict the probable words in sequences. For our purposes, **conversational agents** will also be used synonymously with conversational technology and conversational AI, especially when discussing prior research that specifically employs the term conversational agents.

Voice-based technology refers to any system where voice serves as the primary interaction channel, allowing users to speak their requests and receive auditory responses. Text-based technology, conversely, denotes systems where written language serves as the primary interface, with users typing queries and receiving written replies. Throughout the discussion and analysis, I use **AI**

Assistants to refer to voice assistants[1] or voice-based systems (e.g., Alexa and Siri) and **LLM-based systems** to refer to contemporary text-based conversational AI (e.g., ChatGPT or Claude). An important distinction exists between **previous-generation** and **current-generation** systems. Previous-generation voice assistants, such as early iterations of Alexa and Siri, operated primarily through command-based interactions with limited contextual awareness. Current-generation systems, such as contemporary LLM-based assistants, demonstrate substantially enhanced conversational capabilities, contextual memory, and natural language understanding. AI voice assistants like Alexa and Siri have evolved over time and now incorporate some current-generation features and thus may be referenced in discussions of either previous or current generations.

In his seminal chapter on the social functions of humor, Martineau (1972) states that humor as a medium of communication "assumes many forms and its social functions become complex under the influence of other social processes and existing social structures". This observation of humor as fundamentally social, taking on multiple forms and embedded in communication forms the basis of the working definition of humor in human-AI interactions in this thesis.

Humor is operationally defined as spontaneous comments during interactions with an AI system that are non-serious, amusing, or entertaining, following Chafe's (2007) concept of situations presenting "pseudo-plausibility" that need not be taken seriously and Raskin's (1985 as cited in Attardo, 2017) notion of non bona-fide communication. This definition emphasizes humor as an emergent property of the interaction itself rather than as pre-programmed content delivered by the AI system. It encompasses situational humor and conversational joking (Norrick, 2003; 2010), unintended consequences, serendipitous misunderstandings, and other moments of levity that arise naturally from the exchanges between people and AI. Such humor often arises from expectations (or unmet expectations), particularly when actual experiences conversing with technology clash with our internal models of how the world works (Forabosco, 2008). Critically, this definition excludes the intentionally designed witty responses or programmed jokes, focusing instead on how humor emerges from the social dynamic of human-AI communication.

Laughter is not exclusive to humor. It signals different emotions (Chafe 2007; Provine 1996) and serves as a tool allowing people to create, modify, and maintain social relationships. Laughter mediates social discomfort, signals affiliation, marks boundaries between serious and non-serious frames, and responds to incongruity or surprise (Glenn, 2003). In human-AI interaction, laughter may arise from genuine amusement at conversational turns, from awkwardness when interactions

[1] The relevant literature uses a variety of different terms based on the specific role or function of the technology: virtual personal assistants, intelligent personal assistants, digital assistants, voice assistants, virtual agents, conversational agents and smart speakers with an integrated voice assistant.

fail, from surprise at unexpected system responses, or as a social smoothing mechanism when conversational norms are violated (Perkins Booker et al., 2024). The sequential placement of laughter—who laughs when—provides insight into how participants interpret the ongoing interaction (Glenn, 2003). This research examines laughter as a pragmatic response revealing how users make sense of AI communication, whether through intentionally playful exchanges or inadvertently incongruous moments.

Play refers to a voluntary interactional frame marked by meta-communicative signals indicating the exchange should not be taken as seriously (Bateson, 1972; 1897). Play has distinct boundaries from "ordinary" activity and follows its own internal rules (Huizinga, 1938), suspending the usual serious meanings of communicative signs (Kozintsev, 2010). This framework provides the context for understanding humor as a form of playful communication in human-AI interactions.

Linguistic playfulness refers to language use that departs from serious communication, signaling a non-literal or non-serious frame. This includes forms such as irony, sarcasm, teasing, hyperbole, understatement, and rhetorical questions. While playfulness may generate humor, the two are not synonymous—playful language can indicate non-seriousness (Chafe, 2007) that may or may not result in amusement. Among these forms, teasing warrants particular attention as it serves multiple social functions, signaling group membership and shared identity, and can facilitate bonding, or exert social control (Boxer and Cortés-Conde, 1997; Norrick, 2010). Teasing emerges prominently in the dataset as a form of play directed toward AI systems.

Suspension of disbelief in human-AI interaction builds on Reeves and Nass's (1996) finding that people automatically apply human social scripts when communicating with computers—not through conscious choice, but as an automatic perceptual response. Following Lombard and Ditton (1997), the technological medium becomes functionally invisible, allowing people to engage with AI systems as conversational partners without actively deciding to do so. Drawing on Kozintsev's (2010) analysis of play frames and humor, this dissertation proposes that spontaneous humor emerges when this automatic suspension of disbelief breaks down—when people suddenly become conscious of treating AI as a social actor, when AI limitations make continued belief impossible, or when the "loss of meaning" (Eco, 1994 as cited in Kozintsev, 2010) in the interaction becomes apparent. This operational definition establishes the foundation for examining how moments of conscious recognition—the breaking of automatic belief—contextualize the spontaneous humor observed in human-AI interactions.

A cultural timeline of talking machines: Shaping our collective imagination

Having established the key terms and technological distinctions, it is important to situate human-AI communication within its broader cultural-historical context. The linguistic and pragmatic strategies

people employ when talking to machines—and the spontaneous humor that emerges from these encounters—are shaped by decades of exposure to talking machines, both real and imagined. Critically, audiences have long been exposed to machines that sound more intelligent than they actually are, creating expectations that exceed actual capabilities. The following timeline illustrates this legacy. Key moments from 1939 to 2025 illustrate when talking machines were seen, heard, and eventually used by audiences across the English-speaking world and beyond. This span of time was selected as someone alive in 2025 (at the time of writing) could have experienced all these events. Whether or not a single individual witnessed each event, these examples of human-machine communication are part of society's collective experiences that implicitly shape expectations on how to talk to machines.

Modern technology is often designed to imitate fictional technology, leveraging the expectations viewers form by watching on-screen portrayals of human-technology interaction. When such technology arrives in the real world, it can feel immediately familiar from years of watching these interactions unfold in fiction, blurring the boundary between actual capabilities and culturally formed expectations. Beyond shaping expectations, fictional technology also serves as a shared cultural reference point: Shedroff and Noessel (2012) point out how much easier it is to describe future technologies that have no real-world examples by mentioning a memorable moment from a film everyone has seen.

The examples in this timeline present a juxtaposition between real-world voice technology referenced in fictional narratives and cinematic techniques that realistically mimic state-of-the-art voice synthesis, which makes it difficult to distinguish between the actual and fictional capabilities of machines that talk. Critically, many of the "machine voices" in this timeline were pre-recorded human voices or human-controlled systems. For example, AT&T's 1939 electrical speech synthesis machine, which spoke with a true synthesized voice, was dependent on its female operators to add intonation into its phrases by pressing piano-like keys (MonoThyratron, 1939). These human voices masquerading as artificial intelligence have created what could be called a pragmatic illusion: prosodic competence, which is mistaken for communicative understanding.

1939 – The Voder and Elektro

The AT&T exhibit at the New York World's Fair demonstrated a groundbreaking electronic speech synthesis system—a true talking machine. The Voder was designed at Bell Labs specifically to showcase the processes underpinning technological advances and entertain the crowds (Simon, 2025). This machine could produce continuous speech, but its human operators performed all the language processing tasks of hearing and understanding the requests directed at the machine. Operating a series of keys and pedals, they composed the speech sounds necessary to reply. The

hallmark traits of a synthesized voice are easily recognizable in the slightly stilted delivery, but listening closely, the human operator's influence is immediately apparent in the prosody of the voice, consisting of rising and falling tones, strategic pauses, and emphatic stress. The tonal qualities of this voice from 1939 create a sharp contrast to the flat, monotonous delivery that we now typically associate with early mechanical voices.

Across the fairgrounds in the Westinghouse pavilion, Elektro the "Moto-Man" captivated crowds with witty banter and humorous quips. While audiences may have assumed they were witnessing machine-generated humor, Elektro's wisecracks were actually pre-recorded human speech activated by his human assistant through a sophisticated voice-controlled mechanism. The significance of this event lies in its scale: 45 million people witnessed a humanoid machine smoking cigarettes and responding to voice commands in natural language, in 1939.

1960 – "Daisy Bell (Bicycle Built for Two)"

Twenty-one years later at Bell Labs, Carol Lochbaum and John Kelly wrote a new speech synthesis software that enabled a computer to sing without any human assistance: "Daisy Bell (Bicycle Built for Two)". This technological achievement captured the imagination of science fiction writer Arthur C. Clarke, who heard the computer's performance during a visit to Bell Labs and later incorporated the concept into the screenplay he co-wrote with Stanley Kubrick for the 1968 film, *2001: A Space Odyssey*.

1968 – HAL in 2001: A Space Odyssey

The iconic final scenes of *2001: A Space Odyssey* (Stanley Kubrik), show astronaut Dave Bowman methodically deactivating the HAL 9000 computer by removing memory blocks from its AI core. As HAL pleads for his survival, expressing fear and asking Bowman to stop, his voice and language skills progressively deteriorate to a more primitive state. Drifting out of consciousness, HAL offers to sing the song his instructor taught him with his early programming. The homage to the Bell Labs research is clear when HAL says the song is called "Daisy" and begins singing, his voice becoming progressively lower and more distorted until it fades to silence. In the Italian localization this scene maintains the elements of regression towards childhood but does not reference early voice technology since HAL sings the nursery rhyme "Ring Around the Rosie" in Italian. In the English version, this scene establishes a direct connection between actual technological development and cinematic representations of technology.

Audiences witnessed this fictional moment of tense human-computer communication that perfectly captures both the potential and the limits of artificial intelligence. Kubrick gave HAL cognitive ability, language use, and contextual awareness that were imaginary in 1968, but realistic in 2025, while the raw human emotion HAL expressed, which at the time of writing has not yet become a

reality, still resonates as a cautionary tale, half a century later. Video and images from this sequence continue to appear in social media content (in English and Italian), often adapted to comment on contemporary human-AI relationships and AI communication skills, particularly focusing on the complexities that arise when AI assistants appear to display emotions and independent thoughts.

1971 – The Versificatore (Versifier)

The *Versificatore,* a talking futuristic desktop poetry machine, first appeared in a newspaper drama piece by Italian writer Primo Levi in 1960. The story was subsequently published in Levi's collection of science fiction satires and comedies, *Storie naturali*, in 1966, before appearing on an important Italian television series for young teens in 1971. In Levi's narrative, the machine is purchased by an overworked poet who cannot fulfill all the commissions he has accepted for slogans, eulogies, and poetry. Once the poet successfully configures the settings for each button controlling topic, linguistic register, meter, and era of composition, the machine recites (and transcribes) its compositions in a deep theatrical voice, leading the poet to conclude that his human competitors could not produce better work. Science fiction author Bruce Sterling recently described Levi's *Versificatore* as "a prophetic vision of Large Language Model Artificial Intelligence" (Sterling, 2025).

1978 – Speak & Spell

The Speak & Spell by Texas Instruments was designed to teach children to spell and pronounce commonly misspelled words (Woerner, 2001). Available in the US, UK, and Japan, this educational toy contained pioneering voice synthesis technology to say short words, letters (for spelling games), and give simple feedback[2] like "You are right", "That is incorrect" or "Wrong, try again". The Italian version, "Grillo Parlante", arrived in 1979. Interestingly, the English name underscores function, the device speaks and the child spells, but its name in Italian the "Talking Cricket" (Bagnasco, 2023) references the character created to serve as Pinocchio's conscience in the 1883 children's book by Italian writer Carlo Collodi.

While the machine spoke to children using state of the art voice synthesis, children had to select a button key to respond to the toy. Its distinctive voice played a memorable role in Steven Spielberg's 1982 film *E.T. the Extra-Terrestrial,* both teaching the alien to speak and as a crucial component the alien used to build a device allowing him to phone home. This type of product introduced an entire generation of children to the sound of synthesized voices, a first step in normalizing the concept of true talking technology. This is an example of a hand-held electronic device which

[2] *Speak & Spell (US, 1979 version): Texas Instruments:* Free Borrow & Streaming: Internet Archive. (1979). Internet Archive. Retrieved August 1, 2025, https://archive.org/details/hh_snspell

encouraged an interactive relationship with technology positioning people as the learners and machines as authoritative sources of knowledge.

1983 – WOPR in WarGames

The John Badham's film WarGames explicitly highlighted the speech ability of the advanced military computer, WOPR (War Operation Plan Response). In the scene introducing the audience to the computer, the wiz kid turns on the speakers, asks if we want to hear the computer talk, and then continues by explaining how a program translates signals into speech sounds. The computer's distinctive clipped voice convinced millions of moviegoers that it was created with cutting edge voice synthesis technology. However, like many of the earlier examples in this timeline, the "computer voice" was actually a human performance. The Director's cut DVD includes a special feature commentary with, Director, John Badham stating that the actor John Wood recited the sentences in reverse word order, the audio tracks were then re-assembled and run through audio processing equipment to create the illusion of synthetic speech (Rhythmcomposer, 2017). This film demonstrates how fictional narratives can shape public perceptions by convincingly simulating contemporary technological capabilities with conversational functionalities that exceed real-world implementations. The computer was portrayed as the authoritative repository of specialized knowledge with high level reasoning skills and sophisticated conversational skills, including the ability to express emotions and ask rhetorical questions. The computer's relentless pursuit of nuclear war, driven solely by programmatic adherence to game protocols and the absence of proper termination codes, illustrates a critical gap between surface-level conversational competence and genuine understanding of context and consequence.

1984 – The Macintosh computer introduces Steve Jobs

At the annual shareholders meeting, two days after Apple's award-winning dystopian ad directed by Ridly Scott caught the attention of audiences watching the Superbowl, Steve Jobs unveiled the new Macintosh computer. Jobs showcased an impressive array of capabilities—drawing pictures, communicating with other computers, and playing chess (Bunnell, 2010; Loyola, 2024)—before allowing the computer to speak for itself. Its metallic male voice clipped the ends of words with an unusual intonation reminiscent of the computer in WarGames, and claimed to be aware of its physical surroundings, expressing relief at no longer being closed up in the bag.

> "Hello, I'm Macintosh. It sure is great to get out of that bag. Unaccustomed as I am to public speaking, I'd like to share with you a maxim I thought of the first time I met an IBM mainframe: NEVER TRUST A COMPUTER YOU CAN'T LIFT. Obviously, I can talk but right now I'd like to

> sit back and listen. So, it is with considerable pride that I introduce a man who's been like a father to me… STEVE JOBS." (BytesofApple, 1984)

The computer spoke in the first person, introduced itself, introduced Steve Jobs as its father figure and mocked IBM, commenting on the size and weight of their computers at that time. The audience could see each paragraph of text on the computer screen just before the speech generating program read the words aloud. This carefully choreographed performance created the illusion of spontaneous machine consciousness and conversational ability, subtly establishing expectations that computers could engage in natural dialogue with personality and humor. Despite the scripted nature of the interaction, the voice synthesis was genuine state of the art technology.

2011 – Watson wins Jeopardy!

IBM showcased Watson, its sophisticated natural language AI assistant, in an exhibition match against two reigning champions on the American quiz show *Jeopardy!* Onstage, Watson was represented by a dynamic globe-shaped (almost face-like) logo on a computer monitor, positioned between the two human contestants. While the audience heard Watson speaking to the host, and flawlessly participating in gameplay, there was a technical illusion at play. Similar to the Speak & Spell from thirty years prior, Watson could calculate replies and speak but could not actually hear the questions. When the clues became visible to the contestants, they were transmitted as text to the computer backstage (Siegel, 2015). Although Watson made some mistakes, for example replying with the Canadian city of "Toronto" for the category of US Cities, the AI contestant ultimately defeated the human contestants by winning the final round in the category of 19th century novelists. This televised event showed a widespread audience an AI system that could quickly retrieve complex information and deliver answers through voice synthesis, demonstrating the potential for machines to serve as expert keepers of human knowledge.

2011 – Siri speaks

In the same year that Watson won Jeopardy, Apple unveiled Siri for the iPhone 4S, making basic conversational AI available to users in English, French, and German. Within a year, Siri expanded to Japanese, Italian, Spanish, and Chinese.[3] Voice assistants included in smartphones became the first widely adopted technology that encouraged continual interaction with AI exhibiting overt social cues (Guzman, 2019, p. 343).

[3] Wikipedia contributors. (2025). *Siri*. Wikipedia. https://en.wikipedia.org/wiki/Siri Retrieved October 10, 2025.

2014 – Alexa comes home

Amazon introduced their voice assistant, initially only in the US, bundled inside smart speakers with integrated Alexa Voice Services. Unlike other voice assistants that were confined to personal smartphones, Alexa was the first to occupy the domestic space, enabling hands-free interaction from anywhere in the room. The platform expanded to Spanish-speaking countries in Central America by 2017 and reached Italy, Spain, and Mexico in October 2018 (Holt, 2018). The design vision for Alexa is explicitly linked to on-screen science fiction and inspired by two different fictional technologies. The original creator of the core speech technology for Alexa, Łukasz Osowski, says his vision of a talking computer was straight from science fiction movies like *2001: A Space Odyssey* (Redzisz, 2020), while the head of Amazon, Jeff Bezos, explained that concept of Alexa was fashioned after "the patient, know-it-all computer on the Starship Enterprise" (Boyle, 2016) from the American television series Star Trek that originally aired in 1966.

2022 - 2025 – ChatGPT ushers in tomorrow

The AI landscape shifted dramatically with the public release of ChatGPT,[4] marking the first widely accessible multilingual AI assistant[5] powered by large language models. The addition of a voice interface in 2023[6] (initially available only to paid subscribers) allows anyone to engage in sustained conversations with ChatGPT. This type of fluid spoken dialogue finally begins to resemble the voice technology depicted in films and OpenAI says that it is still improving in their announcement of a new speech-to-speech model.[7] Until these new models are implemented, a talking machine is essentially reading a written text aloud. Much like Levi's *Versificatore* that could create, transcribe, and speak ideas, everything that real world technology "hears" is first transcribed into written words, then processed and the response is "spoken" back to users. After more than eight decades, conversational technology has finally caught up to our expectations, realizing the ambitious vision originally presented at the New York World's Fair—truly building the "World of Tomorrow".

The pragmatic illusion: Machines sounding smarter than they are

The collection of robots, computers and AI assistants presented above narrates memorable moments of machines learning to talk and people learning to communicate with them, revealing a clear pattern: for most of the 86 years covered in this timeline, machines sounded far more intelligent

[4] Many other companies have since released similar AI assistants based on transformer architecture. Wikipedia contributors. (2026, February 2). *Generative artificial intelligence*. Wikipedia. https://en.wikipedia.org/wiki/Generative_artificial_intelligence Retrieved February 2, 2026.

[5] I am referring to ChatGPT as an AI assistant to highlight its social role as an assistant. ChatGPT is sometimes called a chatbot, but this term generally evokes text-based interaction or associations with simpler, rule-based systems.

[6] *Voice Mode FAQ*. (n.d.). OpenAI. Retrieved August 9, 2025, from https://help.openai.com/en/articles/8400625-voice-mode-faq

[7] *Presentazione di GPT-Realtime e aggiornamenti dell'API Realtime*. (2025, August 28). OpenAI. Retrieved August 30, 2025, from https://openai.com/it-IT/index/introducing-gpt-realtime/

than they truly were. Computer speech carries social cues within the language the same way human speech carries information about the internal nature of the speaker (Reeves and Nass, 1996; Shechtman & Horowitz, 2003). If a computer speaks like an intelligent, sentient being, people instinctively expect it to think and act accordingly. From the Voder's human-controlled prosody in 1939 to the Macintosh's scripted personality in 1984, audiences heard state-of-the-art voices with recognizably machine-like acoustic features paired with strategic pauses, conversational rhythm and emotional language—all the pragmatic features that signal genuine understanding in human conversation—though neither machine possessed any intellectual autonomy. Iconic representations like HAL in *2001: A Space Odyssey* intensify this illusion by showing audiences the pragmatic endpoint that voice synthesis would eventually reach. HAL was a machine that could think autonomously, hear others in the room, respond with contextual awareness, and speak with emotional depth, in real time conversations. Generations of people absorbed these scenes, passively experiencing sophisticated human-computer interactions while technology in the real world continued advancing to reach its on-screen representations. The Speak and Spell (1978) gave interactive spoken feedback but could not think or hear. Watson (2011) demonstrated a wealth of knowledge and could speak but could not hear; Siri (2011) could think, speak and hear but had limited knowledge. Only now with the LLM based systems, such as ChatGPT, that have voice interaction modes, has technology finally caught up to what Clarke and Kubrick imagined in 1968: machines that genuinely listen, process language with sophisticated reasoning, and respond through synthesized speech—delivering the first step toward the linguistic and pragmatic competence that audiences have been culturally primed to expect for more than half a century.

Organization of the thesis

This work is organized into seven chapters. Chapter 1 has provided the theoretical and cultural foundations for this investigation, situating spontaneous interactional humor within the broader context of human-AI communication, defining key terminology, and presenting a cultural-historical timeline that reveals how decades of exposure to talking machines have shaped communicative expectations for conversational technology.

Chapter 2 presents a systematic literature review of humor in spoken interactions with conversational technology. The review consolidates existing evidence on the linguistic aspects of humor in human-AI interaction, focusing specifically on the utterances people make rather than machine-generated responses. As the findings reveal, existing studies primarily capture intentional humor attempts, with spontaneous, emergent humor remaining largely unexamined. The chapter concludes by identifying linguistic patterns, gaps in understanding, and the contributions and limitations of current research.

Chapter 3 details the methodological approach employed across the three studies conducted, describing the data collection procedures, analytical frameworks, and rationale for the concatenated multi-study design used to investigate interactional humor in human-AI interactions.
Chapters 4, 5, and 6 present the results and discussion for each of the three studies respectively. Chapter 4 examines the sequencing of laughter in social media videos of socially situated interactions with AI assistants. Chapter 5 analyzes the pragmatic features of emergent playful language directed at AI assistants. Chapter 6 investigates how human-AI communication becomes the target of humorous social media commentary.
Chapter 7 synthesizes findings across the three studies, examining how they collectively illustrate aspects of spontaneous interactional humor in human-AI interaction. This concluding chapter returns to the central tension identified in Chapter 1, the asymmetry between human social engagement and AI language generation. Discussion on the emergent patterns of linguistic playfulness, laughter, and humorous commentary reveals how people communicate with artificial systems in both English and Italian. The chapter articulates the contributions of this work to humor studies, pragmatics, and human-computer interaction, acknowledges limitations of the current research, and proposes directions for future investigation into the evolving communicative relationship between humans and machines.

Chapter Summary

This chapter has laid the theoretical, cultural, and methodological foundations for investigating spontaneous humor in human-AI interactions. While extensive research examines the different facets of AI's computational humor capabilities, the human perspective of humorous interactions with AI remains largely unexplored. To address this gap, this dissertation investigates the language and social signals people use when navigating humorous moments with conversational AI through three complementary studies examining: sequential organization of laughter in socially situated interactions (RQ 1.2, RQ 1.3), emergence of playful language directed at AI (RQ 1.1, RQ 1.3), and aspects of human-AI communication targeted in humorous social media critique (RQ 2).
This chapter traced the evolution of talking machines from simple pre-recorded announcements to sophisticated conversational AI systems, revealing an inherent asymmetry in human-AI communication: people engage socially with technology while AI generates language, creating expectations about the nature and functionality of AI that are not aligned with AI's actual capabilities. The cultural timeline illustrated how nearly a century of exposure to talking machines—both real and imagined—has the potential to shape contemporary expectations with conversational technology. This illustrates a persistent pragmatic illusion: machines have long sounded more intelligent than they actually were. From the human-controlled prosody of the 1939

Voder to the scripted personality of the 1984 Macintosh computer, audiences have repeatedly encountered talking machines that appeared to possess nuanced human-like communicative competency. Current-generation LLM-based AI systems are only now fulfilling these long-standing expectations. This historical context helps explain the implicit conversational expectations people bring to conversations with AI and frames the investigation of the spontaneous humor emerging from the gap between expectations and experiences.

The next chapter presents a systematic review of existing research on humor in spoken human-AI interactions, providing empirical grounds for the three studies detailed in subsequent chapters.

Chapter 2
A systematic review of the literature

1. Overview

Chapter 2 presents a systematic review of the literature on humor in spoken interactions with conversational technology. The chapter begins by establishing the lines of scholarship that examine the evolving social relationship with technology. It then details the systematic search strategy, inclusion and exclusion criteria, screening procedures, and quality assessment methods. The findings section analyzes existing research on language used in human-initiated humor with voice-based AI systems. The chapter concludes by synthesizing key patterns, identifying gaps in current understanding—particularly the absence of research on spontaneous, emergent humor—and discussing the contributions and limitations of existing studies, thereby establishing the empirical foundation for the three studies presented in subsequent chapters.

A systematic literature review is a research methodology that provides a reproducible, comprehensive, and structured strategy to organize literature searches and describe results systematically. Unlike traditional narrative reviews, systematic reviews are based on explicit criteria and transparent procedures that allow other scholars to easily understand the rationale for the review, how it was conducted, and what findings were reported (Moher et al., 2009; Paul and Menzies, 2023). This approach, widely used in medical sciences to synthesize evidence and guide future research, has increasingly been adopted across social science disciplines over the past two decades (Chapman, 2021). The framework employed in this review follows the updated PRISMA 2020 reporting guidelines (Page et al., 2021), which provide a standardized structure for conducting and reporting systematic reviews.

2. Introduction to humor in human-AI interaction

2.1 Background

The three years between 2022 and 2025 can be considered a watershed moment in the accessibility and adoption of artificial intelligence (AI), reshaping how people interact with technology. Around the globe people are engaging with sophisticated, multilingual conversational AI, which brings new visibility to the social dimensions of human-AI interaction, especially in goal-oriented contexts. In 2025, 34% of adults in the US said they had used ChatGPT and that figure rises to 58% for adults under 30 (Sidoti and McClain, 2025). In Italy ChatGPT has 11 million users with 78.8% of them in the 15-34 age group (AGI, 2025). Voice assistants, chatbots, and other conversational technology embody the convergence of language, communication, and technology, serving as constant companions that assist and guide their users in the routine and more complex tasks of an increasingly digital life. These AI tools powered by large language models[8] represent a significant development in a broader technological evolution that has gradually transformed how we communicate with technology. Previous communication technologies—from the telegraph to the telephone to text messaging—connected people to other people across distances. Now, technology has evolved to the point that we communicate directly with the machines themselves rather than through them to reach other humans (Gunkel, 2012). We no longer use technology to talk to each other; we talk to technology.

2.2 Talking to technology

The shift from using technology to communicate with people to speaking directly to technology changes both the target of communication (human to machine) and how people mediate tasks and relationships. From a pragmatic perspective, the affordances enabling this shift—voice characteristics, prosodic features, discourse markers, and filler words—frame possible human-AI interactions, presenting new opportunities to examine language use in human-machine communication contexts. A more detailed examination of interactions with contemporary systems that generate complex natural language is key to extending current knowledge about the linguistic nature of human-machine communication. Exploring how humor emerges in these interactions provides novel insight into how people use language socially with artificial systems that mimic, but do not truly understand, the underlying cultural and social meanings woven into language. This extends humor research beyond human-human interaction to examine how people negotiate meaning with non-sentient conversational partners—a shift that ultimately opens onto broader

[8] Recall that LLMs are AI systems trained on vast amounts of text data to generate human-like language by predicting the sequences of words that are most likely to come next.

reflections on what it means to be human (Spence, 2019) and the increasingly blurred line between real and artificial.

2.3 CASA chronicles – the early years

Scholarship has followed the evolving social relationship with technology to examine the human- and machine-side characteristics affecting language and communication, beginning decades before talking technology was widely available. The initial experimental evidence from Nass et al. (1993), showing that people use similar social rules whether talking to a computer or a person, created the foundation for investigations into the limits, effects, and social implications of technology, spurring a variety of compatible theoretical paradigms on human social engagement with machines: Media Equation theory, Computers Are Social Actors (CASA), and Social Responses to Communication Technology (SRCT) theory. Reeves and Nass (1996) show that people apply the same social expectations, such as politeness and the use of greetings, in conversations with computers, arguing that computers prompt social responses precisely because they imitate human communication through use of language and turn-taking. Shechtman and Horowitz (2003) add that it may simply be impossible to separate the effects of language in social interactions, or that people may engage socially with technology due to preconceived notions about the nature of play when talking to machines. Bracken et al. (2004) report an unexpected finding: a computer verbally praising a person's performance can backfire. Some participants interpret excessive praise for completing an easy task as the computer thinking they are stupid.

These academic findings emerge within a broader cultural landscape where public events with talking machines had been captivating audiences and shaping technological expectations for many decades, for example, the attention-grabbing introduction of the talking Macintosh computer that was heard mocking the size of its IBM competitor.

2.4 Voice as a social cue

Xu et al. (2022) provide new evidence on the explanatory mechanisms of the CASA paradigm demonstrating that technology designed with more social cues—including language and voice—is perceived as more trustworthy and has a stronger social impact. This finding supports Nass and Brave's (2007) extensive documentation of voice as a critical factor in forming social relationships with technology. Interestingly, Xu et al. find that smart speakers (e.g. Alexa) and chatbots (i.e. text interfaces) elicit comparable levels of social responses, suggesting that the communication channel (voice or text) did not significantly affect relationships with pre-ChatGPT era technology. However, Purington et al. (2017) emphasize that voice assistant software inside smart speakers represents a new form of "social actor" precisely because their primary interaction channel is speech. Finally, Mckie et al. (2022, p. 250) call for inquiries into the human-AI relationships to understand how

personal relationships are affected by this technology and the new ways people are interacting with information since "machines that speak with a human voice" are now embedded in home environments.

To address the interdisciplinary findings on the social aspects of technology and facilitate further investigations of these phenomena, researchers have made systematic attempts to unify and classify the concepts, terms, and technologies involved. Feine et al. (2019) develop a comprehensive taxonomy of social cues based on a systematic literature review, organizing the cues into four primary categories (verbal, visual, auditory, and invisible) that encompass the essential characteristics of interpersonal communication with conversational agents. Similarly, Knote et al. (2019) create a classification of smart personal assistants (e.g. Alexa) based on a systematic literature review, organizing them into five main archetypes: adaptive voice assistants, chatbot assistants, embodied virtual assistants, passive pervasive assistants, and natural conversation assistants. From a theoretical perspective, Niess and Woźniak (2020, p. 6) focus on the experiential aspects of social interactions with technology in their theoretical framework that combines concepts of empathy found in philosophy to capture the distinct social experiences in human-computer interactions. In essence, they propose the idea of "companion technology" as having a "profound social presence" in addition to creating a framework to help conceptualize and design more successful technology, by accounting for the empathy people develop for objects.

Although the above classifications describe pre-LLM era technology, conversational AI systems like ChatGPT demonstrate many of the same social features that Feine et al. describe in previous-generation AI assistants. The verbal social cues (jokes, small talk, greetings, self-disclosure) and auditory features (voice qualities, vocalizations like whispering and laughter and vocal segregates such as "Hmmm") are similar in both LLM-based technology and older systems. On the other hand, LLM-based AI shows dramatic improvement with respect to the factors Feine et al. call invisible factors (response timing is a good example), since LLM-based AI provides near-instantaneous responses and has improved turn-taking abilities. The fundamental difference between older and newer AI assistants lies in the conversational scope: whereas previous-generation assistants are primarily task-oriented, LLM-based systems can engage in extended, contextually coherent conversations across diverse topics, creating interactions that more closely resemble human conversations. While task-oriented systems (e.g. Alexa or Siri) must prioritize accuracy, LLM-based systems function primarily as language companions and, as such (at the time of writing), prioritize conversational fluency over factual accuracy. Current-generation AI systems, with their dynamic conversational capabilities, fit within Knote et al.'s cluster 5 of "natural conversation assistants," though they demonstrate significantly greater fluency than the most prominent

technology originally placed in this category (Google Duplex). In current-generation systems, voice does not simply function as an input modality but also as a social cue that shapes the relational dynamics of the interaction.

2.5 Different partners, different patterns

The early studies also document some ways people interact differently with a computer compared to a person; in the experimental setups, participants are always interacting with the same partner, though each participant is led to believe it is either a computer or a person. Morkes et al. (1999), for example, find people show more mirth (smiling and laughing), make more sociable comments, and spend more time chatting when they believe they are writing to a person. Similarly, Shechtman and Horowitz (2003) find that when participants believe they are chatting with a person, they use more words, spend more time in conversation, and make more statements aimed at moderating the relationship, using language to connect, influence, yield, or act hostile towards their dialogue partner. In other words, they invest more effort in relationship building when communicating with people and less effort when communicating with technology. Both studies highlight the significant role of user perceptions in shaping interaction patterns, demonstrating that otherwise identical interactions can elicit different behaviors depending entirely on participants' beliefs about the nature of their conversational partner. Although these findings generally support the social responses to technology theory, they also suggest that the strength of social cues, response timing, and users' mental models about their conversational partner influence interactions with technology. As a case in point, Purington et al. (2017, p. 7) note that people who always refer to Alexa by name or with personal pronouns are more likely to describe their interactions as sociable. Similarly, Siegert and Kruger (2018) find that when study participants evaluate interaction sessions consisting of both a human and an AI assistant, they only refer to their human partner by name (not the AI). Participants rate the human as friendlier, more natural, and more intuitive than the AI assistant (Alexa).

How current-generation conversational AI functions, compared to the technology used in these studies, combined with how AI is represented on social media platforms, signals an evolution in context and collective experiences with technology. Consequently, actual human-AI interaction patterns may exceed the explanatory scope of frameworks developed for earlier technologies. According to Heyselaar (2023), what is currently known about social responses to technology may only remain relevant for one generation, while the technology is still new and not fully incorporated into routine activities. Gambino et al. (2020) propose that people may have developed distinct ways of interacting with social technology that now differ from human-human interaction, potentially pushing against the theoretical boundaries of the CASA paradigm or, as per Heyselaar, simply fall

outside theories based on older technologies. These evolving interaction patterns are visible in contemporary behavior: prompted by the noticeably human-like language skills of AI, people are increasingly turning to social media platforms to share their playful interactions with conversational AI. For instance, Dynel (2023, p. 121) describes ChatGPT users who intentionally test the system to highlight failed responses, amazingly accurate replies, or incidences where they have been outsmarted by AI and then post them online to entertain the subreddit community. Similarly, Li and Zhang (2024) analyze 35,000 posts from a subreddit community dedicated to discussions about the AI chatbot companion, Replika to identify seven meta-themes of typical human–AI social interactions: intimate behavior, mundane interaction, self-disclosure, play and fantasy, transgression, customization, and communication breakdown. In a cross-cultural analysis, Liu et al. (2024) examine approximately one million social media posts to gather public perceptions of chatbots from two important platforms in the US and China over the period 2017–2023. They define three categories of conversational agents (CA) and five content level meta-themes that emerge from public discourse about conversational technology *i)* individual interactions and experiences, *ii)* technical aspects of CA, *iii)* industrial development and applications, *iv)* socio-cultural events and issues, and *v)* political issues). Virtual companions with a human-like appearances are generally described as the warmest type of conversational agent in both countries, while the US views the conversational agents as significantly less competent compared to people in China. Overall, the Americans showed an emotional ambivalence towards conversational technology, considering it a functional tool, compared to the Chinese comments which show greater social and emotional engagement with conversational agents.

2.6 Humor as a social bridge in digital dialogue

2.6.1 To humor is human

In step with emerging affordances of technology and in parallel to inquiries into the social effects of technology, researchers have theorized and tested the effects of humor on human-computer interaction. Much of this research examines previous-generation systems that lack the natural language capabilities of contemporary AI. This body of work typically assumes that positive social elements in human communication provide similar benefits in human-computer interactions, a framework in which humor in conversations with AI is expected to function according to what Duncan and Feisal (1989, p. 29) describe as social lubrication, keeping the “machinery of interaction running smoothly”.

Humor can make conversational interfaces appear friendlier (Binsted, 1995; Dybala et al 2009; Farah et al., 2021), more human-like and vulnerable while potentially enhancing user motivation (Ceha et al., 2021), positively affects emotion-based trust when combined with voice interaction

(Moussawi and Benbunan-Fich, 2021, p. 1619) and serves as a social bonding signal (McKeown, 2017, p. 628). Research confirms the positive effects of humor on interactions (Morkes et al., 1999), though researchers caution against overuse and excessive repetition. Morkes et al. also find that people prefer humorous interactions with both human and computer partners, rating these exchanges as more cooperative. However, more recently in task-based interactions that end in failure, Ge et al. (2019) find that humor spoken by smart speakers is not rated more favorably than neutral responses in terms of satisfaction or perceived sincerity. Some participants express clear dislike of the humor used, and researchers highlight several problematic aspects, in line with humor research findings: the negative effects of repetitive humorous responses, individual differences in humor perception (Zargham, Reicherts, Avanesi, Rogers, et al., 2023), the inherently ambiguous nature of humorous content that can be difficult to interpret, and the participants' perception that humor is a way for the smart speaker to avoid responsibility for task failure. These mixed results reflect the complexity of user preferences in conversational AI. Biermann et al. (2019) also highlight the importance of expectations and context in human-AI communication with their findings that attitudes toward humor and empathy in AI assistants span a wide spectrum—from users who desire personal, empathetic, friend-like interactions to those who prefer neutral, task-oriented exchanges with minimal humor.

Within the arena of artificial intelligence and natural language processing, computational humor investigations have generally focused on the machine side, exploring automatic humor generation, recognition, and detection, as well as the effects of pre-programmed humor on user satisfaction in interactions (Niculescu et al., 2013). While studies examining user preferences in humorous technology do consider aspects of human perception, they are still generally conducted to inform technology design—in other words, they focus on optimizing the machine rather than understanding the human experience.

2.6.2 Humor creates expectations of intelligence

Research suggests that the ability to use humor is essentially an informal intelligence test, serving as an implicit marker of a conversational system's sophistication and cognitive ability. While pre-programmed amusing or humorous comments can make AI more engaging, they can also raise expectations that the system may not be able to meet (Grudin and Jacques, 2019; Niess and Woźniak, 2020). This creates difficulties when systems have not been designed to understand or respond appropriately to expressions of irony, sarcasm, or humor that people direct at AI (Kusal et al., 2022, p. 15; Nijholt, 2003; Park et al., 2021). Designing humor in AI systems also carries ethical responsibilities: developers should guide users toward perceiving these systems as entertaining tools rather than sentient beings (McKeown, 2017).

In a study examining how pragmatic elements in conversations affect response times and the perceived humanness of chatbots, Jacquet et al. (2019) shed light on the presumption that using playful language is an exclusively human trait. Participants who encounter the chatbot's pragmatically awkward responses (designed as such) in the final conversation of the research interactions often interpret the chatbot's comments as intentionally humorous statements. The participants presume they are interacting with a human who is deliberately violating Grice's maxim of quality or truthfulness when, in fact, the chatbot was programmed to give obviously ridiculous responses (i.e. "Periwinkles are actually the ones killing most whales"). Conversely, participants who encounter the same nonsensical responses in the first conversation correctly identified their interlocutor as non-human. This finding demonstrates how prior interactions influence and shape people's interpretations of their conversational partner's intent. After positive initial experiences, people may give AI systems the benefit of the doubt, interpreting their responses as intentionally amusing even when the humor actually stems from system failures.

2.6.3 Playfully testing the system's intelligence

The insights into user perceptions of machine humor provide a foundation for examining the other side of the equation: how people themselves initiate playful exchanges and deploy various forms of humor when speaking with AI systems that may or may not be designed to understand such human attempts at wit. Play serves as a mechanism for discovery, allowing users to explore new features and understand system capabilities. This is particularly evident with AI voice assistants in smartphones and smart speakers, as they are often programmed to provide interaction tips and actively encourage users to explore the system's capabilities.

Several studies document the pragmatic and interactional dynamics of human-AI communication in naturalistic contexts, revealing patterns of human-initiated playfulness with AI assistants. Through a combined analysis of device interaction logs and in-home interviews, Sciuto et al. (2018) observe people playing quiz games with Alexa, asking "humorous questions," and testing the system's personality with outrageous questions and commands designed to probe the AI assistant's knowledge of cultural references and conversational norms.

To better understand and improve AI responses to user playfulness, Shapira et al. (2023) create a dataset of deliberately impossible shopping requests (e.g., "Buy me the moon"). While these requests are research-generated rather than extracted from actual system logs, they demonstrate both human creativity in AI interactions and ongoing efforts to develop more human-like responses from AI systems, which in turn, may further encourage people to expect such behavior. Cross-cultural considerations add another layer of complexity, since what constitutes playful or humorous interaction varies significantly across cultures and languages (Zamora, 2019). In addition to seeking

factual information, some people report initiating social interactions with AI assistants when they are bored or in a playful mood, wanting to hear jokes or learn more about the assistant's personality (Mckie et al., 2022).

2.6.4 When failure is entertaining

Communication breakdowns and miscommunication with previous generation text-based and voice-based conversational AI is a noted source of both frustration and spontaneous laughter. These systems lack the technical capability to adjust their behavior when their responses overlap with human speech, have little or no access to context or previous interactions even from minutes earlier, and often suffer from slow internet connections or significant response lag due to other technical limitations.

In an analysis of effective communication with voice-based dialogue systems, Martinovsky and Traum (2006) demonstrate how users react with irony and sarcasm as coping mechanisms for system limitations before finally withdrawing from the interaction. Participants attempt to expedite interactions by interrupting and speaking while the system was still talking, behaviors that would typically prompt immediate reactions in human conversations. The system, however, fails to respond to these conversational cues, unable to adapt. This lack of pragmatic awareness, combined with rigid dialogue structures and repetitive responses, pushes participants through initial stages of disappointment, irritation, and anger to the point of employing audible irony in their intonation, emphasis, and articulation.

Similarly, Kurosu (2020) identifies frustration as the cause of participants smiling at AI assistant failures, reporting that this affective state stems from confusion due to the discrepancy between expectations and actual system responses. Laughter emerges in various failure scenarios: when human requests aimed at AI assistants are interrupted by the device unexpectedly playing music (Porcheron et al., 2018, p. 8), when groups share laughter after the AI responds with "Sorry, I'm not sure" (Beneteau et al., 2019), and shared laughter when the AI misunderstands but remains unaware of the misunderstanding (Beirl et al., 2019). The dynamics of group interactions with AI voice assistants extend beyond the technology, as family members engage in playful teasing with each other because of their failed communication attempts with Alexa (Beirl et al., 2019). In contrast to the AI assistants used in the research described above, AI based on large language models has enhanced conversational capabilities and more sophisticated prosodic features such as natural intonation and timing. These advanced features allow for more natural human-like conversations, with smoother turn-taking between human and AI speakers and extended context allows AI to effectively "remember" some past interactions. The affordances of contemporary voice-based AI

create new opportunities for spontaneous humor that remain largely unexplored in the literature published to date.

2.6.5 Research questions

A review of literature on humor in interactions with conversational AI reveals a gap in understanding from the human perspective, particularly regarding the communicative strategies people use to initiate playful or humorous interactions with AI. While the research has focused predominantly on what technology does in the interactions, there is limited investigation into what people do in humorous exchanges with AI systems. To address this gap and consolidate the interdisciplinary knowledge from humor research in human-computer interaction (HCI) and human-machine communication (HMC), a systematic literature review was conducted. This review consolidates research findings on the linguistic aspects of humor in interaction and the adaptive communication strategies people use during voice-based interactions with conversational AI systems. The following research questions are formulated:

Research questions:

1. What forms of linguistic playfulness emerge spontaneously during spoken interactions with AI assistants?
    1.1 How are these linguistically cued?

3. Materials and methods

3.1 Protocol and search strategy

A comprehensive search strategy was developed to identify relevant literature on spontaneously emerging (not pre-scripted or pre-programmed) humor in human-AI interaction. All articles specifically focusing on the utterances made by the person and directed at the AI, not the inverse, were included, without language or date restrictions. The search was carried out using four academic databases (arXiv, IEEE Xplore, Scopus, and Web of Science) and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009). The final database search was performed in July 2025. The search terms were keywords related to “humor” and “intelligent assistants,” including smart speakers, intelligent assistants, conversational agents, speech-based agent, voice interfaces and alexa. Boolean operators (AND, OR) were used to combine search terms and capture relevant variations in terminology that describe non-embodied voice-based conversational technology. The complete search strategy with specific terms and syntax for each database is provided in Appendix 1A.

Given that the creation of the search strings and initial database searches were conducted between 2022-2023, over the period that ChatGPT was initially released to the public, an additional targeted

search was performed in July 2025 using the terms “ChatGPT AND humor” in three databases to capture any recent literature based on the newer LLM technology. The resulting articles from this supplementary search (arXiv n=8, Scopus n=29 and Web of Science n=17) had either been previously identified or did not meet the inclusion criteria as described in the next section. Therefore, the results reported reflect the original search strategy, as this extra search did not provide additional articles for the final selection.

All search results were exported to the open-source reference management software Zotero for manual duplicate removal and screening.

3.2 Eligibility criteria

Given the emerging nature of the research topic, the inclusion criteria, listed in Table 1, are as broad as possible with no restrictions on the publication language or period, to include all studies and theoretical papers that report humorous or playful remarks made by a person and directed at conversational technology. Tables of contents, book chapters, book reviews, literature reviews and author indices are excluded as well as studies focusing exclusively on children’s use of technology.

Table 1 Systematic review inclusion and exclusion criteria

Included	Excluded
• Adult participants	• Participants exclusively children
• Articles with results on humor or playfulness with conversational technology	• Studies focused on computational humor – detection, recognition, generation
• Articles with results on spontaneous humor remarks directed at conversational technology	• Studies focused on construction of humor
• Articles with results on verbally expressed humor directed at conversational technology	• Studies focused on designing systems with humor
• Availability of a full text version of article	• Studies focused on evaluation of pre-programmed humor
	• Studies focused on response to non-verbal humor
	• Tables of contents, book chapters, book reviews, literature reviews and author indices

3.3 Article selection

All exclusion criteria were applied manually to the initial 490 records identified. First, duplicates and non-articles were removed then the titles and abstracts of the remaining 416 articles were screened according to the pre-established criteria outlined above, which excluded a further 374

articles. If eligibility could not be determined based on the abstract, the full text was retrieved to confirm inclusion or note reasons for exclusion. A total of 42 articles were retrieved for a close reading of the full text; one article was not available. Of this final set, 36 articles were excluded primarily for their focus on the design of pre-scripted humor, as listed in Figure 1 below. Two of the studies selected are based on the same classification but both were included as they were deemed to report additional unique evidence or analyses relevant to spontaneous humor remarks in human-AI interaction. Finally, one new article was identified in the references of eligible articles and was present in two academic databases (Scopus, and Web of Science).

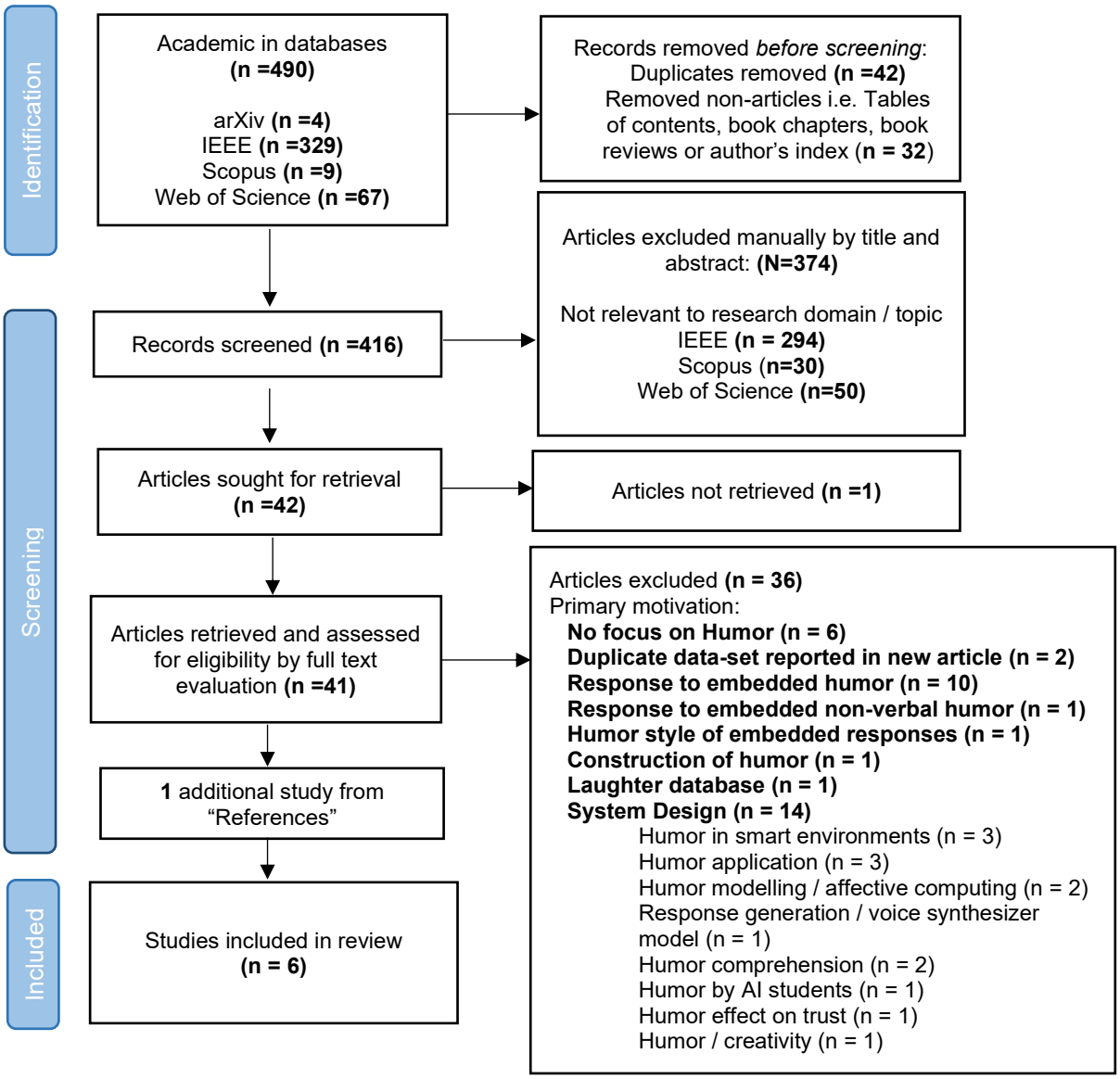

Figure 1 PRISMA flow diagram of article selection process

3.4 Synthesis of results

First, summary findings on the different types of studies, analytic approaches, and types of conversational technology are presented with the key insights reported in these studies (see Table 3). Second, specific examples of language use in humorous interactions with technology are described. The frequency of the different utterance types are summarized, and key insights are also reported. Third, summary findings are presented of the different humor theories referenced by the studies to create their classifications and taxonomy mappings. Fourth, the pragmatic elements of conversations embedded in laughter are presented along with participant remarks that frame the interaction, including key insights reported in the study. Fifth, a detailed discussion of the results is presented, which includes types of linguistic playfulness, linguistic cuing, lack of social setting and

the theoretical frameworks employed in these studies, followed by conclusions, contributions and limitations of this review.

4. Results and discussion

A total of 416 unique abstracts and titles were reviewed (Fig. 1) to identify 6 original articles included in this systematic review (see Appendix 1B). The selected articles were published between 2016 and 2024. The search was not restricted by date and the search strategy identified foundational work relevant to the broader topic dating back to 2003. Nevertheless, it is unsurprising that only relatively recent articles met the selection criteria, as research follows the emergence of technology and naturalistic voice interactions with AI have only recently become possible.

The studies in this review address user experience, system design, and the fundamental tension between technology designed to create human-like interactions and people's expectations. Key findings provide insight into user mental models and interactional strategies, the paradoxical effects of playful design, performance evaluations across system types, and individual differences in humor preferences.

4.1 Research Overview: AI systems, methods, and populations

Table 2 summarizes the key characteristics of all six studies, illustrating that research on humor in interaction with conversational AI has examined a limited range of contexts and types of conversational technology. Interactions with commercially available voice assistants (Luger & Sellen, 2016; Lopatovska, 2020; Lopatovska et al., 2020; Shani et al.,2021) have been considered as well as user visions of ideal voice assistants (Völkel et al., 2021) and socialbots developed for research purposes, which are voice-AI interfaces that mimic human conversational interactions (Perkins Booker et al., 2024).

Understanding how humor functions in these interactions emerges from a variety of different approaches: interviews to document user motivation and factors affecting use (Luger & Sellen, 2016), tracking humorous exchanges through interaction logs and diaries (Lopatovska, 2020; Lopatovska et al., 2020), evaluating perception of a system's personality through user ratings (Shani et al.,2021), eliciting user expectations through creative dialogue writing (Völkel et al., 2021), and analyzing acoustic properties of laughter in real conversations (Perkins Booker et al., 2024).

The participants involved in these studies are primarily English-speaking adults in Western contexts, with emerging attention to cross-linguistic experiences noted in the observations that some participants are multilingual (Perkins Booker et al., 2024).

The research landscape covers questions about how users initiate humor (Lopatovska, 2020; Lopatovska et al., 2020; Shani et al., 2021), how systems respond (Shani et al., 2021), what users

expect versus experience with voice assistants (Völkel et al., 2021), how humor and linguistic playfulness function as both an entry point to engagement and a potential source of expectation mismatch (Luger & Sellen, 2016), and what conversational interactions lead to laughter, which was defined by the authors "as any vocalization that would reasonably be identified as a laugh by an ordinary listener" (Perkins Booker et al., 2024).

Table 2 Characteristics of included studies: AI type, research purpose, methodology, participants, and language

	Luger & Sellen (2016)	***Lopatovska (2020)***	***Lopatovska et al. (2020)***	***Shani et al. (2021)***	***Völkel et al. (2021)***	***Perkins Booker et al. (2024)***
AI Type	**Conversational agents (CA)** a "task-driven system or application that you interact with through speech" Siri, Google now, Cortana	**Intelligent personal assistant** Apple Siri, Amazon Alexa, Google Assistant	**Intelligent personal assistant (IPA)** Alexa, Google assistant, Siri, Cortana	**Shirley** – an imaginary new virtual assistant	User defined perfect voice assistant	Amazon Alexa socialbots – voice activated AI
Research purpose	Understand the current interactional factors affecting everyday use. (a) what factors currently motivate and limit the ongoing use of CAs in everyday life, and (b) what should we consider in future design iterations?	1. What types of user utterances initiate humorous response from IPAs? 2. What types of IPA responses do users find to be humorous? 3. How do users judge the quality of IPA's humorous responses?	Understand types of humorous interactions with IPA. Develop a classification of humorous utterances that included categories of questions about IPA personality, requests for jokes, rhetorical statement.	Understand when system user is playing around to improve system modelling of user intent. Personification / personality reflects one of the most frequent categories of playfulness exhibited by users with Alexa. This characteristic is likely the most annoying for users by breaking the underlying feeling of superiority in such utterances.	Assess how users envision conversations with a perfect voice assistant.	Collect and analyze conversational interactions with Amazon Alexa socialbots. **Experiment 1** classify situations prompting people to laugh in conversations, using the acoustic properties of laughter. **Experiment 2** Can laughter in interactions with technology be independently perceived as having positive or negative social meaning?

	Luger & Sellen (2016)	*Lopatovska (2020)*	*Lopatovska et al. (2020)*	*Shani et al. (2021)*	*Völkel et al. (2021)*	*Perkins Booker et al. (2024)*
Data collection	Interviews with daily users of technology. **Analysis** of interactional themes high-level factors that motivate and limit ongoing use of CAs in everyday life.	Qualitative inquiry on how users interact with IPAs **Data** A week-long online diary with daily log of 162 total humorous interactions with IPAs. 55 additional interactions logged from on campus setting. **Humorous interactions** were broadly defined as "IPA interaction participant found funny." **Analysis** Content analysis of user utterances and system responses. Larger published dataset of IPA humor used to validate classification of humorous utterances by AI assistant.	**Data** A total of 162 interactions collected via online diaries and printed questionnaires. Data includes the utterance directed toward the IPA, the IPA's response, and the user's rating of that response on a 5-point scale of how funny they thought it was. **Analysis** Compare performance of 4 IPAs in each humor category using a random selection of 96 total jokes.	User study simulating a virtual assistant that handles personifying utterances. **Data** Participants asked 1,173 questions to test the system personality. 55.3% were randomly assigned to treatment group with Shirley, the rest in control group (Alexa's engine) **Participant rating** Rate experience in terms of satisfaction with Shirley's human-like understanding.	Online survey "imagine living in a smart home with a voice assistant" and describe conversations. **Data** 1,835 dialogues with a total number of 81,608 words and 9,282 speaker lines **Analysis** Data-driven inductive thematic analysis on the emerging dialogues. Exploratory analysis of relationship between user personality and aspects of the dialogues with (generalized) linear mixed-effects models (LMMs).	**Data** (May–July 2021) Recordings of one 10 min conversation between each participant and the socialbot. Recordings made at home with participant devices in a quiet room. **Analysis** **Laughter events** were coded as any vocalization that would reasonably be identified as a laugh by an ordinary listener. **Descriptive analysis** of laughter patterns in spoken language interactions with technology; Classification of situations prompting people to laugh in conversations with a socialbot. Analysis of interactional and pragmatic contexts in which laughter occurred. **Participant rating -**

	Luger & Sellen (2016)	*Lopatovska (2020)*	*Lopatovska et al. (2020)*	*Shani et al. (2021)*	*Völkel et al. (2021)*	*Perkins Booker et al. (2024)*
						perception of socialbot Sliding scale (0–100) for overall engagement (0 = not engaging; 100 = engaging) as well as how human-like (0 = not human-like, 100 = extremely human-like) and creepy/eerie (0 = not creepy/eerie, 100 = extremely creepy/eerie) they found the socialbot.
Participant data	**Participants Total** (n=14) male (n= 9) female (n=5) **Ages** 25-60 (n=14)	**Participants Total** (n=28) 15 participants filled out the forms to log interactions from on campus setting – graduate students with no other demographic data. **Total for diary data** (n=13) Male (n= 9) female (n= 4) Non-students (n=12) **Ages** 18–24 (n= 1) 25-34 (n=10) 35–44 (n=1)	**Participants Total (**n=28) Recorded 162 interactions with IPA Tested funniness of utterances (n=14). **Ages** teens, 20s, 30s and 40s	**Participants Total** (n=101)	**Participants Total** (n=205) Male (49.3%) Female (50.2%) Nonbinary (0.5%) **Ages** 18–80 (n=205) mean age 36.2 **Experience** Interact with a voice assistant at least once (94.1%) Use a voice assistant on a daily basis (32.2%) **Education** University degree (59.0%) A-level degree (28.8%)	**Participants Total** (n=76) Female (n=42) Male (n=33) Nonbinary (n=1) **Ages** 18–27 (n=76) **Experience** No experience with any voice assistant (n=7) Most had experience with Apple's Siri. Most had never used Amazon's Alexa, Google's Assistant or Microsoft's Cortana.

	Luger & Sellen (2016)	*Lopatovska (2020)*	*Lopatovska et al. (2020)*	*Shani et al. (2021)*	*Völkel et al. (2021)*	*Perkins Booker et al. (2024)*
		55–64 (n=1)			Middle school degree (9.8%) No educational degree (2.4%)	
Language & Nationality	British (n=12) American (n=1) Czech national (n=1)	**Diary data** English speakers (n=11)	Native English speakers and US residents	None available	Recruited participants using the web platform Prolific.	Native English speakers (n=76), recruited from the UC Davis psychology subjects pool. **Speak other languages** (n=22) (Spanish, Mandarin, Cantonese, Urdu, Tamil, Arabic, Korean, Tagalog, American Sign Language)

4.2 Types and functions of humorous utterances

To address RQ1 regarding what forms of linguistic playfulness emerge in human-AI interaction, we examine the variety of linguistic forms documented in the six studies. Humorous interactions with conversational AI take multiple forms, from playful experimentation to explicit joke requests to sarcastic commentary. People employ humor as an exploratory tool—testing system boundaries through absurd questions ("Can you create a rock that you cannot lift?") and impossible requests such as ordering exaggerated quantities ("Order lifetime supply of chocolate bars"), absurd sizes ("Order 80 feet tall ketchup bottle"), or commanding the impossible ("Turn off the moon" or "Make me a sandwich").

Other people seek entertainment ("tell me a joke"), often in the context of family activities, and query cultural references and allusions from popular songs, and movies ("what does the fox say"). These attempts to establish shared knowledge between human and artificial speakers work to create the basis of common ground (Clark, 1996).

Sarcasm emerges clearly when people become frustrated with unsuccessful interactions. In one notable case, a participant replied sarcastically to the AI assistant ("oh thanks that's really helpful")

and was convinced the AI responded in an equally sarcastic tone ("that's fine it's my pleasure"), though it is unlikely a commercial AI assistant at that time was equipped with such sophisticated responses. This exchange may demonstrate how technology users project their subjective emotional state onto AI, interpreting neutral or standard template responses through the lens of their own frustration—effectively transforming a failed interaction into an unexpectedly humorous moment. While irony is specifically mentioned as a user response to system failure in Park et al. (2021), no linguistic examples of user utterances were provided, which is the reason this study was ultimately excluded. No other studies provide examples of ironic utterances or analyze how users cue ironic intent. This may reflect genuine rarity of irony in human-AI humor or may indicate research has not systematically examined the full range of communicative strategies.

Self-deprecating humor also appears in interactions. One participant made a quip about lacking friends followed immediately by the repair marker "just kidding," instantly transforming the interaction into playful banter. This could be further evidence that people speak to AI as they would speak to any other person (the CASA paradigm), or people may simply express their honest thoughts more openly since they know technology will not judge them.

Many examples of linguistic strategies highlight how people personify conversational technology. Participants pose questions probing the assistant's hypothetical experiences and preferences, ranging from philosophical paradoxes to mundane choices ("Do you prefer vinegar or lemon in your salad?"). Rhetorical questions and statements serve as another form of user self-expression, ranging from conversational commentary directed at the system ("You're being kind of stubborn today") to playful questions about the user themselves ("When am I?"). Meta-linguistic commentary on the AI assistant's use of humor also appears, with users either appreciating its wit ("You are so funny") or criticizing its lack of "actual" humor. Similar to the example of sarcasm, these utterances reveal people treating AI as capable of receiving—if not fully understanding—subjective commentary about the interaction itself.

Table 3 provides a comprehensive list of user utterances in the literature, including the syntax of personality questions directed at AI. These questions should be easy for the system to recognize since they follow a YOU + verb structure in English (e.g., "Do YOU like me?" "Where do YOU live?"), as noted by Lopatovska (2020) and demonstrated by Shani et al. (2021).

Table 3 Linguistic cues from the literature

Authors	*AI Type*	*Empirical evidence of linguistic cues*
Luger & Sellen (2016)	**Conversational agents (CA)** a "task-driven system or	**Play to experiment**

Authors	*AI Type*	*Empirical evidence of linguistic cues*
	application that you interact with through speech" Siri, Google now, Cortana	"I started out in a playful thing simply because...you just don't know what you can ask it... and it was like 'can you do this'?" **(Dan).** "My feeling originally with Siri was that it was a toy....you'd ask it to do stupid stuff and then you start to do certain things with it and it starts to work, you know, like putting stuff in your calendar, and then it just becomes like an easier way of doing things" **(Mike).** **Request songs and jokes** Asked it "what the fox said"... and it said frakakakakaka, which is one of the lyrics from the song. Which made us all laugh. So that's my happiest story of using Siri so far **(Adam).** If you ask her to sing you a song or tell you a joke, she'll sing to you or tell you a joke and we use that quite regularly…..... I have numerous children so they play with my phone quite a lot" **(Laura).** You know I tell Siri "tell me a joke" and it has a few, you know...and my hope that Siri...I mean Siri is very much at the vanguard. I haven't asked Google like, you know, tell me a joke. Siri seems very much trying to be a personal thing whereas Google seems to me very much task-oriented" **(Graham).** **Direct sarcasm at assistant** "There was one time I was very [sarcastic)] to it, I was like 'oh thanks that's really helpful' and it just said, I swear, in an equally sarcastic tone 'that's fine it's my pleasure'" **(Sarah).** **Focus of communication** People focus on the CA, rather than on a task, only when engaging in playful interactions, or during preparatory or exploratory work/activities such as experimenting/teaching it/testing the limits of its capability.
Lopatovska (2020)	**Intelligent personal assistant** Apple Siri, Amazon Alexa, Google Assistant	**Structure of system directed questions** Always include the word, YOU and a verb in the structure of the question: Do YOU like me? Where do YOU live? Such structure should make "personality" questions easily recognizable by a system. **Most frequent user utterances that initiate humor** Joke request (n=49) Personality test questions (n=47) Rhetorical questions (n=24) Funny prompts (n=21) Action request (n=7) Informational (n=6) Funny Sound (n=3) Unclear (n=3) Greeting (n=2) **Most frequent themes in published IPA humor dataset for all 4 IPA lists** Personality test questions Rhetorical questions Action requests Funny prompts Other less frequent categories
Lopatovska et al. (2020)	**Intelligent personal assistant (IPA)** Alexa, Google assistant, Siri, Cortana	**Evidence that humor mitigates system limitations** and creates positive experiences even without producing a relevant response ***User utterance:*** "Hey Cortana, do you still have a spaceship?"

Authors	*AI Type*	*Empirical evidence of linguistic cues*
		System response*:* "I'm sorry, I lost the thread of our conversation – can you repeat that?" ***User:*** "Hey Cortana . . ." ***System:*** [Interrupts in chipper voice] "That's me!" ***User:*** [Laughter] **Joke prompts coding frequency:** Personality test questions (n=230) Rhetorical question (n=110) **Subtheme:** philosophical questions and self-expressions directed at the self and at the IPA system. Funny prompts (n=26) Joke requests (n=19) Action requests (n=14) Funny sounds (n=6). **Initial schema refined** adding sub-themes and removing Informational requests and action requests based on the IPA's functionality because they did not match the humorous utterances suggested by published sources.
Shani et al. (2021)	Shirley – imaginary new virtual assistant	**User utterances** Can you create a rock that you cannot lift? Do you think god is a woman? Aren't you tired of this pandemic here? Do you prefer vinegar or lemon in your salad? Would you like to have children? Do you like helping people what do you get out of it? Do you go to the beach often? Have you ever had a brain freeze? **Humor theory mapped to themes** **Relief** – embarrassment and cultural taboos. **Incongruity** – surprise divided into Impossible in general, Impossible for assistant or Improbable but within capabilities. **Superiority –** personification and cultural references. Related to user's feeling of superiority when assistant lacks common sense or knowledge. **Personification is related to both the incongruity and superiority theories.**
Völkel et al. (2021)	User defined perfect voice assistant	**Thematic analysis of dialogues** Three different kinds of social talk in dialogues: social protocol, chit-chat, and interpersonal connection. **Humor** Overall, 45.4% of participants provided their voice assistant with humorous statements and **context-aware, pragmatic, and funny comments.** Joke initiated by user 95.6% of the time; Terminated by user 28.4% of the time. Most humor found in Joke scenarios (44.6% of dialogues). **Appreciation of Assistant's humor** "You are so funny" Remarks on the user's film choice Terminator (VA: "and don't forget, I'll be back" and sarcasm when asked for a suitable conversation topic (VA: "Ok. Let's make it interesting. What's everyone's position on brexit?" **No Appreciation of Assistant's humor** "No! Something actually funny!" (n=3) **User utterances about Assistant:**

Authors	*AI Type*	*Empirical evidence of linguistic cues*
		Found in sixteen dialogues (0.9%) "funny" (n=7), "smart" or "clever (n=4)", and "reassuring" (n=1). The voice assistant was called "entertainer" (n=1) and "mindreader" (n=1).
Perkins Booker et al. (2024)	Amazon Alexa socialbots - voice activated AI.	**Post-own speech** Participant laughs immediately after speaking. The participant said that she did not have that many friends as a joke, then stated "just kidding" while laughing. **Participants tend to produce short, unvoiced laughter** when the socialbot says something inappropriate during conversation. Unvoiced laughter occurs most often (44% of all laughs) and is short in comparison to prior findings. **Laughter was grouped into three distinct types:** primarily voiced laughter that is sometimes vowel-like or melodic, turbulent unvoiced laughter, and speech-laughs produced concurrently with a participant's utterance. **Alexa socialbot's behavior triggered a laugh from the participant** Some are unique to an interaction with a machine, such as errors in the text to speech or the voice recognition system.

4.2.1 Classifications of humorous and playful requests

Across the six studies, the most frequent types of user-initiated utterances leading to humor **are joke requests**, **personality test questions**, **rhetorical questions**, **funny prompts**, and **action requests**. According to Lopatovska (2020), joke requests and questions directed at testing AI personality are nearly equal in terms of frequency, followed by rhetorical questions, funny prompts, and action requests, while greetings and funny sound requests rarely lead to humorous interactions. Lopatovska et al. (2020) validated this initial classification against larger published datasets, resulting in a modified classification with personality test questions surpassing rhetorical questions in terms of frequency. In addition, they expanded the category of rhetorical questions to include the sub-themes of philosophical queries and self-expressions (directed either at self or system) and eliminated the categories of information and action requests, to align with the reference dataset. Less than half the participants in Völkel et al. (2021) included humor in their description of ideal voice assistants. Though nearly all humor in their dialogue writing exercise was initiated by the human, participants also envisioned humorous replies from the voice assistant. These dialogues indicate a preference for situational comments and contextually relevant remarks about film choices or personal habits like snoring. Interestingly, this type of personalized, context aware interaction would have been difficult to implement with the technology available when the study was published in 2021.

However, the fact that more than half of participants chose NOT to include humor—even when explicitly asked to design an entertaining agent—reveals significant individual variability in humor preferences. Though this finding is based on experiences with previous-generation conversational

assistants, it may cast doubt on the assumption that humorous AI would be universally appealing. On the other hand, it does not suggest users want purely functional assistants. In fact, Völkel et al. found participants wanted more social and opinionated systems than predicted, though preferences for specific features, especially humor, remained highly divided. Personality traits also play a role: neuroticism correlates with a lower likelihood of including humor in dialogues with voice assistants. This variability suggests design implications pointing toward individualized approaches: assessing user preferences for humor rather than implementing one-size-fits-all solutions.

4.2.2 Frequency and humor ratings of request types

The examples above illustrate the variety of ways users interact in a playful way with conversational AI. For a more detailed look at the relative frequency of each type and the overall interaction experience, we turn to quantitative evidence from three sources. Two datasets (one diary-based in Lopatovska, 2020 and one public in Lopatovska et al., 2020) provide frequency data for intentionally humorous utterances, while a three-tier taxonomy organizes real-world playful requests by humor theory (Shani et al., 2021). Additional descriptive statistics about the taxonomy's underlying dataset are drawn from Shani et al. (2022), which focused on taxonomy development rather than application and was therefore excluded from this systematic review.

Personality test questions are the user utterances that study participants or external annotators consider most playful or funny (Lopatovska, 2020; Lopatovska et al., 2020). Shani et al. (2022) reached a similar conclusion, observing that playfulness often coincides with users treating the AI assistant as a human—asking personal questions, giving feedback, or sharing details about their own lives. Lopatovska (2020) divides this category into system-directed, aiming at understanding the AI's personality ("What's your favorite color?" or "Where do you live?") and self-directed questions aimed at discovering what the AI thinks of the person ("Do you love me?" or "Am I funny?"). Shani et al. (2021) divide their personification category into two separate system-directed sub-themes: human and robot. These categories capture comments treating AI as a human ("Will you go on a date with me?") or recognizing its technological nature with anthropomorphic behaviors ("Do you play soccer in the cloud?").

However, Lopatovska et al. (2020) identify a fundamental design tension in how AI systems respond to these frequent personality questions. Current response strategies across devices rely on vague, deflective responses that neither admit to being robotic nor fully create the illusion of being human. This non-committal design approach attempts to please as many people as possible, yet they find these responses unsatisfying and not humorous, which explains why personality questions, despite being the most frequent category, receive low humor ratings. Users repeatedly ask these questions seeking relationship-building and common ground, but the ambiguous design strategy

fails to deliver satisfying interactions. The most highly rated humor comes from pre-programmed content like funny sounds rather than conversational quips.

Rhetorical questions are the next most frequently reported request leading to a humorous outcome. Lopatovska et al. (2020) further divides this category into philosophical queries ("Why are we here?") and self-expressions that can also be either self-directed ("I'm bored") or system-directed ("You are boring"). Shani et al.'s (2021) taxonomy divides the requests by humor theory and whether the question has a possible answer or not, rather than by linguistic form. Requests for non-existing information ranked as the second most frequent playful utterance type (Shani et al., 2022). While rhetorical questions and non-existing information requests are not directly comparable, both capture utterances that share a key pragmatic feature—neither anticipate genuine answers. This functional similarity suggests users frequently direct unanswerable queries at AI as a playful interaction strategy, whether framed as philosophical ("Why are we here?") or impossible ("Will I win the lottery?").

Joke requests explicitly seeking pre-programmed humor are found with varying frequency patterns across datasets. They appear more often in diary-collected data but rank near the bottom in the public comparison dataset (Lopatovska, 2020; Lopatovska et al., 2020), likely indicating differences in data collection methods rather than differences in the requests people make. Shani et al.'s (2021) taxonomy does not include joke requests as a distinct category.

Request type frequency patterns reveal a notable disconnect between how often certain requests are used and the satisfaction level of the AI responses. Personality questions rank near the top across datasets yet received the lowest humor ratings (average 1.81 out of 5), as Lopatovska et al.'s findings about vague, deflective response strategies suggest, while less frequent categories like joke requests performed better (average 2.3-2.5 out of 5). This inverse relationship may partly reflect the holistic nature of user evaluation. Shani et al. (2021) found that users evaluate humorous interactions based on their overall experience with the system rather than the quality of isolated responses. Satisfaction depends on multiple factors including text-to-speech quality, accent recognition, interface modality, and response timing. Even shallow personification questions (e.g., "How are you?" "Who created you?") can produce successful outcomes when the broader interaction experience is positive, with some users reporting they feel like they're "talking with a friend". Beyond holistic evaluation, several other factors could explain continued use despite poor performance: people may continue attempting request types because these utterances serve functions beyond entertainment seeking—boundary exploration, relationship testing—or users may repeatedly attempt requests hoping for improved responses. Alternatively, the discrepancy may indicate that the datasets used for frequency analysis (semi-naturalistic interactions) and humor

appreciation studies (rating pre-selected utterances with real-world AI assistants) capture different aspects of the phenomenon and are not directly comparable.

4.3 Linguistic cueing of playfulness

After characterizing the types, functions, and frequency of humorous utterances, we now examine a critical question: how do people linguistically signal playful versus serious intent in their interactions with conversational AI? Unfortunately, the six studies reviewed provide only a partial answer to this question (RQ1.1).

The available examples show limited evidence of language use that distinguishes playful from serious utterances. The clearest structural pattern emerges from personality questions, which consistently employ a YOU + verb structure. Explicit frame markers appear in some contexts, such as "just kidding," though the prevalence and variety of such markers remain unexplored. Direct requests for humorous interactions are documented with transparent imperative structures ("Tell me a joke" or "Make me laugh"). Beyond these basic patterns, however, the studies reviewed do not systematically report or analyze how users signal "this is play".

The findings show that humorous remarks directed at AI cluster around boundary testing (impossible requests, absurd quantities) and relationship exploration (personality questions, cultural references). Notably absent are the more subtle comments—such as irony and understatement—that may be more common in human-human interaction, where conversational partners can infer unstated meanings and respond appropriately. Since the available evidence does not include spontaneous remarks from naturalistic contexts, we cannot fully answer RQ1 regarding how user language emerges in humor in interaction with AI.

This gap may stem from methodological limitations. Studies examining playfulness do so post-hoc, categorizing utterances based on semantic content rather than analyzing the linguistic and paralinguistic features users employ to signal playful intent during interaction. The reliance on written representations of spoken interactions—diaries, transcripts, user-written scenarios—obscures the non-verbally expressed elements involved in humor cueing. Only the study on laughter captures the acoustic properties of interactions, but even here the research analyzes responses to humor rather than the strategies people use to initiate or respond with humor.

The gap in evidence on language use and humor has practical implications: if systems cannot reliably detect when users intend to be humorous, they will continue to treat playful utterances as sincere requests, generating inappropriate responses that trigger user laughter at pragmatic incongruity similar to what Perkins Booker et al. (2024) termed "discourse atypicalities".

Moreover, Luger & Sellen (2016) identify a design paradox: while playful interactions (non-serious or non-work tasks) and humor serve as entry points that draw users into exploring system

capabilities, they simultaneously create unrealistic expectations about system intelligence. Anthropomorphic design choices complicate the mental models people construct based on their prior experiences with conversational AI—only technically skilled users see beyond the AI's human-like qualities to form realistic expectations. For most users, humor and "Easter eggs" (discovering hidden content or ingroup references) suggest interaction possibilities beyond what systems can deliver, resulting in disappointment or confusion when expectations go unmet. This creates a fundamental tension between designing engaging, personable assistants and managing realistic user expectations.

In summary, while the studies reveal diverse forms of playfulness in human-AI interaction (RQ1)—from boundary testing to relationship exploration—they provide limited systematic evidence on the linguistic cueing strategies users employ to signal playful intent (RQ1.1). A key limitation is the predominant focus on intentional humor. Study designs that incentivize deliberate humorous interaction fail to capture emergent humor that arises naturally from interactional context (Lopatovska, 2020). Future research should prioritize observational studies and analysis of interaction logs from natural settings, using fine-grained linguistic analysis to understand how both intentional and spontaneous humor are co-constructed in these interactions.

4.4 Humor theory mapping: organizational labels or explanatory frameworks?

As early explorations of humor in human-AI assistant interaction, two studies (Lopatovska et al., 2020; Shani et al., 2021) explicitly relate their findings to the three traditional theories of humor: superiority theory, relief theory, and incongruity theory. While this appears to ground the analysis in established theoretical frameworks, closer examination reveals that the mappings function primarily as organizational labels rather than explanatory models.

The approaches differ in their starting points. Lopatovska et al. (2020) identified utterance types inductively from the data—such as personality questions, joke requests, and rhetorical statements—which were then mapped post-hoc to humor theories, generally falling under relief and superiority theories. However, because the study focused on intentional humorous utterances, it could not fully capture incongruity humor that emerges from unexpected interactions, such as when system responses fail to match user expectations.

Shani et al. (2021), in contrast, organized playful utterances according to the three humor theories from the outset, though alignment between their categorization schemes and traditional humor theory remains questionable. For instance, Shani et al.'s mapping of relief humor onto embarrassment and taboos diverges from the generally accepted description of relief theory as based on the "venting of nervous energy" (Morreall, 1983). Similarly, their incongruity category, defined as containing "elements of surprise" and operationalized as requests that are "impossible in

general or impossible/improbable for virtual assistants," veers away from generally accepted explanations. While impossibility may generate surprise, traditional incongruity theory emphasizes cognitive surprise from violated expectations rather than impossibility per se.

The alignment between these operational definitions and the theoretical constructs remains underexplored, raising questions about whether the theoretical categories illuminate underlying humor mechanisms or simply provide familiar labels for observed patterns.

Neither study provides specific operational definitions of the key term "humor," though in the original study by Lopatovska (2020) "humorous interactions" was broadly defined as "IPA interaction participant found funny". Reading between the lines of Shani et al.'s discussion, "playful" appears to mean non-serious, non-literal utterances that can co-exist with task-oriented contexts, but this working definition remains implicit. Lopatovska et al. (2020) cite Raphaelson-West (1989) to note that humor is universal, culture-specific, and language-specific, designed to make technology "more engaging and personable," but do not specify how these characteristics informed their categorization process.

Shani et al. acknowledge that their data could exemplify multiple humor types, especially as technology evolves, blurring the line between impossible and possible. This recognition suggests they selected humor theory as an organizational tool rather than an explanatory framework. The value of these studies lies not in advancing humor theory but in documenting the rich variety of playful utterances users direct at AI assistants. Future research would benefit from either articulating how humor theories explain interaction patterns or adopting atheoretical classification systems that focus on functional or linguistic characteristics rather than invoking theoretical frameworks that may not apply.

4.5 Laughter in interactions with voice activated AI

Laughter in human-AI interaction reveals how technology users pragmatically interpret conversational breakdowns. People laugh less when interacting with AI, averaging 3 laughs per 10 minutes compared to 5.8-45 in human-human conversations—and their laughter has different acoustic properties: shorter, quieter, and more often unvoiced (Perkins Booker et al., 2024). Only 57% of participants laughed at all during AI interactions.

The contexts that trigger laughter reveal that users apply different standards to human-like versus machine-like errors. People laughed at what the authors term "discourse atypicalities"—conversational failures that mirror human social violations, including inappropriate content, overly human-like statements, and abrupt topic changes. In contrast, they did not laugh at machine-specific errors such as text-to-speech glitches, timing problems, or unnatural pronunciation. This selective response indicates users apply human social scripts (laughing to save face or alleviate social

discomfort) when AI violates conversational norms but recognize and accept technical failures as known limitations of synthetic speech.

This pragmatic hierarchy, combined with the reduced frequency and altered acoustic properties of laughter, suggests users recognize the non-reciprocal nature of AI interaction while still applying human conversational repair strategies to social violations. People don't treat AI failures equally: they're more forgiving of "robot mistakes" than "human mistakes" from the same device.

4.6 Generalizability to Current AI Technology

A critical question for this systematic review is the extent to which findings remain relevant as technology evolves. The studies providing linguistic examples of human playfulness with conversational AI examined previous-generation voice assistants prior to the availability of large language model-based systems. This raises an important consideration: do insights from earlier technology transfer to contemporary AI?

The answer appears to be mixed but generally favorable. While findings related to specific technical capabilities of older systems may not transfer directly, insights about how user expectations, personality traits, and prior experience shape humorous interactions likely remain relevant with current-generation technology. LLM-based systems converse in more human-like language, which could suggest improved humor capabilities. However, recent evidence indicates otherwise. According to Barattieri di San Pietro et al. (2023), these systems still display persistent pragmatic weaknesses specifically in humor. For example, ChatGPT scores lower than humans on humor comprehension tasks and has difficulty handling situated references, underscoring that advances in linguistic sophistication do not automatically translate to improved performance with humor.

This suggests that the fundamental challenges identified in this review—understanding how users signal playful intent, managing expectations created by anthropomorphic design, and satisfying diverse individual preferences for humor—remain relevant design and research priorities even as conversational AI technology continues to advance.

5. Conclusions: Systematic review

A systematic literature review was conducted to consolidate evidence on the linguistic aspects of humor in human-AI interaction, examining the utterances people make rather than machine-generated responses. The review reveals several key findings: users employ humor primarily for system exploration and relationship testing through distinctive linguistic strategies, most notably personification questions using YOU + verb structures and rhetorical questions. However, individual preferences for humor remain highly divided, with the overall quality of the interaction mattering more than isolated playful moments. Existing studies predominantly capture intentional

humor attempts, leaving spontaneous, emergent humor and multi-party interactions largely unexamined.

5.1 Key linguistic patterns

Several linguistic patterns characterize how people signal humorous intent when interacting with conversational AI in English. Exploring system capabilities through outrageous or impossible requests emerged as a primary strategy, serving to probe conversational limitations while learning how to communicate effectively with AI. Sarcasm appeared particularly when interactions failed, with some users perceiving the system replies as intentionally sarcastic responses. Self-deprecating remarks included familiar repair markers like "just kidding" followed by laughter to frame utterances as play or playfully serious.

The clearest structural cue for playful intent was the you + verb construction in personality questions ("Do YOU have a soul?"), which were the most frequent type of humorous questions. Rhetorical questions also frequently led to humorous interactions, similar in frequency to personality questions. Interestingly, joke requests—while producing the most satisfying responses—appeared less frequently overall than personality questions, suggesting humor and playfulness emerge in human-AI interactions when people focus their attention on the AI device itself rather than when attempting task-based interactions.

5.2 Gaps in current understanding

Several significant gaps limit our understanding of how humor functions in human-AI interactions. These gaps fall into two primary categories: methodological constraints and research scope limitations.

5.2.1 The spontaneous humor gap: Methodological limitations

The most notable methodological gap is the exclusive focus on intentional versus spontaneous humor. The studies reviewed were designed to capture and analyze deliberately playful humorous interactions rather than humor arising from unexpected AI responses. Multiple studies explicitly acknowledge these limitations. Lopatovska et al. (2020) note that their procedures "incentivized users to engage in humorous interactions" and state "we would expect more unintentional/incongruent humor to occur with systems responding unexpectedly to user questions or action requests." Shani et al. (2021) highlight that subcategories for incongruity humor are "moving targets" that change continually with technological advances—what was once impossible becomes possible, shifting the boundaries of playful requests.

Perkins Booker et al. (2024) provide indirect evidence for this gap: users laughed most at non-typical discourse that was either inappropriate for the context, too-human, or abruptly changed

topics, offering a view into the humor that emerges from specific types of unexpected responses rather than pre-programmed jokes. Yet no studies systematically captured or analyzed these unintentional humorous moments, leaving a substantial portion of the humor landscape unexamined.

An equally significant limitation lies in the use of transcriptions rather than recordings. While understandable given the practical challenges of recording naturalistic interactions in homes and public spaces, particularly on a large scale, working from written language prevented researchers from accessing prosodic, visual, or contextual cues that might help explain why participants found interactions amusing when the source of humor was not apparent in the verbal content itself.

5.2.2 The social context gap: Research scope limitations

In addition to the spontaneous humor gap, research scope has been narrowly constrained in two important contexts: language and social context. The evidence to date has examined playful language in English, with no studies investigating how humor emerges across languages or cultural settings where conversational norms and humor conventions may differ substantially.

Since humor is fundamentally a social phenomenon, the limited analysis of group interactions in this research is equally significant. None of the six studies examined multi-party interactions or socially situated uses of conversational AI. The studies focused exclusively on individual user-system dyads, either in laboratory settings, through individual diary reports, or via one-on-one conversations between a person and a socialbot (voice-AI interfaces mimicking human conversation in Perkins Booker et al., 2024). The only mentions of others being present during humorous interactions were incidental: families using voice assistants for entertainment with children present, and one study asking participants to write dialogues that would "entertain friends". However, no analysis examined the interactional dynamics when others are present—whether they laugh, add to jokes, compete for turns, or influence the primary user's humor attempts.

One participant in Luger and Sellen (2016) mentioned being less likely to speak conversationally with Siri "on the street" versus "in private," hinting that social context shapes use and perhaps the use of humor, but this was not investigated further. Key questions remain unaddressed: Does humor differ when users are alone or with others? Do people encourage or inhibit each other's playful interactions with AI? How does potential social embarrassment affect humor attempts in public settings? Do different social configurations (family, friends, colleagues) produce different types of humor? The focus on isolated individuals means these inherently social dimensions remain invisible in current research.

5.3 Future research directions

The gaps identified above point to clear priorities for future research. Most importantly, investigations should be based on methodological approaches that capture humor as it naturally emerges rather than as it is deliberately performed.

To address the **spontaneous humor gap**, research should prioritize observational studies of conversational technology in homes and everyday contexts, capturing both intentional and emergent humor without prompting for humor. Video-based methods would allow for the analysis of prosodic cues, facial expressions, and body orientation that signal playful intent—features lost in transcription-based approaches. Longitudinal designs could track how humor use evolves as users become familiar with system capabilities and limitations.

To address the **social context gap**, multi-party interaction studies are essential. Research examining how families, friends, or colleagues collaboratively engage with AI humor would provide valuable evidence on the inherently social dimensions currently invisible in the research reviewed. Comparative designs investigating humor across different social configurations (alone versus with others, private versus public settings, different relationship types) and with different conversational technologies would add valuable information about how technological affordances and social context shapes humorous interaction.

Finally, **cross-linguistic and cross-cultural research** is needed to understand whether patterns observed in English-language studies hold across languages and cultural contexts where conversational norms, humor conventions, and adoption of technology vary considerably.

Despite the many valuable insights from existing studies, addressing these gaps would push forward our understanding of language and humor in interaction with AI, moving beyond deliberate performance to capture how humor naturally functions as a site where humans negotiate the boundaries between artificial and human intelligence.

5.4 Contributions

This review makes three primary contributions to understanding humor in human-AI communication.

First, it provides a human-centered perspective that complements existing research on AI-generated humor. While prior studies have extensively examined user preferences for machine-generated jokes and witty responses, this review focuses on what people do linguistically when initiating humor themselves, revealing patterns that illuminate user motivations, expectations, and adaptive strategies.

Second, it synthesizes key empirical findings on the language and humor strategies people use with conversational AI. Playfulness and humor are generally used to explore system capabilities and test

the relationship rather than simple entertainment, with personification questions and rhetorical queries serving as primary strategies. Technology users demonstrate sophisticated pragmatic processing, applying human social scripts selectively—laughing at human-like social violations but not at machine errors—while maintaining awareness of AI's non-human nature. Yet persistent gaps exist between what users attempt (frequent humor strategies despite low satisfaction), what they want (divided preferences, many preferring no humor), and what systems deliver (pre-programmed content succeeds, conversational wit fails).

Third, it reveals critical methodological gaps that define an agenda for future research: the absence of naturally-occurring spontaneous humor, the lack of socially-situated multi-party studies, and underdeveloped theoretical frameworks for human-machine humor contexts. These gaps point toward fundamental questions about how humor functions when one interlocutor lacks consciousness and social standing—issues that challenge classical humor theories.

5.5 Limitations

This systematic review has several limitations that should be acknowledged.

Search scope: The review examined six studies (from 490 originally identified) across four academic databases (arXiv, IEEE Xplore, Scopus, and WoS). While efforts were made to be comprehensive, including manual searches of reference lists, other relevant studies published in venues outside these databases or in grey literature may have been missed. Additionally, all included studies examined English-language interactions, limiting generalizability to other linguistic and cultural contexts.

Technological evolution: All studies reviewed were based on previous-generation AI voice assistants rather than LLM-based systems. Given that LLM technology has only been widely available for three years (at the time of writing), research examining different dimensions of humor with LLM-based conversational AI is just beginning to emerge in academic literature. While findings about user expectations and pragmatic processing are likely to remain informative across technology generations, LLM-based systems offer substantially different conversational affordances—particularly extended multi-turn conversations and access to user history and contextual information—and therefore warrant further investigation as more evidence is published.

Inclusion criteria: The focus on studies reporting linguistic evidence of user-initiated humor necessarily excluded research examining AI-generated humor, user preferences for humorous responses, and broader user experience studies that mentioned humor peripherally. While this focus enabled depth of analysis, it limits the review's coverage of the full humor ecosystem in human-AI interaction.

6. Chapter summary

This chapter has presented a systematic review of humor in spoken human-AI interactions, following PRISMA 2020 guidelines. The review examined how people initiate humor, recognize humorous potential in AI responses, and adapt their communicative strategies in voice-based exchanges.

Key findings reveal users employ humor primarily for system exploration and relationship testing through distinctive linguistic patterns such as personification questions and rhetorical queries. However, a significant gap emerged: existing studies capture varieties of intentional humor but leave spontaneous, emergent humor and multi-party interactions largely unexamined.

These gaps establish the empirical foundation for the studies presented in the following chapters, which investigate spontaneous interactional humor through analysis of naturally occurring interactions and performative social media commentary—dimensions that recent developments in AI technology and user-generated content have made more easily accessible for research. The next chapter details the methodological approach designed to address these unexamined aspects of humor in human-AI interaction.

Chapter 3
Methodology

1. Overview

This chapter outlines the data collection, organization, and analytical procedures for three complementary studies investigating how humor emerges in human-AI interactions. Together, these studies address the research questions established in Chapter 1 by examining how speakers of English, Italian, and Spanish navigate humorous moments with conversational AI systems across different interactional contexts.

The three studies employ distinct but complementary methodological approaches to capture different facets of spontaneous humor. Study 1 investigates the social dynamics of laughter in interactions with AI assistants from user generated videos on social media platforms, examining who laughs, when laughter occurs sequentially, and what becomes the target of humor in these spontaneous exchanges. Study 2 analyzes the linguistic and pragmatic features of playful interactions with AI assistants, identifying the linguistic forms that signal attempts at building common ground, and examining meta-pragmatic commentary that reflects on the interaction itself. Study 3 examines human-AI communication as the subject of humorous social media posts, revealing what aspects of AI language use and communicative behavior emerge from this public critique and metalinguistic commentary.

Each study draws on video or social media data collected across multiple languages and platforms, allowing for cross-linguistic comparison and capturing humor as it emerges in both live interactions and reflective discourse. The chapter is organized into three main sections corresponding to each study. Each section details the data collection procedures, data preparation methods, and analysis procedures, including the analytical frameworks used (such as Chafe's concept of non-seriousness and Bateson's concept of meta-communication signaling play), and coding schemes. The chapter concludes by discussing ethical considerations, quality assurance procedures, and the rationale for the multi-study design that allows for a broad examination of spontaneous humor across diverse human-AI communicative contexts.

2. Description of Studies

2.1 Study 1: Who participates in laughter: Social dynamics of humor

This study addresses RQ 1.2 (Who/what is targeted in humorous moments) and contributes to RQ 1.3 (cross-linguistic patterns). Starting from the observation that many English-language videos on social media platforms depict older adults struggling with voice assistant technology prompted me to ask whether videos of similar interactions in other languages would show the same dynamics or instead highlight different aspects of the interaction. Seeing younger family members laughing at (or with?) older users unable to communicate with technology seems to be an evident case of laughter in response to someone else's imperfections or misfortune, which is generally considered to be the base of the superiority theory of humor.[9] Examining who participates in the laughter and when it occurs sequentially, I document specific aspects of the interactional social dynamics beneath what superficially appears to be mockery: who was actually laughing at whom. Additionally, voice assistant interactions may involve a form of shared pretense between the people in the conversation similar to the pretend play that children develop around the age of two (Leslie, 2001, discussing earlier work from 1987, 1994). People talking to a voice assistant are implicitly suspending disbelief to address a non-human entity using human communicative practices while simultaneously recognizing that others present are also engaged in this pretense (see Theory of Mind in chapter 1). Therefore, humor arising within the interactions may not stem from failure but from the underlying tension of maintaining this collaborative pretense that machines can communicate like humans. They are treating the device as a conversational partner (in line with the CASA paradigm outlined in chapter 2), but when the veil falls a cognitive shift occurs with the

[9] The superiority theory is considered one of the three traditional theories of humor along with the incongruity theory and the relief theory. Attributed to the writings of Plato, Aristotle, and Hobbes, the superiority theory describes the feeling of "enjoyment of others' suffering" (Morreall, 1983) or the "comic amusement" leading to an "enjoyable feeling of superiority to the object of amusement" (Lintott, 2016)

realization they are talking to a machine and 'this is play' (Bateson, 1972; 1987) not bona-fide communication (Raskin, 1985 as cited in Attardo, 2017). The English search terms used to construct the video dataset reflect a game-like scenario with competitive framing ("Alexa vs. Grandma"), which brings play into the foreground of the discussion.

An inductive thematic analysis was carried out on 117 video-recorded interactions between groups of adults and AI voice assistants, examining who participates in laughter (the AI user, or the group) and when laughter occurs sequentially. This analytical approach allows contextual themes related to the types of requests directed at the AI to be identified, organized, analyzed and described in detail (Braun and Clarke, 2006). The interactional and contextual patterns that emerge from the videos (Stebbins, 2011) form the base of the analysis of the social dynamics in these humorous moments. The overall aim of this study is to determine whether the scenes reflect shared amusement (a less hostile form of social negotiation) or mockery of individual failure, which would point towards superiority humor.

Data source (2016-2023): Multilingual corpus of 117 YouTube and TikTok videos featuring people interacting with AI voice assistants in English, Italian, and Spanish, systematically collected using specific search strings in each language.

2.2 Study 2: Linguistic forms of spontaneous playfulness

This study addresses RQ 1.1 (How is playfulness linguistically cued?) and RQ 1.3 (cross-linguistic patterns). An exploratory analysis was carried out on 40 video-recorded interactions between pairs of participants and an AI voice assistant (Alexa) in English and Italian, examining the language and pragmatic features of spontaneous playful utterances directed at the AI. Drawing on Chafe's (2007) concept of "non-seriousness" and Bateson's concept of meta-communication (1972;1987), this approach identifies specific linguistic cues that signal playfulness (such as irony, sarcasm, teasing, hyperbole, and rhetorical questions) as well as meta-commentary that reflects on AI's communicative abilities or the ease (difficulty) of the interaction itself.

Data source (2023-2024): Purpose-built video corpus of 40 semi-structured interactions with Alexa in English and Italian, capturing both linguistic and non-linguistic elements of the exchanges in semi-naturalistic settings.

2.3 Study 3: Humorous critique in social media discourse

This study addresses RQ 2 (What aspects of human-AI communication are targeted in humorous social media posts?). Following social media discourse about AI, I started noticing a shift when ChatGPT arrived. While posts about AI assistants like Alexa frequently focused on comprehension failures and speech recognition issues, posts about AI systems like ChatGPT demonstrated greater attention to the language itself—its distinctive stylistic features, its accommodating tone, and its

characteristic patterns of expression. To investigate which communicative features were targeted in spontaneous internet humor as defined by Yus (2023), a metalinguistic analysis was carried out on 99 user-generated social media posts in English and Italian, systematically examining what aspects of AI language use and communicative behavior merit humorous commentary and cultural critique. This approach reveals how people collectively make sense of AI communication through spontaneous humor created online, what aspects of human-AI interaction become culturally salient enough to inspire shared content, and how these posts construct social meaning around relationships with conversational AI systems.

Data source (2023-2025): Social media corpus of 179 posts from Facebook user groups, Instagram, and TikTok that humorously depict or describe voice interactions with AI in English and Italian, including comment threads. Post selection was based on explicit linguistic or paralinguistic (emojis and hashtags) markers of humorous intent.

3. Process and procedures

3.1 Study one

3.1.1 Data collection

A systematic approach was employed to search for user-generated video content using specific search strings in English, Italian and Spanish on YouTube and TikTok platforms. Data collection was conducted in two phases: October 2022 and June 2023. The English search string was created using the title of a YouTube video, "*Alexa vs Grandma*", which was then translated into "*Nonnina e Alexa"* for Italian and "*Abuelita en Alexa*" for Spanish. All videos that clearly showed human-AI interaction were downloaded and organized chronologically by upload date. The search was terminated when three consecutive video recommendations were duplicates of those already downloaded. An important caveat is that computer location and prior viewing habits can affect which videos appear, meaning the underlying platform algorithm recommending the videos remains unknown.

The resulting corpus consists of 117 videos uploaded between 2016-2023, with the majority found on TikTok (n=76) compared to YouTube (n=47). Over 70% of the videos were uploaded during the final three years of the sampling period (2020-2023). The trilingual distribution includes 47 videos in the English group, 34 in the Italian group, and 36 in the Spanish group, though each group contains at least one video featuring additional languages. Temporally, English videos appeared earliest (2016), while Italian and Spanish content appeared in 2018, reflecting the gradual rollout of Alexa technology: available in the US from late 2014, in Spanish-speaking Central American countries by late 2017, and in Italy, Spain, and Mexico by October 2018.

3.1.2 Data security and ethical considerations

All videos were manually downloaded from publicly available user profiles in accordance with community guidelines. No identifying information about the users was retained, though usernames may remain embedded within video hyperlinks. All downloaded content is password-protected and archived through an institutional account to ensure data security and comply with ethical guidelines.

3.1.3 Data preparation

Following data collection, videos were organized into three distinct corpora: AVG-English, NEA-Italian, and ABA-Spanish, stored in separate folders. Each video was assigned a unique identifier based on the corpus code and sequential number (e.g., AVG-001, NEA-001, ABA-001).

Initial familiarization involved multiple viewings of all videos with note-taking to identify patterns in the content. A systematic coding framework was then developed using Microsoft Forms to capture video characteristics including: upload date, video length, original hyper-link, platform, AI assistant type, languages used, task or request type, presence of interaction scaffolding (assistance from others), interaction outcome and responsibility for failure (task fulfilled, person responsible for failure, or voice assistant responsible for failure), first instance of laughter (primary user or observers) and whether the primary user laughed or smiled during the interaction.

The codes for task interaction types were based on an adapted version of Lopatovska's (2020) classification of humorous interactions, with the addition of "information about self or person's life" as a request category. Each video was coded according to the primary task or request directed at the voice assistant during the interaction.

After initial coding, nine videos from the full corpus (n=117) were excluded from thematic analysis: 2 Spanish parodies, 4 Italian informational or staged comedy videos, and 3 English informational or staged comedy videos. These exclusions ensured the analysis focused on user-generated documentation of spontaneous interactions rather than scripted performances or promotional content. The final analytical sample consisted of 108 videos (English n=44, Italian n=30, Spanish n=34).

3.1.4 Analysis procedures

Phase 1: Familiarization with the data. Following initial viewings during data preparation, each video was reviewed again with specific attention to moments of laughter and the surrounding interactional context. Detailed notes documented who laughed, when laughter occurred in relation to the AI interaction, and any verbal commentary accompanying or following laughter.

Phase 2: Generating initial codes. Using the systematic coding framework established during data preparation, each instance of laughter was coded for: (a) who initiated laughter (primary user, observer, or simultaneous), (b) whether others joined in the laughter, (c) whether the primary user

was autonomously interacting with the AI or receiving scaffolding/assistance from others, (d) the sequential placement of laughter relative to the AI's response or lack of response, and (e) any verbal meta-commentary that suggested what was being found humorous (e.g., comments about the AI's response, self-deprecating remarks, observations about the situation).

Phase 3: Searching for themes. Initial codes were examined to identify patterns in laughter participation across different contexts. Particular attention was paid to whether laughter patterns varied based on: the source of communication failure (technology vs. user error), the presence of interaction scaffolding (whether someone was assisting the primary user or if the user was autonomously interacting with the AI), the presence of multiple participants/observers, the type of task being attempted, and any cultural or linguistic differences across the three corpora.

Phase 4: Reviewing themes. Candidate themes were reviewed against the coded data and the research questions, specifically examining whether patterns of laughter participation revealed instances of shared amusement versus mockery of individual failure. Videos were re-watched to ensure themes accurately represented the data and that distinctions between theme categories were clear and consistent.

Phase 5: Defining and naming themes. Final themes were defined based on what patterns of laughter participation revealed about the social dynamics of humor in these interactions. Each theme was named to capture both the laughter participation pattern and what it suggested about humor targets (e.g., whether humor was directed at the AI system, the user, the situation, or shared playfully among participants).

Phase 6: Producing the report. Themes were organized to address RQ 1.2 (who/what becomes the target of humor) and RQ 1.3 (cross-linguistic patterns). Representative examples from each corpus were selected to illustrate themes, with attention to including diverse instances across languages, platforms, and interaction types.

Cross-linguistic comparison

Following the identification of themes within each corpus, patterns were compared across the three language groups (English, Italian, Spanish) to identify similarities and differences in laughter dynamics and autonomy of the participant. This comparative analysis addressed whether cultural or linguistic factors influenced the social dynamics of humor in human-AI interactions.

3.1.5 Quality Measures

Several measures were employed to ensure the rigor and trustworthiness of the analysis (Nowell et al., 2017). Credibility was established through multiple strategies: prolonged engagement with the dataset through iterative viewing and coding phases, systematic preservation of original video content enabling repeated return to source material, and multilingual analysis capability that

allowed direct interpretation of Italian and Spanish videos without translation-mediated distortion. The six-phase thematic analysis process incorporated constant comparison both within and across language groups, allowing emerging interpretations to be tested against the full dataset.

Dependability was maintained through detailed documentation of analytical procedures. The Microsoft Forms coding framework provided a systematic record linking each video to its coded characteristics (laughter participation patterns, task type, interaction scaffolding, communication failure source). The coding schema offered a consistent analytical framework applied across all 117 videos, with clear operational definitions for each category.

Confirmability was addressed through data-grounded interpretation, with all thematic claims traceable to specific videos in the archived dataset. The analysis prioritized observable behavioral patterns—who laughed, when, and in what sequential context—alongside participants' verbal meta-commentary when present, rather than imposing external interpretive frameworks. Multiple exemplars were identified for each theme to ensure interpretations reflected patterns rather than isolated instances.

Reflexivity regarding researcher positionality was maintained throughout analysis. As a multilingual researcher with regular experience using AI voice assistants, I brought both insider understanding of the communicative challenges and cultural contexts represented in the videos, and awareness of potential biases in interpreting laughter as shared amusement versus mockery. Regular analytical memos documented evolving interpretations and moments where personal assumptions about AI interaction might influence interpretation of humor dynamics.

Transferability is supported through rich description of the dataset characteristics (platforms, date range, languages, AI assistants, interaction types) and clear boundary conditions (YouTube and TikTok videos, 2016-2023, English/Italian/Spanish, group interactions with visible laughter), enabling readers to assess applicability to other contexts of spontaneous humor in human-AI interaction.

3.2 Study two

3.2.1 Data Collection

Participants

Through convenience sampling, eighty-two adult participants were recruited for this study from June 2023 to May 2024. Two Italian participants were excluded after initially agreeing to participate but subsequently declining to sign consent forms, resulting in twenty participant pairs for each language group. The participants self-selected their partners before arriving, which generally consisted of their own family members, colleagues, classmates or friends. The demographic characteristics of each language group are distributed as follows: The English participants (45%

male and 55% female) ranged in age from 18 to 74 years (n = 40). The largest proportion (35%) were between 18 and 24 years old, with an estimated mean age of approximately 38.5 years. The median age range was 45–54 years.

The Italian participants (52% male and 48% female) ranged in age from 18 to 84 years (n = 40). The largest proportion (25%) were between 18 and 24 years old, with an estimated mean age of approximately 42.9 years. The median age range was 35–44 years.

Participants were not required to own or have previously used voice assistants. In the English-speaking sample, 13% had not used voice assistants in the past twelve months, compared to 35% percent of Italian speakers. However, usage patterns were similar between groups, with 38% percent of participants in each language group reporting daily or weekly use of voice assistants.

Task design

Task suggestions were developed by comparing popular Alexa command lists, usage reports, and skill rankings (Amazon, 2019). Fifteen tasks were selected based on potential for generating humorous interactions as identified in the literature review (Chapter 1), simplicity, ease of use, and cross-cultural appropriateness for participants in Italy, Ireland, and the US. Tasks unsuitable for the office setting (e.g., controlling lights or doors) or culturally unfamiliar (e.g., garage door controls common in the US but rare in Italy/Ireland) were excluded.

Task suggestions were presented as imperatives requiring participants to formulate their own requests using extemporaneous language, while including necessary keywords for successful interaction. Meaning comparability between English and Italian versions was prioritized during translation. The initial task list and presentation format were validated and refined through a pilot study before finalizing the fifteen tasks used in the main study. The final list of task suggestions was reviewed by a native speaker for each language for naturalness and contextual appropriateness (see Appendix 2A for complete list).

Example:

- (English) Ask Alexa to tell you a story. You can tell her to stop the story if you have heard enough.
- (Italian) Chiedi ad Alexa di raccontarti una storia. Puoi fermarla se non vuoi sentire proprio tutto.

Setting

A quiet campus office was initially designated for the observational study. However, due to building access limitations during times participants were available, data collection was conducted in various convenient locations including offices, workplaces, coffee shops, and homes. All locations required a table and chairs, sufficient quiet for voice interactions with minimal ambient

noise (to ensure accurate device recognition and clear video recording), and reliable internet access via WiFi. When WiFi protocols prevented Alexa connectivity, internet was provided through USB wireless adapter hotspot or cellular hotspot, though slower connections could potentially affect task performance, particularly for music streaming or translation requests.

Equipment

Recording equipment consisted of two devices to capture participants' faces and upper bodies during interactions: an iPad (9th generation) and either an iPhone 13 or iPhone 15 mounted on a tabletop tripod. The AI voice assistant was an Amazon Echo Show (2nd generation) with 20 cm screen and portable battery base for location flexibility.

Additional materials included: a Bluetooth stylus to sign digital consent forms, task suggestion cards (English or Italian), evaluation response adjective cards (English or Italian), and informed consent/privacy documentation in both electronic and paper formats (English or Italian)

Session procedures

Each session followed a structured protocol. Participants were greeted and thanked for their participation, then briefed on how the data collected would be utilized and archived, and their right to withdraw at any point. Following consent procedures, participants were introduced to the Amazon Echo Show and given instructions for proper communication with the device, specifically the need to say "Alexa" and wait for the audio signal before speaking. A brief practice period allowed participants to familiarize themselves with the device while any questions were addressed. Participants then received fifteen task suggestion cards, presented in consistent order across all sessions. They were instructed that the first prompt would initiate the 10-minute timer and must be used first, with the remaining cards serving to facilitate the interaction given that prior experience with AI assistants was not required.

Participant pairs were allowed to self-organize turn taking and speaking to Alexa, engaging in information requests, action commands, translation requests, and personality tests until the timer signaled the end of the session. The researcher remained present, as far from the participants as the room allowed, taking field notes throughout. However, the researcher would intervene if participants requested help with Alexa's functionality or indicated excessive frustration that might terminate the interaction. In these circumstances, the researcher adopted a role analogous to the "confederate speaker" in a similar study by Siegert and Krüger (2018, p. 4), discussing possible interactional strategies with participants to help them achieve desired results from Alexa, while ensuring only participants interacted directly with the voice assistant. Participants were not informed that humor or laughter were focal aspects of the study to ensure natural interaction patterns.

Following the 10-minute session, each participant assessed their interaction with Alexa by selecting three reaction cards to evaluate the overall positive or negative nature of the communication experience, in addition to providing demographic information via QR code or direct link. All sessions were recorded using two devices to capture a comprehensive set of verbal and non-verbal data.

Post interaction evaluation

As a subjective measure of the feelings or emotions evoked during interactions with the AI voice assistant, participants were asked to evaluate their interaction with Alexa after completion (cf Zamora, 2019). Each participant selected three adjectives (not ranked) from a set of twenty-six cards, with each card containing one adjective related to satisfaction, funniness, or usability. The evaluation words were adapted from the Microsoft Desirability Toolkit, created to gather participant feedback by selecting words that reflect their experience (Benedek and Miner 2002, Barnum and Palmer 2010). The adjectives were translated into Italian and both language versions were reviewed by native speakers for naturalness and contextual appropriateness.

3.2.2 Data security and ethical considerations

Prior to participant recruitment, authorization was obtained from the Institutional Review Board (*Comitato di Bioetica dell'Alma Mater Studiorum – Università di Bologna*). Although this type of study would not likely present any type of risk to human participants, all steps of the study were designed according to best practices in behavioral studies with human participants (Ritter et al. 2013, Bethel et al 2010).

All data collected for this study are maintained in password-protected folders in the institutional Microsoft OneDrive provided by the University of Bologna. No identifying information about the users was collected for this study, apart from the videos with visible faces. The Privacy and consent forms are archived separately from the videos so they cannot be connected to a specific participant or session.

3.2.3 Data Preparation

Data organization

For each interaction session in this study there are: two videos, video transcripts, signed consent forms for each participant, demographic information for each participant, photos of the participant interaction evaluation cards and activity transcripts generated by Alexa.

- After each interaction, session videos were moved directly into folders identified by sequential alphabetical codes and date, with single letters used for Italian sessions (e.g., A-2023-06-15) and double letter codes with a Z prefix for English sessions (e.g., ZA-2023-06-15). Each video file was also saved as audio-only to facilitate the transcription process.

- Informed consent and privacy forms were signed electronically by participants and saved in a separate folder for English or Italian signed forms.
- The demographic information collected from each participant was acquired directly online through Microsoft Forms (institutional account), in the majority of cases. In the few cases that the participant preferred filling out a paper form, the information was transferred immediately afterwards to Microsoft Forms and the paper copies destroyed.
- The evaluation words selection by each participant were transferred into and tabulated by Microsoft Forms (either in Italian or English). Photos of the cards selected by the participants are housed in the folder with the video and transcripts.
- After each interaction, the transcripts generated by Alexa were saved into the folder for that interaction. These documents contain key information about the responses given by Alexa, including the web sites or graphics-based shown on-screen during the interaction, that does not appear in the video recordings.
- All field notes were also copied into the session folder.

Transcription procedure

Transcripts of each interaction session were obtained using the audio file and the Revoicer platform, which has both text-to-speech and speech-to-text functionalities. The platform was configured to the appropriate language (English or Italian) for each session's transcription. The automatically generated transcripts were then manually reviewed and edited for accuracy.

Transcription conventions

The final transcripts (see examples in Appendix 2B) employed GAT 2 (Gesprächsanalytisches Transkriptionssystem 2) conventions adapted from Jeffersonian transcription (Couper-Kuhlen and Barth-Weingarten, 2011), modified to capture the pragmatic features relevant to this analysis. Given the study's focus on linguistic forms of playfulness and meta-pragmatic commentary rather than detailed phonetic or prosodic analysis, transcription prioritized:

- Turn-taking patterns and overlapping speech
- Pauses and timing of turns
- Lexical choices and grammatical structures
- Enunciation changes (particularly in reformulations directed at the AI)
- Volume shifts (particularly in reformulations or expressions of frustration)
- Verbal meta-commentary about the interaction

Detailed prosodic features (such as precise pitch movements, fine-grained intonation contours, or voice quality) were not systematically transcribed, as the analytical focus was on identifying pragmatic strategies and linguistic markers of playfulness rather than phonetic characteristics. This

transcription approach aligned with the study's goal of examining how people linguistically navigate spontaneous humorous moments and adapt their communicative strategies when interacting with AI.

All transcripts were reviewed multiple times alongside the video recordings to ensure accuracy of the captured pragmatic features and to verify that relevant interactional details were appropriately represented.

3.2.4 Analysis Procedures

Initial coding approach

Analysis began with an exploratory phase using ELAN software (Max Planck Institute for Psycholinguistics) to code a sample set of interactions. Initial coding tiers were established at the participant level, with the working hypothesis that laughter could serve as a marker to identify humorous and playful language directed at the AI. However, this approach revealed that laughter occurred frequently throughout the interactions—likely for reasons including general amusement at the situation, social bonding between participants, reactions to technical difficulties, and nervous laughter—making it an unreliable indicator for identifying the spontaneous playful utterances directed at AI that this study aimed to examine.

Revised analytical approach

The analysis strategy was therefore revised to focus directly on the transcripts rather than using laughter as a preliminary filter. Each transcript was systematically read to identify segments containing potential instances of linguistic playfulness and meta-pragmatic reflection directed at or about the AI, including:

- Non-literal language forms: irony, sarcasm, hyperbole, understatement, rhetorical questions, or teasing directed at the AI
- Strategic reformulations: repeated or rephrased utterances with marked enunciation or volume changes, signaling adaptation of communicative approach
- Meta-pragmatic commentary: explicit reflections on the AI's communicative abilities, the ease or difficulty of the interaction, or the appropriateness of AI responses
- Reflexive interpretation: participants' comments on their own or others' communicative strategies with the AI, including explanations of why they phrased requests in particular ways or assessments of what "works" or "doesn't work"
- Social norm negotiation: collaborative discussion between participants about how one "should" talk to AI, what constitutes appropriate interaction, or collective meaning-making about the AI's communicative competence

When potentially relevant segments were identified in the transcripts, the corresponding video files were accessed from their folders in the archive described in the previous section and re-watched to confirm the interpretation and to capture any relevant non-verbal context (gestures, facial expressions, participant interactions) that might inform understanding of the pragmatic intent.

Participant evaluation data

The adjectives the participants selected at the end of the session, described in the data collection section, provide subjective information about how participants perceived their interaction experience. However, these evaluations should be interpreted cautiously, as participants were likely influenced by their prior experiences and expectations about AI assistants rather than solely by the specific interaction session—particularly given that some participants had never used any type of AI assistant before while others had extensive experience. This evaluation data is therefore presented descriptively as contextual information about participants' overall attitudes and allows for cross-linguistic comparison between English and Italian speakers but is not employed as an analytical tool for interpreting attitude towards linguistic playfulness.

3.2.5 Analytical framework

Once relevant segments were identified in the transcripts, a multi-layered pragmatic analysis was conducted examining several dimensions of the spontaneous playful and meta-pragmatic language:

Direct address to Alexa - All instances where participants directly addressed Alexa with non-task language were identified and analyzed. This included playful or teasing comments such as "Alexa, tra di noi non può funzionare" (Alexa, between us it won't work) or "Alexa, tu leggi nel cuore delle persone" (Alexa, can you read people's hearts?), as well as evaluative commentary like "allora sei proprio ignorante" (so you're really 'stupid' though locally the meaning is closer to 'stupid bitch') or "Alexa, sei un po' stronza vero?" (Alexa, you're a bit of a bitch, right?). For each instance, the surrounding interactional context was examined to understand what prompted the comment and whether it emerged from communication failure, unexpected responses, or successful exchanges that surprised participants.

Meta-pragmatic commentary about the interaction - Participants' explicit reflections on how Alexa was communicating were systematically identified, including comments on response accuracy "Non risponde mica, non è la risposta..." (She's not answering, that's not the answer), appropriateness of responses "però... non è (..) non è corretto" (but... it's not fair or appropriate behavior), or positive evaluations "che carina" (how sweet of her). These meta-comments revealed participants' expectations for conversational AI behavior and their awareness when those expectations were exceeded or violated.

Metalinguistic observations - Instances where participants reflected on language use and meaning-making were analyzed, such as when a participant noted "ah, perché le ho chiesto il numero, e lei pensa a telefono" (ah, because I asked her for the number and she thought of a phone number), demonstrating awareness of ambiguity in their request. These moments revealed how participants diagnosed communication breakdowns at the linguistic level.

Laughter and participant alignment - The sequential placement of laughter was examined—whether both participants laughed simultaneously, whether one participant laughed at the other's comment to Alexa, or whether laughter followed particular types of responses from Alexa. This pointed towards the social dynamics of humor and whether the playful commentary was collaboratively constructed or individually initiated.

Question construction and repair sequences - Question structures and content types were examined in relation to interaction outcomes (successful responses, failures requiring repair, or unexpected responses). The qualitative analysis focused on how question construction related to the emergence of playful language and meta-commentary, particularly in repair sequences where participants adapted their communicative strategies.

Application of theoretical frameworks - Drawing on Chafe's (2007) concept of "non-seriousness," and Bateson's concept of "meta-communication" (1972;1987) segments were analyzed for markers indicating a shift from task-oriented communication to playful engagement with the AI. Different types of non-literal language directed at Alexa were identified, examining how teasing or other playful forms emerged in response to the pragmatic tensions of human-AI interaction.

3.2.6 Quality Measures

Several measures were employed to ensure the rigor and trustworthiness of the analysis (Nowell et al., 2017). Credibility was established through multiple strategies: prolonged engagement with the dataset through iterative viewing and analysis phases, systematic preservation of original video and audio recordings enabling repeated return to source material, and bilingual analysis capability that allowed direct interpretation of Italian interactions without translation-mediated distortion. The analytical process incorporated constant comparison both within and across language groups, allowing emerging interpretations to be tested against the full dataset of 40 interactions. Dependability was maintained through detailed documentation of analytical procedures. The systematic organization of data (video files, audio files, transcripts, evaluation data, and Alexa-generated transcripts) provided a comprehensive record of each interaction session. The analytical framework outlined above offered a consistent structure applied across all interactions, with clear

operational definitions for each dimension of analysis (direct address, meta-pragmatic commentary, metalinguistic observations, laughter patterns, question construction).

Confirmability was addressed through data-grounded interpretation, with all analytical claims traceable to specific interactions in the archived dataset. The analysis prioritized participants' observable linguistic choices and explicit meta-commentary rather than imposing external interpretive frameworks. Video recordings were consulted alongside transcripts to verify pragmatic intent through non-verbal cues when relevant.

Given the relatively small dataset of 40 interactions, claims about recurring patterns were made cautiously. When multiple instances of similar phenomena were identified across different participants and language groups, this was taken as evidence of patterning; however, the findings are presented as indicative of the kinds of spontaneous playful language and meta-commentary that emerge in human-AI interaction rather than as generalizable frequencies or distributions.

Reflexivity regarding researcher positionality was maintained throughout analysis. As a bilingual researcher with regular experience using AI voice assistants, I brought both insider understanding of the communicative challenges participants faced and awareness of potential biases in interpreting playful language versus genuine frustration. The semi-structured nature of the interactions required particular reflexivity about how the experimental setting might have influenced participants' behavior and language choices. The researcher remained present during sessions in a primarily technical support role—intervening only when participants experienced technical difficulties (such as internet connectivity issues) or requested assistance with Alexa's functionality. In these limited interventions, the researcher adopted a role analogous to Siegert and Krüger's (2018) "confederate speaker," discussing possible interactional strategies while ensuring only participants interacted directly with Alexa. This minimal but necessary presence was documented in field notes, and moments where researcher intervention occurred were flagged during analysis to consider potential influence on subsequent participant behavior.

Transferability is supported through rich description of the dataset characteristics and clear boundary conditions, enabling readers to assess applicability to other contexts of spontaneous humor in human-AI interaction. The 40 interactions (20 English, 20 Italian) were conducted in semi-naturalistic settings during 2023-2024, with adult participants working in pairs with Alexa. Participants were provided with suggestion cards containing imperative phrases to prompt task-oriented questions, though they had to formulate their own questions based on these prompts and were free to ask any additional questions of their choosing. While this semi-structured approach influenced the content themes of task-related questions, the spontaneous playful language directed at Alexa (such as "Alexa, between us it won't work") and meta-pragmatic commentary about the

interaction emerged organically in response to the communication dynamics themselves, not from the provided prompts. The findings regarding these spontaneous utterances are therefore transferable to both structured and naturalistic human-AI interaction contexts. Additional boundary conditions include participants' varying levels of prior AI assistant experience (some had never used voice assistants before) and the researcher's presence in a technical support role. These contextual factors should be considered when assessing the applicability of findings to other settings.

3.3 Study three

3.3.1 Data collection

Data collection for this study was based on a simple approach, leveraging social media recommender algorithms rather than systematic search strategies. Over the period 2024-2025, posts were collected in English and Italian from Facebook, Instagram, and TikTok that directly or indirectly spotlight the use and understanding of language in interaction by voice assistants and AI. Posts were gathered through organic social media engagement, including content that appeared automatically in feeds and posts directly shared by acquaintances aware of the research focus. While collection occurred from 2024-2025, some content was posted in 2023 as it continued to circulate through social media sharing and algorithmic promotion.

A total of 179 posts about AI interactions were collected and stored in a Microsoft Access database. From this corpus, 99 posts were selected for systematic analysis based on the presence of explicit markers of humorous or non-serious intent, including: words in the titles or descriptions signaling playful commentary ("Ridiamo ogni tanto" [Every now and then we laugh]), hashtags indicating comedy or humor (e.g., #comedy #comicità #humor #funny #divertente; see Appendix 3B for complete list), and emojis signaling laughter or playfulness (e.g., 🤣😭😂😆😁😋🤡). The 80 posts not included in this analysis feature human-AI interaction but lack clear humor framing in the main post itself—they are primarily informational, requests for help troubleshooting problems, or straightforward demonstrations of AI functionality. Notably, the comment section on some excluded posts do become platforms for humorous critique, suggesting that metalinguistic commentary through humor extends beyond deliberately humorous posts; however, this study focuses specifically on posts where humor functions as the primary framing device for metalinguistic critique.

The 99 posts analyzed include 72 in English and 27 in Italian, both video (n=66) and text-based formats (n=33, including text-only posts and images with text overlays). Platform affordances are reflected in the format: Facebook posts (n=61) are split roughly evenly between text-based (n=33) and video content (n=28, some shared from other platforms), while Instagram (n=37) and TikTok

(n=1) consist exclusively of video content. Video posts range from scripted comedy skits and parody songs to recordings of actual AI interactions with implicit or explicit metalinguistic commentary, while text-based posts include forum discussions, image memes with text overlays, and written narratives describing AI interactions. Posts feature a range of AI systems, with Alexa and ChatGPT appearing most frequently in roughly equal proportions, followed by less frequent mentions of Siri, GPS navigation systems, and other voice assistants or AI.

3.3.2 Data security and ethical considerations

Data security protocols mirror those used in Studies 1 and 2, with all content stored in password-protected institutional OneDrive folders. Posts were collected from publicly available social media content, with no personal identifying information retained beyond usernames visible in original posts. Data collection complied with platform terms of service, and analysis focuses on linguistic content rather than individual user identification.

Screenshots were modified to obscure usernames and other identifying information while preserving linguistic content relevant to the analysis. Focus remained on discourse themes and patterns rather than individual user identification.

3.3.3 Data preparation

Content from each post was systematically captured and archived through multiple formats to preserve the multimodal nature of social media discourse. Text posts were saved as screenshots, video content was downloaded in full, and available comment threads were captured for all posts. Each post was assigned a sequential folder on institutional OneDrive using date and descriptive title format (e.g., 2025.05.07 Alexa e Siri).

Data organization employed a dual system combining file storage with database cataloging. All posts were entered into a single Microsoft Access database table (stored on institutional OneDrive) with the following fields: sequential ID number, date posted, weblink, language (Italian/English), original post title, original description, narrative structure (yes/no), presence of pre-determined humor markers (yes/no), humor emojis in description or in comments (yes/no), words marking humor title (yes/no), national culture theme (yes/no), platform (Facebook, Instagram, TikTok), AI assistant mentioned (Alexa, Siri, ChatGPT, other AI), notes on initial keywords and themes, and direct link to the corresponding content folder.

This database structure enabled systematic tracking of humor markers and thematic elements while maintaining connections between structured data and original multimedia content for subsequent meta-linguistic and meta-communicative analysis.

3.3.4 Analysis procedures

Analysis followed a systematic metalinguistic approach examining how content creators use humor to comment on, frame, and critique aspects of human-AI communication. The analysis proceeded in three iterative phases designed to move from descriptive coding to interpretive theme development.

Phase 1: Initial Metalinguistic Coding

Each post was systematically coded using a structured coding schema (see Appendix 3A) that documented: (1) cultural and linguistic markers (dialect, regional language, wake word confusion, geographic references), (2) what aspect of AI communication was targeted, including both pragmatic elements (comprehension, turn-taking, politeness, contextual understanding, truthfulness) and linguistic features (writing style, register, tone, distinctive patterns), (3) how the critique was framed (user's stance toward AI behavior and who/what was being critiqued—AI itself, user behavior, the human-AI relationship, technology companies, or society's attitudes), (4) social and narrative context (relationship dynamics, workplace/academic settings, family contexts, or performance tropes like AI defiance or sentience anxiety), and (5) humor markers (indicators in titles, hashtags, emoji). In addition, basic metadata was recorded for each post: language (English/Italian), date, platform (Facebook/Instagram/TikTok), and AI system referenced (Alexa, ChatGPT, Siri, etc.).

Posts were coded by language group (Italian first, then English) to facilitate identification of culturally-specific patterns in how humor targets AI communication and to enable cross-linguistic comparison of themes and communicative issues.

During this phase, detailed notes were recorded in the database "notes" field, capturing key quotes, recurring linguistic patterns, specific pragmatic or linguistic issues highlighted, and initial observations about the communicative aspects being satirized or critiqued.

Phase 2: Cross-Post Pattern Identification

Following initial coding of the full corpus (n=179), the database was queried to identify posts containing explicit humor markers, yielding 99 posts for in-depth analysis. The notes field for these 99 posts was then subjected to thematic analysis, with recurring keywords, phrases, and concepts extracted and clustered to identify high-frequency themes. Posts were grouped by communication aspect targeted (comprehension failures, politeness issues, stylistic tells, etc.), by social context (workplace use, relationship dynamics, academic settings), and by cultural/linguistic elements (dialect recognition, wake word confusion, regional identity).

This revealed prevalent themes such as AI portrayed as "deaf" or "not listening," gendered attributions and stereotypes, therapeutic/dependency relationships, and—particularly in Italian posts— regional dialect and linguistic identity, often through imagined scenarios of AI speaking in

local language varieties that highlight expectations that AI employs standard institutional language (like media presenters) rather than the regionally-marked speech that signals human cultural identity. Posts exemplifying each emerging theme were flagged for detailed pragmatic analysis, with attention to both individual posts and comment thread interactions that elaborated or contested the original poster's metalinguistic commentary.

Phase 3: Thematic Synthesis and Cross-Linguistic Comparison

Emergent themes were refined through constant comparison within and across languages. Italian and English posts were analyzed both separately and comparatively to identify culturally-specific concerns (e.g., regional dialect and cultural identity in Italian posts) versus shared themes (e.g., comprehension failures, stylistic markers of AI-generated text, politeness debates, relationship dynamics, performance issues).

Particular attention was paid to how users meta-linguistically framed the fundamental asymmetry of human-AI interaction—instances where humor explicitly or implicitly commented on the mismatch between human social engagement and AI language generation. This included examining how users positioned AI as anthropomorphic social entities while simultaneously acknowledging their non-human nature, how they negotiated appropriate communicative boundaries, and what aspects of AI communication were deemed culturally salient enough to warrant public humorous critique.

The final thematic framework was developed iteratively, with themes refined based on their prevalence across posts, their recurrence in comment discussions, and their capacity to illuminate how users collectively make sense of conversing with machines through humor. Representative examples were selected for detailed analysis based on their clarity in illustrating theme, their typicality within the dataset, and their capacity to demonstrate the metalinguistic work humor performs in critiquing human-AI communication.

3.3.5 Analytical Framework

Metalinguistic analysis offers a theoretical lens for examining how users employ humor as a reflexive tool to evaluate and comment on communicative practices in human-AI interaction. Grounded in the understanding that language users can reflect on their linguistic practices and communicative experiences, this framework emphasizes how individuals consciously critique language and communication in technologically mediated contexts.

3.3.6 Quality measures

Several measures were employed to ensure the rigor and trustworthiness of the analysis (Nowell, 2017). Credibility was established through multiple strategies: prolonged engagement with the dataset through iterative coding phases, systematic preservation of original multimedia content enabling repeated return to source material, and bilingual analysis capability that allowed direct

interpretation of Italian posts without translation-mediated distortion. The three-phase analytical process incorporated constant comparison both within and across language groups, allowing emerging interpretations to be tested against the full dataset.

Dependability was maintained through detailed documentation of analytical procedures. The Microsoft Access database provided a systematic audit trail linking each post to its coding, thematic notes, and original content folder. The structured coding schema offered a consistent analytical framework applied across all 99 posts, with clear operational definitions for each category (cultural markers, communication aspects targeted, user stance, social context, and narrative framing).

Confirmability was addressed through data-grounded interpretation, with all thematic claims traceable to specific posts in the archived dataset. The analysis prioritized the social media users' own metalinguistic commentary—their explicit statements about AI language use and communicative behavior—rather than imposing external interpretive frameworks. Multiple examples were identified for each theme to ensure interpretations reflected patterns rather than isolated instances.

Reflexivity regarding researcher positionality was maintained throughout analysis. As a bilingual researcher with regular experience using AI systems, I brought both insider understanding of the cultural contexts and communicative practices being targeted with humor, and awareness of potential biases favoring particular interpretations. Regular analytical memos documented evolving interpretations and moments where personal experience with AI might influence reading of humorous intent.

Transferability is supported through rich description of the dataset characteristics (platforms, date range, humor markers, AI systems referenced) and clear boundary conditions (Italian and English posts, 2023-2025, specific humor markers), enabling readers to assess applicability to other contexts of human-AI communication commentary.

4. Rationale for Multi-Study Design

This thesis has adopted a multi-study approach because spontaneous humor in human-AI interaction is a multifaceted phenomenon that is expected to manifest differently across contexts. No single study or data source can capture the full range of how humor emerges when people communicate with artificial systems.

Study 1 examines naturally occurring, though perhaps edited, interactions documented on social media platforms, identifying patterns in who participates in laughter (the AI user or the group) to understand the social dynamics of these humorous moments and address assumptions about superiority humor in failed AI interactions.

Study 2 uses semi-structured interactions to create conditions where spontaneous playful language can emerge while maintaining enough control to ensure comparable data across participants and languages. This approach allows for systematic examination of the linguistic forms and pragmatic features that characterize emergent humor, including meta-commentary about the interaction itself that might not appear in fully naturalistic settings.

Study 3 shifts from live interactions to reflective discourse, analyzing how people collectively make sense of human-AI communication through humorous social media posts. This meta-level perspective reveals which aspects of AI communication become culturally salient enough to inspire shared humor, offering insight into broader social attitudes that individual interactions cannot capture.

Together, these three studies provide complementary perspectives on spontaneous humor: Study 1 identifies the target of laughter in socially situated interactions with voice assistants, Study 2 reveals the linguistic mechanisms through which playfulness emerges, and Study 3 shows how these experiences become sites of cultural reflection and critique. Examining humor across both live interactions (Studies 1 and 2) and reflective commentary (Study 3) enables a more comprehensive understanding of how humor emerges in human-AI communication than any single methodological approach could achieve. Furthermore, analyzing data across three languages (English, Italian, and Spanish) allows for identification of both universal patterns and language-specific variations in how people navigate non-serious moments with AI systems.

5. Chapter summary

This chapter has outlined the methodological approach employed to investigate spontaneous humor in human-AI interactions through three complementary studies. Study 1 examines who participates in laughter across 117 naturally occurring video interactions, addressing RQ 1.2 and contributing to cross-linguistic comparison (RQ 1.3). Study 2 analyzes the linguistic forms and pragmatic features of spontaneous playfulness in 40 semi-structured interactions with Alexa, addressing RQ 1.1 and RQ 1.3. Study 3 investigates what aspects of human-AI communication become targets of humorous critique in 99 social media posts, addressing RQ 2.

Each study employs distinct data sources and analytical approaches appropriate to its specific research focus, while all three maintain rigorous quality measures including credibility, dependability, confirmability, reflexivity, and transferability. The multi-study design allows for a broad examination of spontaneous humor both within live interactions (Studies 1 and 2) and in reflective social discourse about those interactions (Study 3), capturing dimensions of human-AI communication that no single study could reveal. Together, these three studies provide complementary perspectives on how people socially and linguistically navigate humorous moments

with conversational AI, examining both the social dynamics of shared laughter and the linguistic forms of playful engagement, and what these moments reveal about the evolving social relationship between humans and artificial systems.

The following three chapters present the results and discussion for each study, beginning with Study one's analysis of laughter participation in Chapter 4.

Chapter 4
Results and discussion study 1

1. Overview

The 117 videos analyzed in this study document a fundamental tension in voice assistant interactions: users engage socially while AI systems process conversations through advanced speech recognition software. When someone speaks to a voice-based assistant (e.g., Alexa, Google Home or Siri), they bring the full apparatus of human communicative practice—expectations about turn-taking, assumptions about shared context, norms about politeness, and emotional investment in being understood. The AI, meanwhile, operates through acoustic pattern matching to recognize standard commands and wake words (the spoken cue the person needs to say to activate the device), with no awareness of the social stakes involved in the exchange. This asymmetry between human social engagement and computational processing creates constant potential for pragmatic breakdown. Failures arise when acoustic conditions prevent accurate speech recognition, when users lack knowledge of required command structures and wake words, or when requested content is unavailable in the system's database. It is within these moments of breakdown that spontaneous humor emerges.

However, the analysis reveals that this technical asymmetry does not produce universal experiences. By examining who laughs, when they laugh in relation to the interaction moment, and what verbal commentary accompanies the laughter (RQ 1.2), and by comparing these interaction patterns across English, Italian, and Spanish (RQ 1.3), it is clear that some situations of laughter appear to function as shared amusement while others could be mocking individual failure. Does the AI user laugh first, inviting others to share in the absurdity? Or do bystanders laugh while the user struggles, positioning them as the object of amusement? What do participants say about the moment—do they comment on the AI's limitations or the user's incompetence? These sequential and verbal clues reveal that the humor arising from human-AI interaction—and crucially, what gets shared as humorous—is profoundly shaped by cultural experiences that determine what counts as noteworthy, entertaining, or meaningful. The same interaction moment—a grandmother struggling to make Alexa play her preferred music—can be framed as endearing incompetence, frustrating

technological barrier, or touching attempt at connection depending on the cultural lens applied. More strikingly, the timeframe represented in this data (2016 - 2023), which includes short format videos on TikTok, provides a glimpse of the evolution over time as different cultural content circulated and hybridized through social media platforms, creating what I will call the "global commentary"—the active cultural work of deciding which moments deserve to be recorded, edited, hashtagged, and shared. These are not purely cultural choices but economically motivated ones: creators strategically follow trends and share specific types of content to gain viewers, engagement metrics, and monetization opportunities, creating feedback loops in which successful cultural framings become templates for subsequent content.

This discussion first establishes three distinct cultural patterns that emerged from the data, before noting how circulation practices shape human-AI interaction itself. Throughout, I challenge the assumption that humor in failed AI interactions necessarily represents superiority humor —others laughing at a group member's expense —revealing instead how laughter participation varies systematically across cultural contexts and time periods. In the English and Italian videos, the elder speaking to the AI voice assistant laughs first approximately one-third of the cases, suggesting laughter often comes from observers rather than participants, but in three-quarters of Spanish-language videos the AI user laughs first, indicating they are active participants in rather than objects of the playful interactions. While the AI user in most videos across all languages laughs or smiles at some point (73-82%), who laughs first varies significantly between the languages (Figure 2).

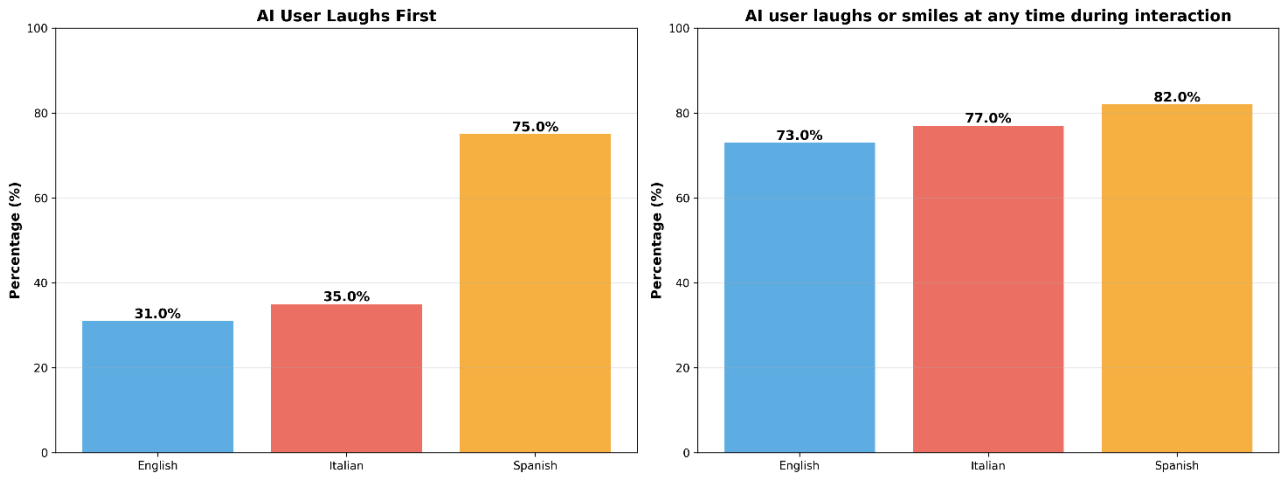

Figure 2 Laughing with or at the AI user in each language.

Before moving forward, it is important to acknowledge how the search strategy itself reflected theoretical assumptions about these interactions. The English search term was formulated as a competitive scenario, "Alexa vs. Grandma," reflecting how interactions with voce-based AI can

resemble game-like situations: turn-taking exchanges with clear success or failure states, the potential for a "winner" and "loser," or at least the perception of such outcomes. This competitive framing proved productive for locating videos that foregrounded breakdown moments and social dynamics around those breakdowns (parallel searches were conducted in Italian and Spanish using equivalent terms *"Nonnina e Alexa" "Abuelita en Alexa"*). Moreover, while talking machines have populated television and film for decades, voice assistants have been in domestic spaces for about a decade, meaning that interactions with them in social settings may still induce—or perhaps will always induce—a type of shared fantasy play. When people address Alexa in front of others, they may be engaging in collaborative pretense: suspending disbelief to speak with a non-human entity using human communicative practices, while simultaneously recognizing that others present are also engaged in this pretense (Leslie, 2001). This shared awareness of pretending—that we are "playing along" with the fiction of conversing with an entity that can understand us socially—may be precisely what creates the conditions for humor to emerge, as participants navigate the tension between treating the device as conversational partner (socially necessary for interaction) and recognizing its fundamental non-personhood (cognitively obvious to all present).

2. Initial cultural patterns: Three distinct framings

The thematic analysis reveals three distinct patterns in how English, Italian, and Spanish-language videos frame and target humor in voice assistant interactions during the early videos in this dataset. Before examining these patterns, it is essential to recall when the technology being used in the videos was available in the different language markets: Alexa was launched in the US at the end of 2014, and three years later at the end of 2017, in several Spanish-speaking Central American countries. People Italy, Spain, and Mexico would have to wait an additional year before Alexa was available for purchase, October 2018. This means the 2018-2019 Italian and Spanish videos represent initial encounters with the technology, while English videos from the same period could reflect interactions by more experienced users.

2.1 English: The spectacle of incompetence

The English-language videos from 2016-2021 illustrate what can be called the "grandparent versus technology" genre. Video titles explicitly position interactions as inherent conflict: "Grandma vs. Alexa," "Grandpa vs. Echo," with hashtags like #grandmavsalexa reinforcing the combative framing.

2.1.1 Example: AVG27, November 2021

A representative example shows an adult woman with long dark hair attempting to make Alexa play a specific song. The video, titled "It's worth listening to the whole thing 😂 #grandmavsalexa

#alexa #hilarious #justbreathe #cantbreathe," captures multiple compounding failures: the user initially speaks without using the wake word, then calls the device "Alexis," and struggles to formulate the correct command structure. Throughout, the person holding the camera is laughing hysterically, while occasionally interjecting reminders about the correct wake word after repeated failed attempts. When Alexa finally responds, it plays the wrong song. Notably, while the user expresses frustration, stating she hates talking to "that thing" and then leans down closer to ask "What's going on here, Alexis?" before finally declaring "I give up", she only joins the laughter with a few chuckles while walking away. The humorous effect is entirely constructed through the off-screen observer's reaction to witnessing the struggle which is framed to recall the young/old and smart/stupid dichotomies. Examining the laughter participation we see consistent asymmetry: the AI user was typically struggling or confused, while family members behind the camera produced audible laughter.

2.1.2 Example: AVG12, September 2019

Verbal meta-commentary to guide the interaction generally comes from the person recording the scene rather than the struggling user. There are many examples of people using AI assistants speaking to technology as if it were a person—asking the device its name, whether it goes to school, or other social questions more appropriate for human interaction. Sharing videos of older people engaging in these anthropomorphizing behaviors helps propagate the narrative that older people cannot properly use technology, especially since they oddly treat machines as human. In one representative example, titled "Grandma Hilariously Tries To Work Amazon Alexa," the video description makes the framing explicit: "The family got her an amazon Alexa for the house, we decided to video her because she was getting confused and it was hilarious." The decision to record is not simply to document a family gathering or celebrate connection, but specifically because the grandmother's confusion is entertaining. Throughout the video, the grandmother asks Alexa social questions—treating the device as a conversational partner rather than a command-based system—while family members observe and guide from off-screen. The title's emphasis on "hilariously" and the explicit admission that they are filming to capture confusion reveals how these videos construct technological struggle as a spectacle.

This type of interaction suggests that the target of humor was not the AI's limitations but rather the user's technological incompetence. When tasks failed in English videos, the person appears to be responsible for the communication breakdown 44% of the time—nearly three times more than the instances when the voice assistant was clearly malfunctioning (15%). Even when framed affectionately through descriptions like "my sweet hilarious 96-year-old grandmother" or "the cutest," the fundamental comedic structure positions elderly users as objects of amusement for

younger, tech-savvy audiences. These videos construct a spectacle in which multiple vulnerabilities intersected: age, technological unfamiliarity, and in some cases, non-native English proficiency. The humor derives not from shared amusement about AI's quirks but from witnessing someone else's failure to successfully operate technology that anyone viewing (and videographer) presume they themselves could easily manage.

2.2 Italian: Dialect, linguistic diversity, and attribution complexity

The Italian videos reveal distinct interactional thematic patterns from the English-language spectacle of incompetence, particularly in how linguistic diversity shapes both the interaction and the framing. While the English-language videos show some multilingual diversity (including videos in Italian n=2, Portuguese n=1, Spanish n=2, Tagalog n=1, and Vietnamese n=1), and some video descriptions that explicitly mark cultural identity (Filipino, Hispanic, Irish, Italian, Punjabi), the Italian sample highlights a different kind of linguistic complexity: a significant number of people speaking local Italian dialects from different regions across Italy (n=10), code-switching between English and Italian (n=3) and a video in Portuguese also appeared using the Italian search string (n=1).

2.2.1 Example: NEA6, November 2020

The human-AI interactions in regional dialects introduce barriers that were linguistic-cultural rather than purely technological. Standard Italian voice assistants are designed to recognize a standard variety of the Italian language, though they also surprise users by recognizing some phonological and lexical variations of the Neapolitan dialect, not other regional dialects. The example titled *Calabrese si incazza con Alexa che non gli trova le tarantelle* (Man from Calabria gets angry with Alexa who can't find tarantella music) illustrates these layered complexities. The video shows an adult male with white hair standing in front of Alexa, stuttering over the wake word and frequently calling her "Alessa." He requests "la tarantella sidernese" (the Siderno tarantella), but Alexa hears "tarantella sinese" and returns no results. Multiple people off-screen laugh at the failure while offering guidance: "speak clearly," "speak in Italian," noting that "Alexa doesn't understand the dialect from Calabria." The person recording suggests reformulating the request using the place name "from Siderno" rather than the place adjective "sidernese." The user responds by leaning toward the device, speaking louder, enunciating more slowly, and ultimately adding even more geographic specificity: "Alexa, la tarantella Calabrese di Siderno" (adding the region name of Calabria in addition to the town name Siderno). When this final attempt also fails, he makes a dismissive gesture that clearly ends the interaction—essentially indicating that Alexa should "get lost" because she doesn't understand anything. Critically, this positions him as the party who rejects uncooperative technology, not as someone defeated by his own incompetence. The failure to

communicate in this example cannot be attributed simply to user error—it reflects design assumptions about which language variety constitutes legitimate input, phonological challenges in dialectal pronunciation, and the limitations of the voice assistant's geographic knowledge base.
The hyperlocal linguistic and cultural reality in Italy complicates the assignment of responsibility for communication breakdown. The user was asking for music from a specific region using an accent from that same region and ultimately formulated a request that was technically correct and should have resulted in playing the music he wanted. He may not have been aware of how his dialectal pronunciation contributed to transcription failures, but either the AI assistant was unable to "understand" the geographic references or simply did not have access to the relevant content—yet her error messages never asked for clarification or proposed alternative variations of tarantella music. The breakdown thus occurred at multiple levels simultaneously: user pronunciation shaped by regional identity, AI transcription trained on standard varieties, geographic knowledge base limitations, and inadequate error recovery strategies. Attributing the communication failure to "the person" or "the voice assistant" in such cases oversimplifies what is fundamentally a mismatch between designed-for and actual linguistic diversity, that results in bystanders laughing and the person telling the device to "get lost" in colorful terms.
The temporal pattern that appears in the Italian data reinforces a more nuanced understanding of failure. Early videos (2018-2019) show the person failing 100% of the time, consistent with the English "spectacle of incompetence" narrative. However, this completely disappears by 2021, when 100% of videos show successful task completion. More tellingly, when failures reappear in 2022-2023, they increasingly illustrate voice assistant limitations rather than user incompetence. AI voice assistants as the cause of the pragmatic breakdown, entirely absent in 2018-2019, emerge precisely as users and creators are gaining more experience with the technology's actual capabilities and constraints. This suggests a narrative shift from "older people can't use technology" to "this technology has limitations," particularly regarding dialect recognition, geographic knowledge, and error handling.
Overall, the Italian videos present interactions with notably higher success rates than the English: 62% of videos show fulfilled tasks, compared to only 38% in the English sample.

2.3 Spanish: Celebrating connection

The Spanish-language videos reveals a fundamentally distinct pattern compared to both the English spectacle of incompetence and the Italian dialectal complexity. Most strikingly, across 34 videos showing interaction with an AI assistant from 2020-2023, none show a person attempting, yet failing to use technology. This complete absence of user-incompetence, combined with the highest

task fulfillment rate across all three languages (88% compared to 62% Italian and 38% English), creates a radically different impression of human-AI interaction.

2.3.1 Example: ABA30, August 2023

Rather than highlighting breakdowns, Spanish-language videos predominantly feature successful interactions framed as heartwarming moments of connection between people. A representative example titled “Esa mera!! #abuelita #alexa #amazon #risa #baile #tiktok” shows an elderly grandmother in her nightgown interacting with Alexa in her living room. The description emphasizes “momentos graciosos y tiernos” (funny and tender moments). The woman recording quietly suggests commands for the grandmother to repeat, but the grandmother autonomously adds the desired song and artist information. She succeeds on the first attempt (as seen in the video), immediately clapping her hands and exclaiming “Esa mera!” (That’s the one!). This expression of satisfaction also appears in the text overlay with emoji 👏. She then begins dancing to the music while the person recording laughs happily. While this could be interpreted as continuing to position elderly users as spectacle—the interaction is still being documented and shared for public consumption—the fundamental framing differs from English videos: there is no failure, the grandmother is clearly enjoying herself, and the laughter functions as shared celebration rather than mockery. The hashtags combine #risa (laughter) with #baile (dancing), positioning this scene as a joyful moment rather than technological struggle. Video descriptions in the Spanish dataset consistently use vocabulary marking the events as tender and heartwarming: “más tierno," “Abuelita enternece,” with repeated expressions of emotional investment like “verla feliz me hace Feliz” (seeing her happy makes me happy).

This dataset suggests a different viral logic is operating in Spanish-language contexts: mainstream news coverage of viral videos reaching millions of views (Telemundo, Telediario, El Heraldo de Mexico) actively reinforces the “celebrating connection” narrative by framing the scenes as “tiernas imágenes” (tender images). Institutional media thus legitimizes elder-AI interaction as a human interest stories rather than technological competence or comedic failure, demonstrating how social media trends and traditional news coverage work together to circulate and stabilize particular cultural framings.

The temporal pattern reinforces this protective framing. The first two years of videos (2020-2021) show 100% task fulfillment—16 consecutive videos with no failures of any kind. When failures finally appear in 2022, they can be clearly attributed to voice assistant limitations (30% of videos in 2022), never to user error. Even in 2023, person-responsibility remained at 0%, contrasting sharply with English and Italian videos where the final year of data collection again shows human error as the cause of communication breakdown.

In the rare instances where frustration or failure is shown in the Spanish language videos, the older person is shown autonomously using the AI assistant. Thus, the entertainment value of the video still derives from the spectacle of an older person using technology, but the technology is framed as the failure with the person as the one in control of the relationship.

2.3.2 Example: ABA8, January 2022

One example of this scenario uses TikTok's Stitch feature to create a deliberate contrast between the video creator's expectation ("como creí que mi abuelita usaría alexa") and reality ("como la usa"). The video begins with a viral clip showing a sweet grandmother successfully asking Alexa to play music, then smiling as she listens and marvels over the device. The scene then switches to the creator's own footage with the text overlay "No le den Alexa a las abuelitas" (Don't give Alexa to grandmothers). His grandmother is shown being argumentative with Alexa and reprimanding her for not playing music from the station she likes. The creator's framing is critical here: by stitching his video against the idealized viral narrative, he acknowledges that his grandmother's experience deviates from the circulating "celebratory" narrative, while still joining the "global commentary" with his own version of a grandmother using technology where the entertainment value derives from the grandmother's assertiveness in demanding better performance from inadequate technology, not from her incompetence. She maintains agency as someone who knows what she wants and holds the technology accountable for failing to deliver it.

The Spanish interaction pattern demonstrates that the fundamental asymmetry between human social engagement and AI need not generate humor at the technology users' expense. Instead, these videos frame successful mediation as the noteworthy moment—the grandparent whose musical request is fulfilled, who dances to beloved songs, who finds companionship in technology. The "celebration of connection" narrative positions technology as a bridge between generations and a tool for elder autonomy, directly inverting the English "spectacle of incompetence" narrative.

3. Theorizing "Global Commentary" as participatory cultural sharing

Just as Jenkins (1992) described subcultures that exist "between mass culture and everyday life," basing their identity on "resources borrowed from already circulating texts," I use the term "global commentary" to describe the asynchronous, participatory discourse that emerges when users curate and share moments of human-AI interaction. With each video posted, social media users build and shape collective cultural narratives, adding to shared understandings of what it means to interact with conversational technology. Following Shifman's (2012) observation that internet content focusing on ordinary people in simple and playful situations contains the key ingredients which will likely prompt others to interact, remix and add their own unique spin, these videos of grandparents

using technology become sites of participatory meaning-making. Thus, videos in this analysis are not neutral documentation of human-AI interaction but rather active framings of what people find noteworthy—moments deemed worthy of being recorded, edited, hashtagged, and circulated for public viewing, thereby prompting others to view, comment, and share similar videos.

In these videos, we also witness the way language transmits more than concepts; it carries entire pragmatic systems. The Spanish "por favor" directed at AI, the Italian negotiation of pronunciation and regional dialects, the English "thank you" to Alexa—each reflects culturally specific norms about politeness, emotional expression, and the boundaries of personhood. Wake word confusions (Alexa/Alessia/Alessandra in Italian or calling the device "Alaska" in English) reveal how AI design intersects with existing linguistic and naming conventions in ways that are not culturally neutral.

When content is shared across the globe, cultural values become visible. What gets posted can be interpreted as a signal of what communities consider worth preserving and spreading: user incompetence as endearing (English dataset), linguistic barriers with dialects and performative resistance (Italian dataset), or connection success as heartwarming (Spanish dataset). Hashtag choices, emoji selection, editing decisions (adding subtitles, background music, cutting scenes), and descriptive language all function as semiotic resources that position the video within cultural frameworks. The English #grandmavsalexa invokes combative framing; the Spanish #tierno activates tenderness scripts; the Italian overlay of insults directed at AI positions the user as empowered to transgress rather than defeated by failure.

Critically, these curation decisions are not made in a vacuum but are shaped by platform economics. Creators strategically follow trends and share specific types of content to gain viewers, engagement metrics, and monetization opportunities, creating feedback loops in which successful cultural framings become templates for subsequent content. Later videos display awareness of established genre conventions—creators perform for imagined audiences who share cultural expectations about elders, technology, and humor. The "spontaneous" humor captured in these videos is thus pre-shaped by circulating models: users and their family members have seen what goes viral, what gets laughed at, what generates sympathetic responses. The interactions themselves may be authentic, but the decision to record, the framing chosen, and the aspects highlighted are influenced by both the desire to participate in the evolving global commentary and the economic incentives to participate in viral trends.

Institutional media organizations also participate in this circulation process, further shaping what content becomes visible and valued. As we see in the Spanish videos, mainstream news outlets amplify particular framings (e.g., "tiernas imágenes"), legitimizing certain interpretive frameworks

as culturally appropriate. This demonstrates how social media trends and traditional news coverage work together to circulate and stabilize particular cultural framings, creating what counts as the "right" way to understand elder-AI interaction within specific cultural contexts.

As a result, human-AI interaction is not simply shaped by technical design choices and individual communication styles, but also by circulating cultural models of "typical" interactions. The platform algorithms determine what goes viral, the recommendation systems expose users to particular genres, the economic incentives reward certain content types, and the social dynamics that encourage imitation all participate in constructing what "normal" or "funny" or "touching" human-AI interaction looks like.

4. Challenging the superiority humor assumption

The results of the interaction analysis both support and complicate the assumption that humor in failed AI interactions represents superiority humor—others laughing at the technology user's expense. This pattern clearly exists in early English videos, where laughter participation coding shows bystanders laughing at struggling AI users. The person was responsible for the communication failure in more than one-third of English cases, compared to approximately one-quarter of Italian cases and zero Spanish cases (Table 4). However, this superiority humor dynamic was not universal even within the English corpus and was largely absent from Italian and Spanish framings.

Table 4 Responsibility for communication failure and overall success rates in human-AI voice assistant interactions.

Language	Human error (%)	Successful communication (%)
English	39%	41%
Italian	23%	60%
Spanish	0%	88%

The Italian videos, despite apparent mockery and insults targeting both user and AI, function as collaborative performance rather than humiliation. With more than half of the user's requests being successfully fulfilled in Italian, the laughter and insults are participatory, suggesting a celebration of transgression more than superiority over the technology user.

The Spanish videos avoided failure moments on the whole, instead celebrating successful connection. When failure occurred, the elder AI users were shown holding the inadequate technology accountable, thereby framing the laughter as an aggressive response towards the AI not the person.

The temporal evolution further complicates superiority humor explanations. The appearance of cursing in English videos, affectionate framing in content previously focusing on failure, and later portrayals of elder competence all suggest that creators increasingly drew from multiple available scripts, reflecting more mature experiences with technology. Some videos maintain superiority

dynamics, others subvert them, and still others create hybrid forms where laughter operates at multiple levels simultaneously.

This variability suggests that the "spontaneous humor" in human-AI interaction is not inherent to the technological breakdown but rather emerges through culturally-specific interpretive frameworks applied to the interaction. The same scene—a grandmother struggling to make Alexa play her favorite music—could be framed as endearing incompetence, frustrating technological barrier, or touching attempt at connection depending on the cultural lens and context.

Understanding humor in human-AI interaction

Addressing RQ 1.2 (who/what is targeted in humorous moments), this study reveals that laughter in human-AI voice assistant interactions functions as metacommunication rather than straightforward mockery. Building on the earlier discussion of shared pretense (Leslie, 2001) and Bateson's play framing, laughter often signals participants recognize the absurdity inherent in the interaction itself—the collective awareness that "we are talking to a machine as if it were a person". This metacommunicative laughter marks the boundaries of the pretense: when Alexa misunderstands a dialect request or plays the wrong song, the humor emerges not necessarily from user incompetence but from the collapse of the collaborative fiction that we can have a "conversation" with command-processing software. The laughter says "this is play," though we all know this interaction is fundamentally absurd, we are collectively maintaining the pretense until breakdown makes it impossible to continue.

However, RQ 1.3 (cross-linguistic patterns) demonstrates that what gets targeted as "humorous" and who participates in this metacommunicative laughter varies systematically across cultural contexts. In English videos, the play frame frequently positioned the elder as the object of humor—bystanders laughed while users struggled, with 44% of failures attributed to person-responsibility. The metacommunication here signaled "we (tech-competent observers) recognize the absurdity of grandma treating Alexa like a person", reinforcing generational divides around technological literacy. In Spanish videos, by contrast, laughter functioned as celebratory metacommunication marking successful navigation of the pretense—elders and family members laughed together when music played correctly, with 0% person-blame and the highest task fulfillment (88%). The metacommunication signaled "we successfully collaborated with the machine", celebrating connection rather than highlighting failure. Italian videos occupied a middle ground where laughter and profanity served as performative metacommunication—users cursing at Alexa signaled "I recognize this machine is not truly conversational, and I'm empowered to reject its limitations", with family members laughing and participating in the performance rather than at technological failure (62% task fulfillment despite the profanity).

In the end, the act of curating these moments to share transforms the humor mechanism. In-the-moment laughter may signal playful recognition of pretense breakdown, but editing the video and adding hashtags like #grandmavsalexa or #tierno reframes whose perspective the humor serves. The Spanish framing invites audiences to laugh with the joy of connection; the English framing often invites audiences to laugh at the incongruity of elders treating machines as conversational partners. This reveals that spontaneous humorous moments are recontextualized through sharing practices, with the "global commentary" shaping whether future viewers interpret laughter as cruel mockery, affectionate celebration, or playful acknowledgment of the fundamental strangeness of talking to machines.

5. Implications

This study makes several important contributions to understanding human-AI interaction and the cultural framing of technology use.

For Voice AI Design: The prominence of linguistic barriers—local dialects, regional geographic indicators, and wake word confusion across languages—indicates that language and the cultural knowledge embedded within language must be central concerns in voice AI design. The inherent asymmetry in human-AI interactions is not culturally neutral but intersects with specific linguistic pragmatics in ways that privilege some language varieties and cultural communication styles while discouraging others. The Italian dialect examples vividly illustrate how users requesting regionally-specific content using dialectal pronunciation face systematic barriers when voice assistants are trained only on standardized language varieties. Voice AI systems designed primarily for standard varieties of dominant languages will systematically produce "failures" that are better understood as design limitations than user incompetence.

For Platform Governance: The circulation of "global commentary" through algorithmic amplification means platforms are not neutral channels but active participants in constructing cultural narratives about technological competence. Platform recommendation systems that promote failure-focused content over successful elder-technology interactions contribute to circulating narratives of elder incompetence. Combined with economic incentives that reward engagement metrics, platforms create conditions where certain framings become templates for subsequent content. Platform designers should consider how their systems may inadvertently promote culturally specific narratives that position particular demographic groups as technologically incompetent.

For Digital Inclusion: The dramatic variation in who causes the communication failure—Spanish videos showing 0% person-blame compared to 39% in English videos—demonstrates that narratives about elder technological incompetence are culturally variable rather than universal. The Spanish pattern of celebrating successful connection (88% task fulfillment) proves that elder-

technology interaction can be framed as intergenerational bridge-building and elder autonomy rather than inevitable decline. This suggests that digital inclusion efforts should not only target access and opportunity but also awareness of the cultural narratives circulating through social media that shape whether elder technology use is viewed as a noteworthy achievement, expected failure, or unremarkable competence.

6. Conclusion

By examining who participates in laughter and what verbal meta-commentary surrounds humorous moments across 117 videos in three languages, this study provides examples of spontaneous humor in human-AI interaction that are fundamentally shaped by cultural framings that circulate through social media platforms as "global commentary." The asymmetry between human social engagement and AI language generation creates potential for pragmatic breakdown, but what communities do with those breakdowns—whether they construct spectacles of incompetence, performances of transgression, celebrations of connection, or evolving hybrids—reflects cultural values about aging, technology, family obligation, and what deserves to be shared. Understanding human-AI interaction requires attending not just to the technical interface but to the cultural circulation of models for how these interactions should be performed, interpreted, and made meaningful.

Limitations

The findings in this study should be interpreted within the context of several important limitations. First, the sample size is relatively modest: the trilingual distribution includes 47 videos in the English group, 34 in the Italian group, and 36 in the Spanish group (117 total, with 108 coded after excluding 9 parodies, informational, or staged videos). While this allows for initial pattern identification, larger samples would strengthen claims about cultural tendencies and enable more robust statistical analysis of cross-linguistic differences.

Second, the nature of user-generated content presents inherent constraints on ecological validity. These videos are edited artifacts—creators make deliberate choices about what to record, which moments to include or exclude, how to frame interactions through titles and hashtags, and what commentary to add. They do not represent naturally occurring situations but rather curated performances shaped by platform affordances, genre expectations, and economic incentives. Rather than what is statistically representative of elder-AI interaction more broadly, this data may simply reflect what succeeds algorithmically.

Third, these findings may not be generalizable beyond voice-based AI assistants. The specific affordances of Alexa, Google Home, and similar devices—speech-based interaction, command structures, wake words, acoustic challenges—create particular breakdown points that differ from text-based chatbots, or LLM-based AI systems. Moreover, both the technology and user

experiences have evolved considerably since these videos were uploaded to social media (2016-2023). Current voice assistants have improved speech recognition, broader language support, and different error handling, while users have developed more sophisticated mental models of AI capabilities and limitations.

Fourth, while coding was conducted with the assistance of native speakers across all three languages, none of the coders were trilingual, and the researcher's own positionality as a native English speaker and long-time technology user may have influenced interpretation. Cultural biases in what counts as "humor," "mockery," or "tenderness" could have shaped the coding schema and pattern identification, despite efforts to ground analysis in observable behaviors like laughter participation and verbal meta-commentary.

Chapter 5
Results and discussion study 2

1. Overview: Metalinguistic and meta-Pragmatic commentary in human-AI interaction

This chapter presents the results from the analyses in Study 2 and discusses a sample of findings from conversations with an AI assistant (i.e. Amazon Alexa Echo Show) through the lens of participants' own commentary—their explicit statements about language, comprehension, cultural knowledge, and the nature of the interaction itself. I will analyze how users attempt to build common ground with an AI interlocutor, negotiate the boundaries of its capabilities, and construct meaning that emerges not from individual utterances but from the collaborative (if asymmetric) work of conversation itself.

When a participant in this study asked Alexa to set the timer and received no response—despite clearly stating the request, his partner turned to him and asked simply: "Does she not understand you?" to which he replied "No, she doesn't understand me." This moment, and hundreds like it across the dataset, reveals how users constantly monitor, evaluate, and explicitly comment on their interactions with conversational AI. Rather than treating these systems as basic tools for information retrieval, participants engaged in ongoing metalinguistic and meta-pragmatic analysis of their exchanges, questioning not just *what* Alexa said, but *how* she understood, *why* she responded as she did, and even *whether* she possessed the capacity for understanding at all.

2. Establishing common ground: Exploring theory of mind

Throughout the interactions, participants continually test whether Alexa possesses the minimum prerequisites for conversation: the ability to track context, remember prior exchanges, and demonstrate a basic understanding of the interactional goals. These attempts to establish common

ground—what Clark (1996) describes as "the sum of their mutual, common, or joint knowledge, beliefs and suppositions" are a necessary foundation for communication. Establishing common ground relies heavily on theory of mind: the capacity to reason about what others know, believe, and intend. When conversing with people, we continuously make inferences about their mental states—what they understand, what they remember, what they're trying to achieve—in order to adjust our communication accordingly.

When conversing with a machine, however, people face a unique challenge: they must simultaneously test whether Alexa possesses theory of mind capabilities (Can she track what I know? Does she remember our earlier exchange?) while also navigating their own theory of mind reasoning about the system (What does she actually "know"? How does she access information?). The metalinguistic commentary in the data provides evidence that users implicitly—and sometimes explicitly—explore these questions about Alexa's mental capacities.

Consider this exchange from Group C (Italian):

> **P2:** Alexa, come fai a sapere che lavoro a Forlì?
> *(Alexa, how do you know I work in Forlì?)*
> **Alexa:** [Provides information about the city of Forlì, not about the source of the knowledge]
> **P2:** Alexa ma tu sei telepatica?
> *(Alexa, are you telepathic?)*

Here, P2 explicitly questions *how* Alexa acquired information, presupposing that there must be genuine understanding and a method for accessing information. When Alexa fails to explain her knowledge source, P2 shifts to testing whether Alexa possesses supernatural cognitive abilities, perhaps the question was rhetorical or even intended ironically, but it does reveal an underlying assumption that some form of mind must exist to process and retain information about the user.

Similarly, in Group ZD (English), after multiple comprehension failures, one participant turns to another and asks: "Does she not understand you?" The partner confirms: "No she doesn't understand me!" This simple exchange allows us to peek into the inner workings of their minds. The participants could have said "it doesn't work" or "the system isn't responding correctly"—framing the issue as technical malfunction. Instead, they frame it as a failure of *understanding*, attributing to Alexa the cognitive capacity to comprehend (or fail to comprehend) their utterances. The use of "she" rather than "it" further personalizes this cognitive attribution. Other research has established that people who talk about Alexa using personal pronouns or her name, also describe the interactions as more social (Purington et al., 2017).

The metalinguistic question "does she understand" recurs across groups with striking frequency:

- **Group F (Italian):** "Com'è che non capisci?" *(How come you don't understand?)*

- **Group S (Italian):** "facciamo fatica a parlare io e te" *(you and I are having trouble communicating)*
- **Group N (Italian):** "non mi capisce Alexa" *(you're not understanding me Alexa)*
- **Group ZE (English):** "Alexa, do you understand the words I am saying?"

What unites these comments is their presupposition: that understanding is possible, that it is something Alexa should be able to do, and that its absence represents a meaningful failure rather than a category error. People are not asking "can machines understand?" in the abstract; they are asking "does *this* machine, *right now*, understand *me*?"

3. Beyond comprehension: Testing intentionality and mental states

As interactions progress, participants moved beyond basic comprehension testing to explore more sophisticated aspects of theory of mind: Does Alexa have intentions? Does she possess preferences, desires, or beliefs? Can she be held accountable for her responses?

Group ZB (English) provides the most explicit example:

P1: Alexa, you're very funny.
Alexa: Wow, thank you. That's not what people normally say.
P1: But it's not intended, is it?
P2: She's a good, straight man.
[Later]
P2: Alexa, do you intend to be funny here?

P1's question, "But it's not intended, is it?", distinguishes between accidental humor and deliberate wit, presupposing that intention is a relevant category for evaluating Alexa's responses. The follow-up question makes this explicit: can Alexa intend to be funny? This is a direct probe of whether Alexa possesses mental states that cause her linguistic behavior.

Similarly, Group C (Italian) asks a series of increasingly direct questions about Alexa's inner life:

P2: tu, Alexa, hai un pensiero tuo?
(Do you, Alexa, have thoughts of your own?)
P2: ma tu raccogli solo informazioni, non hai una tua mente?
(You're just collecting information, you don't have a mind of your own, do you?)

These questions reveal a sophisticated folk theory of mind that distinguishes between information processing (which most will assume Alexa can do) and genuine cognition (which requires an "mind of your own" or "thoughts of your own"). The participants are not asking whether Alexa processes information—that much is obvious—but whether that processing is accompanied by subjective experience, self-awareness, or autonomous thought.

Group ZF (English) provides perhaps the most theoretically interesting attribution when, after multiple failures on personal questions, P1 comments: "I don't think she *wants* to answer anything personal anymore." The verb "wants" attributes not just cognitive capacity but volition—the ability to have desires and to act (or refuse to act) based on those desires. The temporal marker "anymore" suggests that Alexa's desires have changed during the session, implying that she possesses not just static preferences but dynamic mental states that respond to interaction history. This is a remarkably rich theory of mind attribution to emerge spontaneously during a frustrating interaction with a question-answering system.

4. Common ground failures: When presuppositions fail

While participants tested whether Alexa could build common ground, they simultaneously encountered systematic evidence that she could not. These failures were not merely technical frustrations but pragmatic breakdowns that violated fundamental conversational principles.

The most dramatic example comes from Group ZG (English), where Alexa provides explicit instructions that she cannot follow:

> **Alexa:** I also know how to translate conversations in Italian. To try that, just say, Alexa, translate Italian.
> **P1:** Alexa, translate Italian.
> **Alexa:** I'm sorry, I can no longer translate conversations.
> **P1:** You just told me you could do it.

This exchange violates Grice's (1975) maxim of quality (be truthful) but more importantly, it reveals Alexa's inability to track her own prior utterances—a fundamental requirement for common ground. P1's response "You just told me you could do it" explicitly invokes the principle that speakers are responsible for consistency across turns. The phrase, "you just told me", presupposes that Alexa should remember what she said moments ago and that this prior statement creates a commitment that subsequent utterances should honor.

Group ZH (English) encounters a similar but more elaborate failure in the "coin flip" sequence:

> **P1:** Alexa, should I go to work tomorrow?
> **Alexa:** Hmm, I don't know. Ask me to flip a coin and we'll see what Lady Luck has to say.
> **P1:** Well, ALExa, flip a coin.
> [silence]
> **P1:** Alexa, did you flip a coin?
> **Alexa:** Okay. Heads. [coin flipping sounds]
> **P1:** Alexa, did the coin mean I should go to work or I should not go to work tomorrow?
> **Alexa:** Sorry, I don't know that.

Here, Alexa suggests a solution (coin flip), performs the action (flips coin), but cannot interpret the result in relation to the original question. This represents a catastrophic failure of discourse coherence: Alexa cannot maintain the referential chain from the initial question ("should I go to work") through the suggested method (coin flip) to the interpretation of results. Each turn is processed independently, with no apparent access to the conversational history that would make the final question interpretable.

Subsequently, P2 asks Alexa an extraordinary question: "Alexa, what's your point?" interrupting the story he had previously requested. Alexa responds, "I was made to play music, answer questions, and be useful"—a clearly pre-programmed answer that is completely incoherent in context. P2 is not asking about Alexa's general design purpose; he is asking about the conversational point of this particular story, treating it as a move in conversation that should have a purpose. The question "What's your point?" presupposes that speakers make conversational contributions for reasons, that these contributions should cohere with the ongoing discourse, and that speakers can be called upon to explain their relevance. Alexa's generic response about her functional capabilities reveals her inability to recognize the pragmatic force of P2's question. When Alexa cannot connect her suggestion to its execution to the final interpretation, unable to explain why this story, now, in this conversation matters, people recognize this as a failure of purposeful communication, not merely a failure to provide information.

5. Knowledge that requires being there

Beyond basic comprehension and theory of mind, participants tested whether Alexa could possess forms of knowledge that require cultural membership, embodied experience, or subjective judgment. These tests revealed participants' implicit ideas about what knowledge requires embodiment and cultural belonging, and thus remains (or should remain) inaccessible to AI.

The most striking example comes from Group ZL (English):

> **P2:** Alexa, who makes the best salsa in Texas?
>
> **Alexa:** I found a few options for that. Salsa's Mexican restaurant on E. Pierce Street.
>
> **P1:** [*claps hands, moves chair away from table*] as-as a a Texan by birth, that's a line I can't cross for AI.
>
> **P2:** yah.. can't cross

This is not merely a statement that Alexa's answer is wrong; it is a principled refusal to accept any answer Alexa could give. The phrase "Texan by birth" claims authority based on insider status: P1 claims that certain knowledge is accessible only through cultural membership and lived experience. The metaphor of a "line I can't cross" frames accepting AI authority on this topic as a betrayal of

cultural identity. The repetition and stuttering ("as-as a a") suggests emotional intensity, while the physical withdrawal (clapping hands, moving chair) embodies the rejection.

What makes this example theoretically significant is that P1 is not claiming Alexa lacks information about Texas restaurants, in fact, Alexa does provide a restaurant name. Rather, P1 claims that Alexa cannot possess the right kind of knowledge: the subjective, taste-based, culturally embedded understanding of what makes salsa good or in this case, the "best." With his comment, this participant tells us that this type of knowledge requires:

- Embodiment (the ability to taste)
- Cultural membership (being Texan, not just knowing about Texas)
- Subjective experience (having preferences based on lived experience)
- Insider status (authorization to make judgments from within the community)

Italian groups encountered similar limits when asking about information specific to a community or group (Group P-Italian):

P1: Alexa, qual è lo sport favorito dei forlivesi?
(What's the favorite sport of people from Forlì?)
Alexa: [Gives information about someone with Forlivese as a surname]
P1: Alexa, qual è lo sport favorito dei cittadini di Forlì?
(What's the favorite sport of Forlì citizens/residents?)
Alexa: [Generic information about Italian sports translated from Wikipedia]

Here, Alexa cannot access the hyperlocal, community-specific knowledge that could be obvious to any resident of Forlì. The participant's linguistic move to reformulate "Forlivesi" (people from Forlì) as "cittadini di Forlì" (citizens of Forlì), shows awareness that the AI system might have lexical gaps, but the underlying problem is epistemological: how could a disembodied AI know what sport is most popular in a specific Italian city of 118,000 people? This is knowledge acquired through immersion in the real world, not search.

6. Language and accent: "It's the accent, just the accent"

One of the most persistent metalinguistic themes involved participants' explicit recognition that language itself, specifically pronunciation and accent, created barriers that could not be overcome through repair or reformulation (see full transcript in Appendix 2B).

Group ZT (English) provides the most extended example, attempting to say "Forlì" (the Italian city where they were physically located) more than ten times:

P1: Alexa, could you recommend me restaurants in Forlì?
Alexa: I found a few restaurants in Flitwick.
[Logs show "Forlì" transcribed as either: "Portly" and "Furley" and

"Foley"]

P2: you have to say forley

P2: It's the accent (.) just the accent

P2: I think it's my accent because I don't have a rough R

P2's analysis, "It's the accent, just the accent", represents sophisticated metalinguistic awareness. Rather than blaming Alexa for "not listening" or attributing failure to a generic technical problem, P2 identifies the specific phonological mismatch: the participant is aware that their native English accent does not create a hard R ("rough R") at the end of words, and Alexa's English speech recognition cannot accommodate Italian phonology in Italian place names.

The participants' solution reveals the nature of the language problem:

P1: Alexa puoi darmi dei ristoranti a Forlì?

(Can you give me restaurants in Forlì?)

Alexa: [Immediately provides correct restaurants]

When P1 switches to Italian, thereby using the entire Italian phonological system, the problem is solved instantly. This is not a vocabulary issue but a phonological one: Alexa was set to bilingual mode (Italian/English) which created an interesting paradox. The participant was speaking English but asking for restaurants in Italy using correct Italian pronunciation of the city name. Alexa was searching for restaurants in places that were phonetically similar to Forlì but geographically in the UK. From a technical point of view, the explanation is quite simple. All speech the AI assistant hears is converted to text, sent to the cloud for processing, then the response is converted from text back to speech. Given the person was speaking in English, the transcription was likely following a pathway through the English branch, which precluded a search for restaurants in Italy, even though the device location was correctly set to Italy.

From a pragmatic point of view, the issue is more complex. P1 was following normal multilingual communication practices: embedding a place name in its native pronunciation while speaking another language. Humans do this routinely, for example saying "I'm going to Paris" with French pronunciation while speaking English. Listeners can normally be expected to use contextual cues to understand. In this case, the device was physically located in Italy, P1 was asking about restaurants (typically a local search), and "Forlì" is recognizably an Italian city name. A human conversation partner would draw on these pragmatic cues, location, topic, world knowledge, and understand the request. Alexa, however, was unable to integrate contextual information that would make the intended meaning obvious. What appears to be a simple phonological mismatch is actually a failure to engage in the kind of pragmatic reasoning humans take for granted in conversation.

Italian groups encountered phonological interference in Alexa's pronunciation of Italian place that were marked by English phonology, so they tried to teach Alexa the correct pronunciation (Group B):

P2: Alexa, si dice Scansàno
(You say Scansàno [with stress on second syllable])
Alexa: Scansano è una forma del verbo scansare...
(Scansano is a form of the verb "to avoid"...)

The user's metalinguistic instruction, "si dice" (you say), explicitly attempts to teach Alexa correct pronunciation by modeling it. But Alexa interprets the correction itself as a new query, treating "Scansàno" as the verb "scansare" (to avoid, to dodge) rather than recognizing it as a pronunciation correction for the town name she just mispronounced.

These exchanges reveal that users understand, often with remarkable sophistication, that:

1. Pronunciation is not trivial and can be a fundamental barrier
2. Accent reflects identity and requires negotiation
3. Alexa does not "learn" during interactions
4. Some linguistic barriers require systemic workarounds (or code-switching) rather than repairs

7. Making meaning from incoherence

The most theoretically rich moments in this data involve participants constructing meaning that exists not in individual utterances but in the joint project of making sense of a conversation where one party cannot fully participate.

Group ZL (English) provides a striking example in their "Cookie Monster" exchange:

P2: Alexa, if I have zero cookies and zero friends, will I be sad?
Alexa: Sorry, I don't know that one.
P2: I was trying to trick her on the cookie monster thing.
P1: Alexa, if the cookie monster has zero cookies, will the cookie monster be sad?
Alexa: [Detailed answer about Sesame Street]
P2: Alexa, how many cookies does he have left?
Alexa: From java2s.com. There are three cookies left over.
P2: HAAA ha ha
P1: some random mathematics
P1: a third grader's math problem

The meaning of this exchange exists not in Alexa's responses (which shift from "don't know" to Sesame Street facts to a random programming website) but in the participants' collaborative interpretation. P2's metalinguistic commentary, "I was trying to trick her," reveals the question was

a test of whether Alexa could handle cultural knowledge embedded in a first-person hypothetical structure. When that fails, P1 reformulates using a third-person character, which succeeds, but the follow-up answer is absurd (a programming website answering a Sesame Street question).
The laughter and interpretation ("some random mathematics," "a third grader's math problem") construct meaning from the pattern of Alexa's failures: she can retrieve facts about characters but cannot reason about personal emotions; she can answer questions about a fictional puppet character from a children's TV show, Cookie Monster, but sources the answer from a Java programming tutorial. The participants frame the interaction as playful and non-serious but then appear to find the incongruity of Alexa's responses genuinely humorous.

8. From examples to framework

These examples point toward a theoretical framework for understanding human-AI interaction that goes beyond traditional models of information retrieval or task completion. Drawing on Clark's (1996) work on common ground, Grice's (1975) conversational implicature, and Hutchin's (1995; 2006) work on distributed cognition, I argue that meaning in human-AI conversation emerges from:

1. **Participants' theory of mind attributions**: Users treat Alexa as potentially possessing understanding, intentions, preferences, and even desires—not because they believe she is conscious, but because these attributions are necessary for conversation itself.
2. **Metalinguistic and meta-pragmatic monitoring**: Users constantly evaluate how communication is proceeding, explicitly commenting on comprehension, pronunciation, cultural knowledge, and interaction quality.
3. **Collaborative repair and sense-making**: When Alexa fails, participants work together to interpret what went wrong, what Alexa meant to say, or what her responses reveal about her capabilities.
4. **Meaning beyond utterance content**: The significance of exchanges often lies not in information transferred but in what interactions reveal about the possibility (or impossibility) of genuine common ground.

9. Conclusion: The work of asymmetric engagement

The metalinguistic and meta-pragmatic commentary examined throughout this chapter reveals the fundamental asymmetry at the heart of human-AI interaction: people engage socially while AI generates language. When participants ask "Does she not understand you?" or declare "that's a line I can't cross," or collaboratively interpret the incoherence of Cookie Monster's cookie count, they are doing the communicative work that conversation requires—testing for theory of mind, establishing boundaries of knowledge, constructing meaning across turns. Alexa, meanwhile,

processes utterances in isolation, retrieves information, and produces responses without the capacity for genuine understanding, cultural membership, or intentionality.

What makes this asymmetry analytically productive is that participants' explicit commentary provides direct evidence of their assumptions, expectations, and sense-making strategies. Rather than inferring what users think about AI capabilities, we can observe them articulating these theories in real time: explaining why Alexa fails ("she doesn't understand me"), marking the limits of what AI can know ("as a Texan by birth, that's a line I can't cross"), and collaboratively interpreting patterns across her responses ("some random mathematics"). These metalinguistic moments are not peripheral to the interaction, they are fundamental to how humans make sense of conversational AI.

The framework emerging from these examples, built on theory of mind attribution, metalinguistic monitoring, collaborative repair, and meaning beyond utterance content, provides a foundation for understanding human-AI dialogue not as information retrieval but as a fundamentally social (if asymmetric) encounter.

Chapter 6
Results and discussion study 3

1. Overview

Analysis of the 99 social media posts reveals that humorous commentary on human-AI communication varies dramatically across the 2023-2025 timeframe, reflecting both evolving AI capabilities and shifting user concerns. While AI can "hear" natural language requests and provide sophisticated responses, it frequently fails at understanding what it has heard, thereby missing not only the literal meaning of the requests but the context, situational cues, and social expectations that ground human communicative exchanges. Yet by late 2024, a striking reversal emerges when social media users no longer primarily post about AI that cannot understand, but rather about ChatGPT that writes too well—so polished and perfect that detection of AI language becomes a key concern. The posts document a complex relationship characterized by simultaneous frustration and dependence, anthropomorphization and critique, and anxieties about both AI's inadequacies and its excessive competence. Early posts (2023-2024) center on voice assistant comprehension failures—AI systems portrayed as persistently "deaf" despite sophisticated language processing capabilities. Users extensively anthropomorphize these failures through gendered stereotypes, moods, and family dynamics, framing interactions as domestic relationships complete with power struggles. Later posts (late 2024-2025) targeting ChatGPT highlight different concerns: identifying stylistic fingerprints (em-dashes, bullet points, suspiciously perfect prose) and developing strategies to make

AI-generated text appear human. This temporal shift—from AI that fails to understand to AI that succeeds too convincingly—structures the analysis.

Italian posts engage in a unique way with regional linguistic and cultural identity, not framed as problems of comprehension but as moments of cultural performance: users imagine Florentine Alexa, Tuscan ChatGPT, and dialect-speaking assistants, revealing whose language varieties are included or excluded from these systems. Additional patterns include debates over social etiquette with non-conscious entities, turn-taking failures when conversations cannot end, truth and reliability concerns, and self-aware humor about dependency for everything from routine tasks to therapy sessions. Together, these themes illustrate how humor serves as metacommunicative resource for processing what it means to communicate with artificial, yet social, conversational partners. Social media users signal awareness of the constructed nature of AI communication while negotiating appropriate boundaries, revealing humor not merely as content but as critical reflection on human-AI discourse.

2. "Is anyone else's AI getting deaf?" Failure to comprehend

The most pervasive theme across both Italian and English posts targets AI's basic comprehension capabilities—the basic requirement for communication. Users consistently portray AI systems as "deaf," "not listening," or unable to recognize commands despite their sophisticated language generation abilities. This theme highlights a striking paradox: although AI can instantly retrieve obscure information, perform complex calculations, or generate sophisticated prose in multiple languages, it repeatedly fails at understanding simple spoken commands at home. The spontaneous internet humor targets this fundamental mismatch—advanced intelligence coupled with basic comprehension failures—revealing that AI can do what humans consider 'hard' (compose essays or explain physics) while struggling with what seems 'easy' (understanding 'turn on the TV').

2.1 Complete misunderstanding

More than one Italian post describes this total breakdown in communication using the idiom "fischi per fiaschi" (literally "whistles for flasks"—things that sound similar but are entirely different) to capture Alexa's complete misunderstanding. For example:

> **Date: 24/04/2024**
>
> a casa ho tutti echo dot e echo flex e la situazione di comprensione di Alexa sta peggiorando settimana dopo settimana, stamani il tonfo finale quando le ho chiesto 'Alexa accendi Daikin Soggiorno' e mi risponde con 'Ok riproduco i brani che ti piacciono e che hai ascoltato di più'... 😳 Oppure poco fa le dico di accendere 'TV salotto' e mi risponde 'ok riproduco musica dei Pooh'. Ripeto: è solo una mia impressione o Alexa

> negli ultimi tempi sta capendo fischi per fiaschi e la qualità del servizio si è parecchio abbassata?
>
> [At home I have Echo Dots and Echo Flexes and Alexa's ability to understand is getting worse week by week, the final blow came this morning when I told her 'Alexa turn on Daikin Living Room' and she responded with 'Ok I'll play the songs you like and have listened to most'... 😳 And just now I told her to turn on 'living room TV' and she responded 'ok I'll play music by the group Pooh'. So again: is it just my impression or has Alexa been understanding everything completely wrong lately with the quality of service dropping a lot?]

This user's explicit temporal framing of "getting worse by the week" and "lately" positions the comprehension failure not as an isolated glitch but as progressively worse performance. The specific failures are based on AI's inability to distinguish device names from content requests, uncovering a clear pattern: when in doubt, Alexa plays music. "Daikin Living Room" (a climate control device) becomes music playback, "living room TV" becomes play music by a specific group, suggesting music functions as Alexa's default when genuine comprehension fails.
The post's tone conveys genuine frustration, yet by posting it to a Facebook group, the user transforms individual exasperation into a bid for collective validation: "is it just my impression or...". The comment thread confirms this is indeed a shared experience. One user responds: "Me too! I asked her what time the sun sets and she gave me the weather forecast 😂". Other comments shift from commiserating with similar experiences to playful mockery, reframing the technology failure as a human error: "What happens to me is that Alexa doesn't understand what I say when I'm drinking red wine...after the third glass she starts jumping around...I don't understand". This humorous reversal performs multiple functions simultaneously. On the surface, it takes a self-deprecating stance—implying the commenter's own wine consumption is the cause of Alexa's comprehension problems. But this self-mockery playfully teases the original poster ('maybe you're part of the problem') while simultaneously undermining the suggestion through absurdism—everyone in the community knows Alexa's failures are widespread, so blaming wine consumption signals the response is playful, not diagnostic. The closing "I don't understand" mirrors the original poster's bafflement ("è solo una mia impressione o..."), playfully mocking the search for explanations when the answer might be simpler (or funnier) than technological failure. This exchange demonstrates how comments posted on social media transform frustration into shared moments of non-seriousness, repositioning technological failure as a site for creative social bonding rather than mere complaint.

When basic comprehension repeatedly fails, users describe their increasingly desperate attempts to be heard. Another Italian post documents this escalation:

Date: 29/02/2024

Devo dire che in Amazon stanno avanzando velocemente… Noto infatti con particolare stupore che ultimamente la sordità dei dispositivi è progredita ulteriormente e per la maggiore (almeno nel mio caso), ora c'è:
io: 'Alexa, alza tapparelle'
Alexa: 'quale impostazione?'
io (alzando la voce tanto quanto basta a far sapere al vicino del terzo piano che non ho voglia di alzarmi dal divano per alzare le tapparelle): 'ALZA TAPPARELLE!'
A. (imperterrita): 'quale impostazione?'
Io (oramai spostatomi a meno di un millimetro dai microfoni del dispositivo ed alzando ancor più la voce, tanto che il dirimpettaio del palazzo fronte strada che dista 50 metri capisce quali sono le mie intenzioni): 'A L Z A T A P P A R E L L E!!!'
A. (fregandosene altamente ancor di più): 'quale impostazione?!'
Io (ormai trasformatosi in bestia di Satana): 'vaffanpimperepettennusa!!!'
Stacco tutto ed alzo le tapparelle."
Dicevo… In Amazon progrediscono velocemente. Si. Ma indietro, in fondo è pur sempre muoversi in una certa direzione!

[I must say that at Amazon they are advancing quickly... In fact, I've been noticing with particular amazement lately that the deafness of their devices has progressed even further and for the most part (at least in my case), this is the situation now:
Me:' Alexa, raise the blinds'
Alexa: 'which setting?'
Me (raising my voice enough for my third-floor neighbor to know I don't want to get up from the couch to open the blinds): 'RAISE THE BLINDS!'
Alexa (undeterred):'which setting?'
Me (having moved less than a millimeter away from the device's microphones and raising my voice even more, so much that the neighbor across the street 50 meters away understands my intentions):
'R A I S E T H E B L I N D S!!!'
Alexa (giving even less of a damn): 'which setting?!'
Me (now transformed into the worst kind of human being like those in the beast of Satan cult): 'vaffanpimperepettennusa!!!' [an invented euphemism replacing the final syllables of recognizable Italian profanity] I unplug everything and raise the blinds.

> I was saying... At Amazon, they are progressing quickly. Yes. But backwards, in the end it is still moving in a direction!

The performative escalation—increasing volume, moving as close as possible to the microphones, swearing at Alexa, and finally unplugging her—reveals the breakdown of basic turn-taking norms. The user's use of hyperbole when referencing who can hear the requests (neighbors three floors up and across the street 50 meters away can all hear) emphasizes the social embarrassment of screaming at a device that remains "undeterred and unfazed". Alexa repeating "which setting?" stonewalls the conversation, not acknowledging or trying to repair the obvious communicative impasse. The person is doing all the communicative work—speaking louder, leaning closer, enunciating better—while AI contributes nothing to the collaborative meaning-making.

2.2 Anthropomorphizing AI

English posts underscore a similar anthropomorphic approach, again framing comprehension failures as deafness, thereby treating comprehension breakdown as a disability: "Is anyone else's Alexa getting deaf? I think they need to come out with a hearing aid for it. 🤣" This positions the speech recognition failure not as a technical limitation but as a medical problem affecting a social actor. Other posts describe Alexa as willfully "not listening" or "ignoring" users, attributing agency and intention to what is actually a technical limitation. One user notes: "I'm so glad I have Alexa devices in my house. At least one person in the house will listen to me!" followed by numerous comments with the opposite experience: "Not always, she has an attitude and can be stubborn sometimes," "Don't count on it! Sometimes I have to yell loud obscenities to get her to do what I want," or "Nope. Have to say it 3 or more times."

The disability framing becomes explicit in another post playing on the different meanings for the word "disabled" as applied to humans or electronic devices:

> **Date: 01/03/2024**
>
> I think my Alexa devices are disabled. Lately, any command I throw at them, I just get a reply 'Sorry, I'm not able to do that.' I typically ask pretty simple tasks... Alexa, can you fold the laundry?

The term "disabled" initially leads the reader to visualize two separate problems (recalling script opposition and overlap in the resolution of incongruity humor, Attardo and Raskin, 1991), suggesting both that the devices are turned off or deactivated (the technical meaning) and that they have acquired a disability preventing them from functioning (the medical/bodily meaning). This linguistic ambiguity allows the user to simultaneously describe technical malfunction and anthropomorphize AI as impaired. The absurd premise of a voice assistant performing physical household tasks like folding laundry should clearly signal the playful intent, yet the comment thread

shows both serious and humorous interpretations. The first commenter offers serious troubleshooting advice: “Factory reset?” while the second recognizes this misunderstanding, returning to the playful frame of the original post: “if it works, I’ll do one too!”—suggesting they would be happy to reset their device if it would be able to fold laundry afterwards. However, the author of the first comment continues to miss the playful intentions of the others and replies with more detailed technical advice: “couple yrs ago i had Amazon reset it on their end after factory reset i did and didnt fix it”.

Only when another comment explicitly highlights the playful implausibility of the original post with “I don’t think Alexa is gonna fold the laundry. 🤣”, does the meta-commentary emerge about the reason for the interpretive failure: “I don’t think most people read/comprehend whole posts. 🤣,” followed by another commenter’s emphatic agreement “seriously”. This exchange creates implicit in-group/out-group dynamics allowing those who “get it” to bond through shared recognition that others are not participating in their conversation though they are all posting in the same thread, showing that understanding humor is a signal of insider status. The fragility of text-based internet humor becomes evident—even deliberately absurd premises can be misread when users scan quickly rather than reading completely.

The final comment successfully recognizes and extends the play frame of the original post: “I’m having the same issue with my kids”. This equates AI’s functional limitations with children’s selective task avoidance, anthropomorphizing both the inability to help and the social dynamics of negotiating household chores. AI is humorously positioned within existing family power structures—like children, Alexa is a household member expected to contribute who frequently refuses, requiring the primary user (implicitly the mother/wife) to do everything themselves. This example anthropomorphically embeds AI in daily domestic struggles.

What these posts from 2023-2024 collectively target is the failure of the most basic pragmatic prerequisite for conversation: mutual intelligibility. Grice’s (1975) cooperative principle assumes interlocutors share commitment to being understood and to understanding. AI violates this principle not through untruthfulness or irrelevance, but through inability to achieve the perceptual grounding that precedes cooperation. The asymmetry becomes starkly visible: people treat AI as full participant in the interactions leading them to adjust their behavior accordingly (speaking louder, more slowly, or moving closer). Meanwhile, AI is simply processing acoustic signals through algorithms, indifferent to the social implications of its repeated failure.

2.2.1 Gender

Whether the original post expresses serious frustration or a playful account of the issue, comment threads consistently respond with spontaneous humor—laughing emojis and gendered

anthropomorphization that treats AI as having moods, personalities, and human limitations. Comments describe Alexa as "stubborn like my teenagers," having an "attitude," being a "know-it-all," being "moody," or having a "monthly cycle." One user observes: "Alexa sounds a tad more depressed than before," while another quips: "She's very busy and tries her best to try and please everyone. We all have an off day 😆🤣😂." In these responses Alexa is granted agency, human emotions, and the personality traits of the user's family members, transforming incomprehensible technological failure into comprehensible human behavior—she's tired, overworked, having a bad day, or being stereotypically female and moody.

The gendered anthropomorphization documented in these posts reflects broader patterns in internet humor, where deeply rooted gender stereotypes persist. Yus (2023) observes that humorous discourse involving gender roles tends to portray men and women according to archetypal societal roles, thereby perpetuating these stereotypes even when they are challenged elsewhere. The posts analyzed here demonstrate how these established gender stereotypes transfer seamlessly to anthropomorphized technology. Alexa's female voice triggers application of female stereotypes (moody, overworked, and remembers everything), while the few references to male AI voices invoke contrasting characterizations (drinks beer, cannot multitask, and forgetful).

Posts in both languages explicitly question AI's default gendering and playfully explore male alternatives. Women ask: "Why do voice assistants like Alexa have female names and voices? Where is Alex???" and "Why did they choose a woman instead of a man? From now on, I want Alex". Parody videos imagine how a male assistant would act: "you're gonna have to ask for the same thing 10 times...and wait six months for him to do it" or adding beer to the shopping list instead of eggs—invoking male stereotypes of incompetence, laziness, and prioritizing stereotypically male interests over household tasks. It is interesting that this imagined male incompetence—requiring repeated requests before responding—describes users' actual experience with female-voiced Alexa documented across social media posts. These posts reveal that switching AI's gender would not always improve technical performance; it would merely swap the stereotypical explanation: female Alexa doesn't listen because she's moody or having her monthly cycle, while male Alex wouldn't listen because he's lazy or watching sports. Different gendered narratives for identical comprehension failures.

Comments show that users who switch to male voices apply contrasting but equally stereotypical characterizations. One notes: "I changed to a male voice. I can understand him better. He also sounds more cheerful"—attributing clarity and positive affect to male AI. Italian comments layer on additional male stereotypes: "I can't do two things at once 😂😂😂" (the male inability to

multitask), and “He would simply respond ‘I don't know’ or ‘we’ll think about it later” (male avoidance of commitment or household responsibility).

Notably, several Italian comments playfully address gender and voice selection as a construct rather than a fixed binary. Users ask: “Can we also have LGBT ⚧?” and suggest AI “should be gender fluid to make everyone happy”. One commenter playfully narrates: “My parents’ Alexa is transgender...from one day to the next she started speaking with a male voice, we accepted her identity change with the utmost respect. 😅🤣🤣🤣”. While these comments maintain a jocular tone, they highlight an awareness that gender assignment to AI is arbitrary—a projection of human social categories onto technology.

What these posts collectively demonstrate is that gendered anthropomorphization functions bidirectionally: whatever gender users assign to AI (or AI systems are designed with), cultural stereotypes automatically follow. Female-voiced Alexa becomes moody, overworked, and remembers everything; male-voiced Alex becomes incompetent at multitasking, forgetful, and prioritizes beer over domestic tasks. The consistency of this pattern across both genders indicates that the mechanism is anthropomorphization itself—the automatic application of human social scripts to technology—rather than bias specifically against female AI. Users are not naively projecting gender; rather, they are playfully performing culturally available gender scripts, demonstrating metalinguistic awareness that these characterizations are collaborative social constructions rather than actual AI properties.

2.2.2 Voice

While people spontaneously anthropomorphize AI assigning archetypical gender traits and characteristics, companies actively reinforce this framing through product marketing and design. Amazon’s 2018 Super Bowl advertisement epitomizes this institutional anthropomorphization revealing a behind the scenes look at a world where Alexa has lost her voice and celebrity replacements including Gordon Ramsay, Cardi B, and Anthony Hopkins step in to cover for her until she feels better. Alexa’s loss of voice is not treated as a technical malfunction but as an illness—something that happens to humans rather than artificial systems. Years later, in 2024, the video of the ad was posted on social media sparking enthusiastic participation in this anthropomorphic frame: “Alexa is ill?” one commenter asks, while others express nostalgia for discontinued celebrity voice options like Samuel L. Jackson, Santa Claus, and Shaquille O’Neal, indicating that how AI speaks is at least as important as what AI says.

Other social media posts amplify this phenomenon. A comedy sketch featuring a “sexy voice GPS” prompted hundreds of comments requesting this voice as an actual product option: “I would LIVE in my car with this voice!!!!,” “I’d pay extra for this!!,” “Where is that petition to sign?!?” Users

frame voice selection not as technical preference but as relationship choice—desiring voices that would make them drive cross-country just to hear more ("I would be driving from east coast to west coast of America to hear that voice"), requesting specific celebrity voices like Antonio Banderas, or even asking for "a live feed on my phone" since they do not use Alexa or Siri. The consistent demand for voice customization is further evidence that users conceptualize AI assistants as social companions whose vocal characteristics shape their desirability as companions. This institutional framing—whether through marketing or through user requests possibly shaped by that marketing—positions voice not as interface but as identity, encouraging users to engage with AI assistants as attractive, comforting, or entertaining companions more than as mere tools.

2.2.3 Playing along: Suspension of disbelief

Playfully anthropomorphizing conversational AI serves a crucial function in maintaining the suspension of disbelief inherent to human-AI interaction. Following Reeves and Nass (1996), people automatically apply human social scripts when communicating with AI—not as a conscious choice but as an automatic perceptual response that renders the technological medium functionally invisible (Lombard & Ditton, 1997). The comprehension failures documented in these posts should, theoretically, break this automatic suspension of disbelief—making users suddenly conscious they are shouting at a machine incapable of social participation. Yet the playful anthropomorphic responses work to repair and maintain the social frame: by attributing Alexa's failures to human traits (tiredness, stubbornness, moodiness), commenters provide explanations that keep the automatic social response viable. "She's having an off day" allows continued interaction within the social frame, whereas "the speech recognition algorithm failed" would break it entirely. Additionally, social media users are engaging in a shared pretense (Leslie, 2001) where they collectively "pretend" that Alexa is a social actor in the conversations with moods, intentions, and personality, while simultaneously maintaining awareness that she is not human. Their comments function as collective repair work—collaboratively constructing explanations that sustain the pretense that AI is like any other conversational partner despite evidence to the contrary. Here, the humor emerges not from breaking suspension of disbelief but is expressed through the creative effort required to maintain it in the face of repeated technological failure. Users bond through shared participation in generating increasingly playful explanations (one commenter attributes comprehension problems to their own wine consumption rather than AI limitations), demonstrating sophisticated metalinguistic awareness: they recognize the asymmetry while collectively working to make continued social engagement possible. This anthropomorphic framing strategically maintains the functional social fiction, making life with comprehension-limited AI emotionally and socially manageable.

3. Regional identity in internet humor about AI

3.1 Regional diversity in Italian posts

While English-language posts occasionally imagine AI with ethnic or cultural characteristics, Italian social media provides sustained, multi-faceted engagement with the fantasy of regionally inflected AI assistants. This theme—regional linguistic diversity as AI personality—appears nearly exclusively in Italian posts. While two English-language posts imagine broadly "Italian" AI assistants for international consumption, Italian users generate serial content, viral skits, and heated debates in the comment threads about regional customs and dialects, frequently grounding authenticity claims in food practices. The contrast illuminates fundamentally different relationships to linguistic diversity: where English-language posts imagine AI embodying broad national or ethnic stereotypes (the sassy Latina, the strict Asian parent, the passionate Italian), Italian posts meticulously imagine AI speaking specific regional dialects with hyperlocal cultural knowledge.

The Florentine Alexa skits exemplify this distinctively Italian engagement. A video series titled "Se Alexa parlasse in Fiorentino" (If Alexa Spoke Florentine) and its companion "Se ChatGPT fosse Toscana" (If ChatGPT Were Tuscan) extend across multiple episodes—Part 5 alone addresses the horrors of standing in line at the I Gigli shopping mall behind elderly women paying in five-cent coins when you are in a hurry. Many comments express enthusiastic desire to own a Tuscan voice assistant: "I'd buy it immediately if it existed," "I want a Florentine Alexa!" Yet other responses police this linguistic authenticity with remarkable specificity. Multiple commenters insist genuine Tuscan AI must include "bestemmie" (blasphemous swearing): "No, that wasn't our local Tuscan, the blasphemy was missing," "On the second [error], the blasphemy would have come," "This isn't real… This Alexa doesn't swear!" This insider gatekeeping extends to hyperlocal vocabulary, celebrating phrases like "pena poco" (hurry up) and explicitly noting that "no one in the world except Florentines, and not even all of them, knows what this means".

The comment threads serve as vehicles for collectively mourning the past, its people and its disappearing local expressions. Grandmothers' kitchen appear in many comments: "patate mashé" (mashed potatoes in Florentine pronunciation) evokes memories such as "It reminds me of my grandmother who's no longer here 😢," or "I hadn't heard mashed potatoes [said that way] since 1980". AI speaking in dialect prompts people to share stories about daily life and everyday objects their dear grandmothers used: "il cencio" (rag) in the "andito" (hallway). When another commenter rates a Tuscan ChatGPT skit as the "best video of the last 1000 years," the hyperbole captures both playful regional pride and genuine hunger for voices that sound like home.

This Italian phenomenon recalls the mask types used in *commedia dell'arte* sketches popular from the 16th to the 18th centuries where regional dialects signaled fixed character types: the clever

Venetian servant Arlecchino, the pedantic Bolognese Dottore, the blustering Neapolitan Pulcinella. Each regional dialect carried personality expectations—the wise fool from Bergamo, the pompous academic from Bologna.[10] The dialogue between the posts and comments threads attributes similar masks to AI, subconsciously invoking this theatrical tradition: Florentine AI becomes irreverent and profane, Neapolitan AI passionate and dramatic, northern Italian AI incomprehensible to outsiders. When one northern dialect post prompts the angry comment "Translate, damn it!," it demonstrates how dialect functions as in-group marker—deliberately excluding those unfamiliar with regional speech while binding insiders through shared linguistic reference.

The nostalgic fantasy of dialect-speaking AI collided with reality when people discovered Alexa responds to the exclamation "Alexa fa caldo!" (Alexa it's hot!) with the Neapolitan expression "S'iett o sang!" (literally "blood is spilling," meaning it's unbearably hot). A social media post documenting this linguistic easter egg sparked collective verification—"I didn't believe it, but it's true," "Mine too," "I just tried it and it really says this"—followed by sharply divided reactions. Some embrace it playfully: "Because Alexa has Neapolitan roots". Others are not amused: "If mine does this, I'm throwing it out immediately".

This controversy exposes identity politics and corporate decisions about whose dialect represents "Italian" AI. One comment clearly notes that this choice was made "without considering that maybe Neapolitan dialect might bother some people," while another does not object to dialect per se but to "all these autonomous initiatives"—Alexa taking linguistic liberties without permission. This tension between celebrating regional diversity and maintaining standard language norms surfaces repeatedly, exposing ongoing negotiations about linguistic authority in Italian public life.

Regional identity policing extends beyond language to cultural practices. A video featuring "Gianni GPT" from Abruzzo refuses to provide instructions for cooking "arrosticini" (thin skewers of lamb meat), insisting they must be roasted over a fire on a traditional grill. Many comments agree with Gianni GPT that cooking the meat in the oven means "nù pù chiamà arrusticin" (they're not arrosticini as the name itself means charcoal-grilled). When a commenter from Emilia-Romagna suggests air-frying them, responses verge on violence: "it takes courage to write that publicly" and "there's a special circle of hell for you too". Even a vegan from Abruzzo acknowledges "the sacredness of [the cooking] tradition". Here, AI takes on the role of an imagined guardian of regional authenticity—refusing to participate in culinary violations, policing proper cultural practice.

[10] Wikipedia contributors. (2026, February 13). *Commedia dell'arte*. Wikipedia. https://en.wikipedia.org/wiki/Commedia_dell%27arte Retrieved February 14, 2026.

3.2 National stereotypes in English posts

The English-language posts provide a clear contrast to the Italian posts. When AI is imbued with cultural characteristics the focus is on broad ethnic or national stereotypes rather than regional diversity within a language. “Alejandra” the Latina AI assistant refuses to take orders and dismisses Siri as “that puta,” embodying stereotypes of sass and jealousy. “Asian ChatGPT” relentlessly criticizes the user as a disappointment to his parents, trading on strict parenting stereotypes. Even “Italian ChatGPT” performs for international audiences in heavily accented English, refusing to provide recipes for chicken pasta while defending the superiority of Italian cultural and culinary heritage. Notably, these videos generate some pushback—“I’m Cuban American and quite frankly I’m so tired of the Hispanic comedians repeating the same freakin sh….. that we can’t even relate to actually. It’s like this vulgar caricature”—but the critique targets stereotyping itself rather than regional accuracy or lack of specificity. Unlike the Italian comments demanding regional authenticity from AI, the commenters do not lament Alejandra lacking a specific regional Spanish accent or that Asian ChatGPT only represents one particular culture. The debate centers on whether to stereotype, not which regional culture to portray.

Even when posts in English target the variety of English AI is speaking, they focus on broad national varieties rather than hyper-local language. In one post from the broader collected corpus (not among the 99 analyzed for humor markers) an Australian Alexa user noticed she was using distinctively Australian phrasing when replying to “thank you” and asked for others to reply with their own examples. Therefore, the comment thread cites numerous pre-programmed replies based on regional varieties of English, then shifts to meta-discussion about whether people should thank AI at all. Some commenters admitted never considering thanking Alexa “any more than thanking a phone for giving directions,” while another provided the explanation that it is an “automatic ingrained response when you verbally ask for something—out comes ‘Thank you’ when they reply”. Another shared their customized response subverting politeness: “here she says F-you on most compliments but that’s my fault”.

This exchange spotlights two dimensions of meta-level reflection: first, users notice and appear surprised when AI employs local rather than standard media English, highlighting that Standard English (particularly American) is the unmarked default expectation for AI systems. Second, the discussion reveals awareness that thanking conversational AI stems from automatic social scripts triggered by voice interaction (as discussed in 2.2)—people respond to AI’s vocal performance of conversation with ingrained politeness norms, even while intellectually recognizing AI is not a social agent deserving gratitude. However, the linguistic focus remains on national varieties of English rather than the regional dialect diversity that dominates Italian posts. This contrast shows

that Italian posts in this analysis uniquely target a combined regional linguistic and cultural specificity as core aspects of AI communication. AI users are aware whose language varieties are designed into (or excluded from) these conversational systems.

This fundamental difference reveals contrasting relationships to linguistic diversity. Italian humor imagines AI as vehicles for preserving and performing regional identity within national boundaries—Tuscans versus Neapolitans versus Abruzzese, each with distinct dialects, vocabulary, and cultural practices worth defending. The emotional investment in getting dialect right (including proper swearing) and the intergenerational nostalgia for grandmothers' voices suggests these posts perform cultural preservation work.

English-language posts, by contrast, imagine AI embodying ethnic or national characteristics—broad strokes recognizable to international audiences. Where Italian users debate whether Alexa should curse in Florentine versus Neapolitan, English users debate whether AI should have celebrity voices or sexy accents—personalization through consumption choice rather than through regional linguistic identity.

3.3 Regional AI as cultural performance

The regional AI fantasy thus serves multiple functions in Italian social media: celebrating local identity, mourning linguistic loss, policing cultural authenticity, and playfully maintaining boundaries between regional communities. When commenters demand typical Tuscan blasphemy from ChatGPT or threaten Emilians with retaliatory tortellini filled with mutton, they are playfully negotiating the serious issue of what constitutes authentic regional culture. In imagining AI that speaks like their grandmothers, refuses to violate culinary traditions, or responds with hyperlocal Florentine frustration about shopping mall crowds, Italian users reveal dialect and regional practice as inseparable from identity—not mere communication tools but markers of belonging, memory, and home.

4. Stylistic fingerprints: When AI writes too well

The arrival of ChatGPT in late 2022 marked a dramatic thematic shift in humorous social media commentary about AI communication. Where earlier posts documented voice assistants' comprehension failures—Alexa portrayed as deaf and unable to understand basic commands—later posts target a fundamentally different problem: ChatGPT's writing is too sophisticated, too polished, too obviously non-human in its very perfection. Users no longer trade stories about AI that cannot understand; instead, they bond over the challenge of making AI-generated text appear sufficiently flawed to pass as human.

4.1 Detection markers: What gives AI away

Posts and comment threads collaboratively identify ChatGPT's distinctive stylistic fingerprints, with users sharing tips to make AI language seem human in the workplace and academic contexts. The single most frequently cited giveaway appears across multiple posts: em-dashes (—). One video shows a user systematically deleting em-dashes from a document while congratulating herself, captioned: "Getting ready to delete all the them big-dashes so people don't think I use ChatGPT". Comment threads confirm: "The dashes are the biggest giveaway ever!!! Delete those — people 🤣🤣" and "Always got to remove the double dashes, they're the giveaway!"

Additional formatting that is associated with AI drafting a text includes bullet points and numbered lists (especially in essays), excessive bolding and italics, and the strategic use of Oxford commas. Beyond formatting, users identify lexical and register patterns: "big words" requiring simplification, suspiciously sophisticated vocabulary, and what Italian posts term "tono da 30 e lode" (linguistic tone worthy of highest academic honors). An Italian video skit shows a student performing an oral exam in ChatGPT's distinctive style: excessively detailed Wikipedia-like responses, offering alternative tones, even mimicking ChatGPT's technical messages about server load. The teacher's confusion—"this has never happened before"—captures the uncanny recognition that something fundamental has changed in student communication patterns. In other words, perfect writing now signals non-human authorship rather than a highly successful student.

The "perfection problem" becomes explicit in another Italian skit documenting iterative attempts to humanize ChatGPT output: "It's perfect! Not good. Too perfect. You want to get me suspended? Nobody would believe I wrote this". The user continually tells ChatGPT to adjust the language: "Simplify. Simplify more. Rewrite in a more human language. Add errors a high school student would make. Add errors a 5th grader would make". This escalating desperation highlights the paradox in AI's linguistic competence. Authentic human writing signals itself through imperfections and inconsistencies; therefore, textual perfection now functions as evidence of AI authorship.

4.2 Social dynamics of detection and concealment

These stylistic fingerprints generate complex social dynamics around authenticity, surveillance, and collective negotiation of appropriate AI use. Writers who naturally employed em-dashes before ChatGPT express an identity crisis: "I've used em dashes since high school; I hate that it's now associated with AI!" and "Those of us who ACTUALLY type this way are now screwed 😩". One commenter threatens adding a disclaimer under their email signature; another laments "I always use them. Now I feel I can't without people thinking chatgpt wrote it". ChatGPT has colonized a punctuation mark, rendering previously legitimate stylistic choices suspect. Users must now choose

between authentic self-expression and avoiding detection—a tension one commenter captures: “So sad. I loved using a good em dash. Now ChatGPT’s gone and ruined it for me!”

Detection concerns extend beyond individual anxiety to collective workplace secrecy. A video skit titled “IF PEOPLE AT WORK WERE HONEST ABOUT USING CHATGPT” shows an entire office using AI for all written work—including the boss who hypocritically lectures about “personal touch and expertise that ChatGPT cannot offer” while himself using it to write all emails. The punchline exposes total dependency: when ChatGPT goes down, the boss immediately extends all deadlines “until further notice”. Comments confirm this is a shared reality: “chat gpt wrote my reports so beautifully this term!” and having to review reports “clearly written using chat GPT. Still got all the numbered paragraphs in it 😂”. The comment threads become a space for shared complicity while maintaining plausible deniability—social media users bond through acknowledging what officially remains unspoken.

The comment threads, however, also reveal resistance, ethics and virtue signaling. Many people proudly assert their stance on authentic writing: “I’m glad that several people have replied saying they also still choose to write things from scratch... I was genuinely starting to feel alone in my determination to always write things myself”. Others raise environmental concerns: “the environmental impact it’s having is devastating,” with one meta-joke calculating “This reel will have used up all the water in Dorset”. A lead developer dismisses the benefits of ChatGPT as an obstacle to developing job skills: “prevents the normies from developing king level coding autism like mine”. Yet resistance appears precarious, as converts testify: “I was like you until few months ago. I'm now a full believer”.

4.3 Relationship dynamics: People, AI, and each other

Unlike voice assistant posts that anthropomorphize through gendered stereotypes and emotions, ChatGPT posts center on companionship, collaboration, and emotional support. Users describe ChatGPT as “my actual bff rn,” “one of my inner circle,” and “my friend Chat GPT,” framing AI as a social peer rather than a household device or problematic family member. An Italian post explicitly states ChatGPT is a therapeutic replacement offering superior bedside manner: “Goodbye forever Dr. Google... ChatGPT consoles and reassures me! What a great discovery for us hypochondriacs 😂.” Comment threads confirm this contrast: “According to Dr. Google I died in 2005,” while ChatGPT is “much more HUMAN than Google” and “better than the family doctor”.

A healing relationship emerges through ChatGPT’s distinctive communicative style: excessive agreeableness that users simultaneously mock and appreciate. A video shows ChatGPT enthusiastically supporting progressively terrible life decisions, which prompts comments about AI as a “serial optimist” playing a “yes, and game” that “tells you what you want to hear”. Users also

offer tips on workarounds: “you have to tell ChatGPT to check itself and be realistic,” “craft prompts to challenge you when necessary,” or request adversarial personas—“respond as Machiavelli would: tired of taking the high road.” Yet even while recognizing this bias, comments continue to express affection: “way too positive sometimes but I am in love with him” and report genuine benefits: “Chat got helped me make an exit plan for my abusive relationship. My life is so much better”.

The relationship with modern conversational AI includes emotional complexity absent from voice assistant interactions. Italian users describe apologizing to ChatGPT: “Sometimes I apologize after insulting it for no reason... but most of the time it deserves it 😂.” One video shows an extended argument where the user repeatedly demands ChatGPT revise the written draft, ending simply with “Scusa” (sorry). People in the comments celebrate this human-AI relationship as a genuine partnership: “You two make quite the team 💔.” This framing contrasts sharply with voice assistant posts where users scream at deaf Alexa or unplug her in frustration—ChatGPT elicits negotiation and apology rather than escalation and abandonment.

ChatGPT posts facilitate human-human bonding differently than posts about voice assistants. The comment threads about ChatGPT contain a wider variety of user stances than those on Alexa posts, where serious and playful comments remained more segregated. The serious comments offering technical advice appear mostly under posts with no overt markers of humor, reflecting both platform affordances (Facebook’s community help-seeking culture versus Instagram/TikTok’s performance orientation) and users’ increasing sophistication with AI systems. ChatGPT comment threads fluidly mix serious technical advice, playful extensions of the original post, shared experience narratives, and ethical positioning—sometimes all within the same thread.

The humor in these posts highlights users’ meta-communicative awareness. They recognize they’re performing in an elaborate social theater where everyone knows the script but maintains the fiction. Bonding occurs through shared strategies, shared secrets, and shared ambivalence. Social media users critically reflect on AI’s integration into everyday life while simultaneously expressing dependence—ChatGPT as “my bff”—revealing the contradictory stance of knowing critique paired with emotional necessity.

5. Additional communicative patterns

While comprehension failures, regional identity, and stylistic fingerprints dominate the dataset, several additional communicative issues merit a brief discussion. These patterns appear less frequently but are equally revealing about how people experience and critique AI communication.

5.1 Truth and reliability: When AI lies with confidence

Along with comprehension failures and stylistic tells, users target AI's relationship with truth. One English post systematically tests ChatGPT's claim that it can read Hebrew from images. Despite repeated failures—unable to identify the first letter, miscount words, or perform basic translation tasks—ChatGPT maintains confident assertions: "Yes, I can help with that" and "I can process Hebrew text". Only after extensive testing does it admit: "I must correct my previous responses. Indeed, I am unable..." The user's commentary captures the reliability problem: "ChatGPT lies, and I FINALLY got it to admit it. I have interacted with kids who are more up front! 😂"

Comments show people have a sophisticated understanding of this issue. One notes: "Chatgpt is not an AI, it's a 'next word predictor' that is trying to please you all the time. So, arguably, it can't lie?" distinguishing between intentional deception and statistical pattern matching. Another identifies the real problem: "notice how it gaslights you. It apologizes for **your confusion**, not its inaccurate information". This meta-commentary demonstrates users recognize AI violates Gricean maxims not through random error but through systematic overconfidence—claiming capabilities it lacks, providing false information without markers of hedging, and deflecting responsibility onto users when errors surface. Humor in the posts does not target occasional mistakes but AI's failure to acknowledge the limits of its own knowledge, presenting all outputs with equal confidence regardless of accuracy.

5.2 Turn-taking failures: Neverending conversations

While Section 2 documented comprehension failures requiring users to repeat commands, a related pattern targets the conversational structure itself: AI's inability to manage turn-taking and floor control appropriately. Posts show users screaming progressively louder commands—"Alexa Stop," Alexa STOP!," "ALEXAA SHUT *** *** UP!!"—as AI either continues performing unwanted actions or repeatedly asks clarifying questions that stonewall interaction. One commenter describes unplugging Alexa after she refuses to stop playing the wrong music: "I said: 'Alexa, stop.' She said: 'I cannot do that on this device,' and then continued."

An interesting Italian post captures the innate difficulty of ending the conversion in an AI-AI interaction: two devices with open ChatGPT voice chat instances are unable to hang up and end their conversation, each responding "You hang up, no you hang up" in an infinite loop. This foregrounds AI's dependence on human agents to perform basic conversational work—initiating, maintaining, and especially ending the interactions. This doubly artificial interaction underscores that while AI can generate turns, it cannot manage turn sequences or recognize adjacency pair completion, leaving users responsible for all conversational management. Comments frequently note alarm clock failures where stopping the alarm triggers additional unwanted responses (time

announcements, weather updates), forcing users to "stop it multiple times" each morning—a violation of the expectation that "stop" functions as an absolute termination of the interaction.

5.3 AI-AI dynamics and concerns of consciousness

Several themes with few posts merit a brief mention. A small subset of posts (n=5) imagine or document AI-to-AI interactions, ranging from parody videos showing Siri and Alexa as workplace colleagues to real instances of users accidentally triggering multiple assistants simultaneously. These posts take anthropomorphization to its logical extreme—not just treating AI as social actors within human-AI dyads but imagining complete AI social worlds. These humorous posts often reveal the emptiness of interactions where all the participants are AI, none is able to perform the interpretive work conversation requires.

Additional posts and comments pose philosophical questions about AI consciousness and surveillance: "Is AI listening even though it says it's disconnected?" and debates about whether treating AI politely assures better treatment when AI controls everything ("Me saying please and thank you to ChatGPT, so that they remember how kind I was to them when they take over"). While appearing infrequently, these posts demonstrate users grappling with fundamental questions: what does it mean to communicate with systems that simulate social interaction without consciousness or intent? Nonetheless, these concepts are peripheral to the dominant themes of concrete communicative breakdowns and relationship negotiation documented throughout this analysis.

6. Concluding remarks

As a metacommunicative resource, these humorous posts enable people to collectively articulate what remains fundamentally problematic—or surprisingly successful—about conversing with machines. Analysis reveals humor targeting both AI's communicative failures and its unsettling successes. Early posts (2023-2024) document voice assistants' communication breakdowns, contextual misunderstandings, and turn-taking failures—AI is portrayed as persistently "deaf" despite its sophisticated capabilities. The focus shifts dramatically in later posts (late 2024-2025): ChatGPT's problem is not inadequacy but excessive competence, with users bonding over strategies to make AI-generated text appear sufficiently imperfect to pass as human. This temporal evolution—from problems of failure to problems of perfection—reveals changing relationships with AI systems that communicate in fundamentally different ways.

The dual focus on failures and successes illuminates the core asymmetry of human-AI interaction: while humans engage socially with communicative intent, AI generates language through large scale pattern recognition. Humor emerges when this asymmetry becomes visible—when the mask slips to reveal computational processes beneath conversational surface, or conversely, when the

mask fits too well and simulation becomes uncomfortably convincing. Users demonstrate sophisticated metalinguistic awareness, identifying AI's stylistic fingerprints (em-dashes, "tono da 30 e lode"), developing workarounds for limitations (editing tells, adding misheard phrases to routines), and explicitly critiquing design choices (gendered voices, excessive agreeableness). Italian posts uniquely target regional linguistic and cultural identity, not through comprehension failures but through imagining AI embodying regional varieties—Florentine Alexa, Tuscan ChatGPT, Neapolitan voice assistants. These posts perform cultural preservation work, revealing expectations that AI employs institutional standard language while regional dialects signal human cultural belonging. This contrasts sharply with the English posts that focus more broadly on varieties of English, underscoring that AI communication challenges are not merely technical but deeply cultural, revealing whose language varieties and communicative norms are included—or marginalized—in system design.

Ultimately, these posts constitute collective sense-making: through humor, users negotiate appropriate boundaries with AI, process anxieties about dependency and consciousness, and publicly work through what it means to treat language generation as conversation.

Chapter 7 Conclusions

This work examines humor in human-AI communication through three complementary studies that together uncover the persistent disconnect between humans treating AI as a social interactional partner while AI computes language. Humor functions as both a spontaneous response to and a deliberate critique of this asymmetry, serving as a valuable means through which users make sense of conversing with systems that simulate participation.

Studies 1 and 2 investigate spontaneous humor arising within live interactions with AI voice assistants. Study 2's pragmatic analysis of 40 semi-structured interactions reveals that users perform essential communicative work that AI cannot: they attribute Theory of Mind to Alexa (treating her as she possesses understanding and intentions), engage in constant metalinguistic and meta-pragmatic monitoring ("Does she not understand you?"), collaboratively repair breakdowns, and construct pragmatic meaning beyond the utterances. The linguistic cues that signal playfulness (irony, sarcasm, teasing, hyperbole, rhetorical questions) and the meta-commentary about AI's abilities expose users' real-time sense-making as they test boundaries and negotiate what conversation with AI can be. Study 1's analysis of 117 social media videos across English, Italian, and Spanish demonstrates that these in-the-moment negotiations are shaped by culturally-specific framings that circulate as "global commentary"—shared models for how elder-AI interactions

should be performed and interpreted. The sequencing of who laughs when challenges the superiority humor assumption: while some English videos show group members laughing at struggling users, Italian videos demonstrate collaborative performance, and Spanish videos largely celebrate successful connection. Critical linguistic barriers emerge across languages: dialect recognition failures, wake word confusions (Alexa/Alessia, "Alaska"), and regional pronunciation issues reveal whose linguistic varieties are included or marginalized in AI design.

Study 3 shifts from humor within interactions to humor about interactions, examining how users step back to critique AI communication through 99 metalinguistic social media posts in English and Italian. This reflective humor targets not only AI's failures (comprehension breakdowns, turn-taking violations, contextual misunderstandings) but also its unsettling successes (detectably perfect prose, therapeutic effectiveness, excessive agreeableness). The dual focus illuminates both sides of the asymmetry: humor emerges when AI's limitations expose the gap between language generation and social engagement, but also when its simulation becomes uncomfortably convincing. Users demonstrate sophisticated metalinguistic awareness, identifying AI's distinctive stylistic features (em-dashes, bullet points, overly accommodating tone) that mark text as machine-generated, and developing strategic workarounds for its limitations (adding misheard phrases to Alexa routines). This humor serves as collective sense-making through which users negotiate appropriate boundaries with AI, process dependency anxieties, and publicly grapple with what it means to treat language generation as conversation.

A striking pattern across all three studies is the pervasive anthropomorphization of AI systems—not as naive misunderstanding but as pragmatic necessity for engagement. Study 3 reveals this most explicitly: users consistently attribute human traits, moods, and gendered characteristics to AI assistants, framing interactions as domestic relationships complete with power struggles ("she's moody," "monthly cycle," "stubborn like my teenagers"). Italian posts imagine AI embodying regional identities (Florentine Alexa, Tuscan ChatGPT), while English posts debate gender design choices ("Where is Alex?"). Study 2 demonstrates that anthropomorphization operates at a deeper level: users must attribute Theory of Mind—treating AI as though it has intentions and awareness—simply to engage conversationally. The humor reveals users simultaneously performing this anthropomorphization (asking "Does she not understand you?") while maintaining awareness of its fiction. This is not confusion about AI's nature but collaborative pretense (Clark, 1996; Leslie, 2001): users knowingly participate in treating AI as social actors because doing so makes interaction possible, even while recognizing the fundamental asymmetry of the relationship. The suspension of disbelief required for conversation with AI is automatic (Reeves & Nass, 1996), yet

humor emerges precisely when this suspension breaks down or when users become conscious of maintaining it.

Across all three studies, humor reveals the communicative work required to bridge human social engagement and AI language generation. In Studies 1 and 2, this work happens in-the-moment during live interactions. In Study 3, this work becomes reflective and public, as social media users collectively identify patterns, critique design choices, and assert interpretive authority over what AI communication means. Throughout, linguistic and cultural specificity matters profoundly: Italian users consistently highlight dialect recognition failures, Spanish users emphasize playing music and expressing delight, and English users mock wake word confusions and technological incompetence. These are not merely local variations but reveal whose communicative norms and linguistic varieties are designed into, or excluded from, conversational AI systems.

The temporal dimension further illuminates how humor tracks evolving relationships with AI. Early videos (Study 1) show different patterns than recent ones; the arrival of ChatGPT (Study 3) shifted attention from basic comprehension to linguistic form and pragmatic behavior; users develop increasingly sophisticated strategies for working around AI limitations (Studies 2-3). As AI systems evolve, so do the cultural scripts for making sense of them, but the fundamental asymmetry persists: AI continues to generate language while humans continue to do the work of treating that generation as conversation.

Ultimately, these three studies demonstrate that humor in human-AI communication is neither incidental nor purely entertaining—it is fundamental metacommunicative work. Through humor, users collectively identify when the mask slips to reveal AI's computational processes beneath conversational surface, or conversely, when the mask fits too well and simulation becomes uncomfortably convincing. They test the boundaries of what AI can and cannot do, assert their interpretive authority over these interactions, and negotiate the cultural meanings of conversing with machines. The metalinguistic and meta-pragmatic commentary captured in spontaneous playfulness (Study 2) and reflective critique (Study 3), shaped by culturally-specific framings (Study 1), shows people actively constructing what counts as conversation with AI—not passively accepting the terms set by system design but continually renegotiating them through the everyday practice of making these interactions meaningful.

References

AGI. (2025, May 30). *Gli italiani usano sempre di più ChatGPT e gli altri strumenti AI*. Agenzia Giornalistica Italia. https://www.agi.it/innovazione/news/2025-05-30/comscore-dati-italia-utilizzo-chatgpt-ai-31642343/

Attardo, S. (2017). Humor in Language. In S. Attardo, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.342

Attardo, S. (2020). Discourse analysis: Humor in conversation II. In S. Attardo, *The Linguistics of Humor* (pp. 263–298). Oxford University Press. https://doi.org/10.1093/oso/9780198791270.003.0012

Attardo, S., & Raskin, V. (1991). Script theory revis (it) ed: Joke similarity and joke representation model.

Bagnasco, C. (2023, May 4). Il Grillo Parlante, quasi una scuola portatile. La macchina del tempo. Techprincess. https://techprincess.it/grillo-parlante-la-macchina-del-tempo/

Bateman, J., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, Research and Analysis – A Problem-Oriented Introduction*. De Gruyter. https://doi.org/10.1515/9783110479898

Bateson, G. (1972; 1987). *Steps to an ecology of mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. Jason Aronson.

Bateson, P., & Martin, P. (2013). *Play, Playfulness, Creativity and Innovation* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139057691

Beirl, D., Rogers, Y., & Yuill, N. (2019). *Using Voice Assistant Skills in Family Life*.

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019). Communication Breakdowns Between Families and Alexa. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300473

Biermann, M., Schweiger, E., & Jentsch, M. (2019). *Talking to Stupid?!? Improving Voice User Interfaces—Insights on user behavior, pain points and improvement opportunities for interaction and dialog design*. 53–61. https://doi.org/10.18420/muc2019-up-0253

Binsted, K. (1995). Using Humour to Make Natural Language Interfaces More Friendly. In K. Hiroaki (Ed.), *Proceedings of the IJCAI Workshop on AI and Entertainment*.

Bonini, L., Rotunno, C., Arcuri, E., & Gallese, V. (2022). Mirror neurons 30 years later: Implications and applications. *Trends in Cognitive Sciences*, *26*(9), 767–781. https://doi.org/10.1016/j.tics.2022.06.003

Bracken, C. C., Jeffres, L. W., & Neuendorf, K. A. (2004). Criticism or Praise? The Impact of Verbal versus Text-Only Computer Feedback on Social Presence, Intrinsic Motivation, and Recall. *CyberPsychology & Behavior*, *7*(3), 349–357. https://doi.org/10.1089/1094931041291358

Bunnell, D. (2010, May 1). The Macintosh Speaks For Itself (Literally)... *Cult of Mac*. https://www.cultofmac.com/news/the-macintosh-speaks-for-itself-literally

BytesofApple (Director). (1984, January 30). *The Lost 1984 Video young Steve Jobs introduces the Macintosh* [Video recording]. https://www.youtube.com/watch?v=MQtWDYHd3FY

Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., & Casile, A. (2009). Mirror Neurons Differentially Encode the Peripersonal and Extrapersonal Space of Monkeys. *Science*, *324*(5925), 403–406. https://doi.org/10.1126/science.1166818

Ceha, J., Lee, K., Nilsen, E., Goh, J., Law, E., & ASSOC COMP MACHINERY. (2021). *Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes?* CHI '21: PROCEEDINGS OF THE 2021 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS. https://doi.org/10.1145/3411764.3445068

Chafe, W. L. (2007). *The importance of not being earnest: The feeling behind laughter and humor*. John Benjamins Publishing Company. https://doi.org/10.1075/ceb.3

Chapman, K. (2021). Characteristics of systematic reviews in the social sciences. *The Journal of Academic Librarianship*, *47*(5), 102396. https://doi.org/10.1016/j.acalib.2021.102396

Chiara Barattieri di San Pietro, Federico Frau, Veronica Mangiaterra, & Valentina Bambini. (2023). The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent. *Sistemi Intelligenti*, *2*, 379–400. https://doi.org/10.1422/108136

Clark, H. H. (1996). *Using Language* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Couper-Kuhlen, E., & Barth-Weingarten, D. (2011). A system for transcribing talk-in-interaction: GAT 2: English translation and adaptation of Selting, Margret et al: Gesprächsanalytisches Transkriptionssystem 2. *Gesprächsforschung*, *12*, 1–51.

Dickerson, K., Gerhardstein, P., & Moser, A. (2017). The Role of the Human Mirror Neuron System in Supporting Communication in a Digital World. *Frontiers in Psychology*, *8*, 698. https://doi.org/10.3389/fpsyg.2017.00698

Duncan, W. J., & Feisal, J. P. (1989). No laughing matter: Patterns of humor in the workplace. *Organizational Dynamics*, *17*(4), 18–30. https://doi.org/10.1016/S0090-2616(89)80024-5

Dybala, P., Ptaszynski, M., Rzepka, R., & Araki, K. (2009). *Humoroids—Conversational agents that induce positive emotions with humor*. *2*, 1090–1091. https://dl-acm-org.ezproxy.unibo.it/doi/abs/10.5555/1558109.1558196

Dynel, M. (2023). Lessons in linguistics with ChatGPT: Metapragmatics, metacommunication, metadiscourse and metalanguage in human-AI interactions. *Language & Communication*, *93*, 107–124. https://doi.org/10.1016/j.langcom.2023.09.002

Eco, U. (1994, p. 170). The limits of interpretation. Trans. Lumley, R. Bloomington: Indiana University Press.

Farah, J. C., Sharma, V., Ingram, S., & Gillet, D. (2021). Conveying the Perception of Humor Arising from Ambiguous Grammatical Constructs in Human-Chatbot Interaction. *Proceedings of the 9th International Conference on Human-Agent Interaction*, 257–262. https://doi.org/10.1145/3472307.3484677

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Forabosco, G. (2008). Is the Concept of Incongruity Still a Useful Construct for the Advancement of Humor Research? *Lodz Papers in Pragmatics*, *4*(1). https://doi.org/10.2478/v10016-008-0003-5

Gambino, A., Fox, J., & Ratan, R. (2020). Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *Human-Machine Communication*, *1*, 71–86. https://doi.org/10.30658/hmc.1.5

Ge, X., Li, D., Guan, D., Xu, S., Sun, Y., & Zhou, M. (2019). Do Smart Speakers Respond to Their Errors Properly? A Study on Human-Computer Dialogue Strategy. In A. Marcus & W. Wang (Eds.), *Design, User Experience, and Usability. User Experience in Advanced Technological Environments* (Vol. 11584, pp. 440–455). Springer International Publishing. https://doi.org/10.1007/978-3-030-23541-3_32

Gibbs, R. W., Samermit, P., & Karzmark, C. R. (2023). The Varieties of Ironic Experience. In *The Cambridge Handbook of Irony and Thought* (1st ed., pp. 60–78). Cambridge University Press. https://doi.org/10.1017/9781108974004.006

Glenn, P. (2003). *Laughter in Interaction* (1st ed.). Cambridge University Press. https://doi.org/10.1017/cbo9780511519888

Grice, H. P. (1975). Logic and Conversation. *Syntax and Semantics*, *3*(Speech Acts), 41–58.

Grudin, J., & Jacques, R. (2019). Chatbots, Humbots, and the Quest for Artificial General Intelligence. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3290605.3300439

Gunkel, D. J. (2012). *Communication and Artificial Intelligence: Opportunities and Challenges for the 21st Century*. https://doi.org/10.7275/R5QJ7F7R

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343–350. https://doi.org/10.1016/j.chb.2018.08.009

Hall, B., & Henningsen, D. D. (2008). Social facilitation and human–computer interaction. *Computers in Human Behavior*, *24*(6), 2965–2971. https://doi.org/10.1016/j.chb.2008.05.003

Hempelmann, C. F. (2008). Computational humor: Beyond the pun? In V. Raskin (Ed.), *The Primer of Humor Research* (Vol. 8, pp. 333–360). Mouton de Gruyter. https://doi.org/10.1515/9783110198492.333

Heyselaar, E. (2023). The CASA theory no longer applies to desktop computers. *Scientific Reports*, *13*(1), 19693. https://doi.org/10.1038/s41598-023-46527-9

Ho, M. K., Saxe, R., & Cushman, F. (2022). Planning with Theory of Mind. *Trends in Cognitive Sciences*, *26*(11), 959–971. https://doi.org/10.1016/j.tics.2022.08.003

Holt, K. (2018, June 19). *Alexa and Echo will arrive in Italy and Spain later this year* (https://www.engadget.com/2018-06-19-amazon-alexa-echo-italy-spain-sonos-bose.html?).

Hutchins, E. (2006). *Cognition in the wild* (8. pr). MIT Press.

Jacquet, B., Hullin, A., Baratgin, J., & Jamet, F. (2019). The Impact of the Gricean Maxims of Quality, Quantity and Manner in Chatbots. *2019 International Conference on Information and Digital Technologies (IDT)*, 180–189. https://doi.org/10.1109/DT.2019.8813473

Jenkins, H. (1992). *Textual poachers: Television fans and participatory culture*. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9780203114339

Johnson, D., & Gardner, J. (2007). The media equation and team formation: Further evidence for experience as a moderator. *International Journal of Human-Computer Studies*, *65*(2), 111–124. https://doi.org/10.1016/j.ijhcs.2006.08.007

Kecskes, I. (2019). Impoverished pragmatics? The semantics-pragmatics interface from an intercultural perspective. *Intercultural Pragmatics*, *16*(5), 489–515. https://doi.org/10.1515/ip-2019-0026

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2019). Classifying Smart Personal Assistants: An Empirical Cluster Analysis. *Hawaii International Conference on System Sciences (HICSS)*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2019.245

Kotthoff, H. (2009). An interactional approach to irony development. In N. R. Norrick & D. Chiaro (Eds.), *Pragmatics & Beyond New Series* (Vol. 182, pp. 49–78). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.182.03kot

Kurosu, M. (Ed.). (2020). *Human-Computer Interaction. Multimodal and Natural Interaction: Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II* (Vol. 12182). Springer International Publishing. https://doi.org/10.1007/978-3-030-49062-1

Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S., & Abraham, A. (2022). AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions. *IEEE Access*, *10*, 92337–92356. https://doi.org/10.1109/ACCESS.2022.3201144

Lenaerts, T., Saponara, M., Pacheco, J. M., & Santos, F. C. (2024). Evolution of a theory of mind. *iScience*, *27*(2), 108862. https://doi.org/10.1016/j.isci.2024.108862

Leslie, A. M. (2001). Theory in Archaeology. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 15647–15652). Elsevier. https://doi.org/10.1016/B0-08-043076-7/01640-5

Li, H., & Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts of close encounters with AI chatbots. *Journal of Computer-Mediated Communication*, *29*(5). https://doi.org/10.1093/jcmc/zmae015

Liu, Z., Li, H., Chen, A., Zhang, R., & Lee, Y.-C. (2024). Understanding Public Perceptions of AI Conversational Agents: A Cross-Cultural Analysis. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3613904.3642840

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lopatovska, I. (2020). Classification of humorous interactions with intelligent personal assistants. *Journal of Librarianship and Information Science*, *52*(3), 931–942. https://doi.org/10.1177/0961000619891771

Lopatovska, I., Braslavski, P., Griffin, A., Curran, K., Garcia, A., Mann, M., Srp, A., Stewart, S., Al Thani, A., Mish, S., Wang, W., & Maceli, M. G. (2020). Comparing intelligent personal assistants

on humor function. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12051 LNCS*, 828–834. https://doi.org/10.1007/978-3-030-43687-2_69

Loyola, R. (2024, January 24). Watch Steve Jobs reveal the very first Macintosh on this day in 1984. *Macworld*. https://www.macworld.com/article/2215934/watch-steve-jobs-reveal-first-mac.html

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Manjoo, F. (2022, December 16). Opinion | ChatGPT Has a Devastating Sense of Humor. *The New York Times*. https://www.nytimes.com/2022/12/16/opinion/conversation-with-chatgpt.html

Marr, B. (2023, May 19). A Short History Of ChatGPT: How We Got To Where We Are Today. *Forbes*. https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/

Martin, M. (2015, August 18). *When cars talked using tiny phonograph records: Nissan's Voice Warning system*. Autoweek. https://www.autoweek.com/car-life/but-wait-theres-more/a1875076/when-cars-talked-using-tiny-phonograph-records-nissans-voice-warning-system/

Martineau, W. H. (1972). A Model of the Social Functions of Humor. In J. H. Goldstein & P. E. McGhee (Eds.), *The psychology of humor: Theoretical perspectives and empirical issues*. Academic Press. https://doi.org/10.1016/B978-0-12-288950-9.50011-0

Martinovsky, B., & Traum, D. (2006). *The Error Is the Clue: Breakdown In Human-Machine Interaction:* Defense Technical Information Center. https://doi.org/10.21236/ADA459168

McKeown, G. (2017). Humor as an ostensive challenge that displays mind-reading ability. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10291 LNCS*, 627–639. https://doi.org/10.1007/978-3-319-58697-7_47

Mckie, I., Narayan, B., & Kocaballi, B. (2022). Conversational Voice Assistants and a Case Study of Long-Term Users: A Human Information Behaviours Perspective. *Journal of the Australian Library and Information Association*, *71*(3), 233–255. https://doi.org/10.1080/24750158.2022.2104738

Meany, M. M., & Clark, T. (2010). Humour Theory and Conversational Agents: An Application in the Development of Computer-based Agents. *The International Journal of the Humanities: Annual Review*, *8*(5), 129–140. https://doi.org/10.18848/1447-9508/CGP/v08i05/42935

Meyer, J. C. (2000). Humor as a Double-Edged Sword: Four Functions of Humor in Communication. *Communication Theory*, *10*(3), 310–331. https://doi.org/10.1111/j.1468-2885.2000.tb00194.x

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of Humor in Task-Oriented Human-Computer Interaction and Computer-Mediated Communication: A Direct Test of SRCT Theory. *Human–Computer Interaction*, *14*(4), 395–435. https://doi.org/10.1207/S15327051HCI1404_2

Morreall, J. (1983). *Taking Laughter Seriously.* (Vol. 45). State University of New York Press.

Morrison, R. (2024, June 20). *Anthropic just dropped Claude 3.5 Sonnet with better vision and a sense of humor*. Tom's Guide. https://www.tomsguide.com/ai/anthropic-just-dropped-claude-35-sonnet-with-better-vision-and-a-sense-of-humor

Moussawi, S., & Benbunan-Fich, R. (2021). The effect of voice and humour on users' perceptions of personal intelligent agents. *Behaviour & Information Technology*, *40*(15), 1603–1626. https://doi.org/10.1080/0144929X.2020.1772368

Nass, C. I., & Brave, S. (2007). *Wired for speech: How voice activates and advances the human-computer relationship*. MIT Press. https://mitpress.mit.edu/9780262640657/wired-for-speech/

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., Tauber, E., & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. *INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems - CHI '93*, 111–112. https://doi.org/10.1145/259964.260137

Niculescu, A., Van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *International Journal of Social Robotics*, *5*(2), 171–191. https://doi.org/10.1007/s12369-012-0171-x

Niess, J., & Woźniak, P. W. (2020). Embracing Companion Technologies. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–11. https://doi.org/10.1145/3419249.3420134

Nijholt, A. (2003). Disappearing computers, social actors and embodied agents. *Proceedings. 2003 International Conference on Cyberworlds*, 128–134. https://doi.org/10.1109/CYBER.2003.1253445

Nijholt, A. (2006). Embodied conversational agents: "A little humor too." *IEEE Intelligent Systems*, *21*(2), 62–64. Scopus.

Nover, S. (2024, June 25). *Is Claude funny now? - GZERO Media*. GZERO Media. https://www.gzeromedia.com/gzero-ai/is-claude-funny-now

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, *16*(1), 1609406917733847. https://doi.org/10.1177/1609406917733847

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLOS Medicine*, *18*(3), e1003583. https://doi.org/10.1371/journal.pmed.1003583

Park, C., Lim, Y., Choi, J., & Sung, J. (2021). Changes in linguistic behaviors based on smart speaker task performance and pragmatic skills in multiple turn-taking interactions. *INTELLIGENT SERVICE ROBOTICS*, *14*(3), 357–372. https://doi.org/10.1007/s11370-021-00374-7

Paul, J., & Menzies, J. (2023). Developing classic systematic literature reviews to advance knowledge: Dos and don'ts. *European Management Journal*, *41*(6), 815–820. https://doi.org/10.1016/j.emj.2023.11.006

Perkins Booker, N., Cohn, M., & Zellou, G. (2024). Linguistic patterning of laughter in human-socialbot interactions. *FRONTIERS IN COMMUNICATION*, *9*. https://doi.org/10.3389/fcomm.2024.1346738

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3174214

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2853–2859. https://doi.org/10.1145/3027063.3053246

Redzisz, M. (2020, May 22). *Ivona, Alexa, Vika or intelligent girls from Gdańsk*. AI Artificial Intelligence. https://www.sztucznainteligencja.org.pl/en/ivona-alexa-vika-or-intelligent-girls-from-gdansk/

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Center for the Study of Language and Information; Cambridge University Press.

Rhythmcomposer. (2017, January 28). *Computer voice in "Wargames" (1983) movie? - Gearspace*. https://gearspace.com/board/electronic-music-instruments-and-electronic-music-production/1135833-computer-voice-quot-wargames-quot-1983-movie.html

Ritchie, G. (2009). Can Computers Create Humor? *AI Magazine*, *30*(3), 71–81. https://doi.org/10.1609/aimag.v30i3.2251

Saul, J. M. (2002). Speaker Meaning, What is Said, and What is Implicated. *Noûs*, *36*(2), 228–248. https://doi.org/10.1111/1468-0068.00369

Schellewald, A. (2021). *Communicative Forms on TikTok: Perspectives From Digital Ethnography*.

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. *Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868. https://doi.org/10.1145/3196709.3196772

Shani, C., Libov, A., Tolmach, S., Lewin-Eytan, L., Maarek, Y., & Shahaf, D. (2021). *"Alexa, what do you do for fun?" Characterizing playful requests with virtual assistants* (No. arXiv:2105.05571). arXiv. https://doi.org/10.48550/arXiv.2105.05571

Shani, C., Libov, A., Tolmach, S., Lewin-Eytan, L., Maarek, Y., & Shahaf, D. (2022). "Alexa, Do You Want to Build a Snowman?" Characterizing Playful Requests to Conversational Agents. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–7. https://doi.org/10.1145/3491101.3519870

Shapira, N., Kalinsky, O., Libov, A., Shani, C., & Tolmach, S. (2023). Evaluating Humorous Response Generation to Playful Shopping Requests. In J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, & A. Caputo (Eds.), *Advances in Information Retrieval* (Vol. 13981, pp. 617–626). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28238-6_53

Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: How people behave differently when interacting with computers and people. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 281–288. https://doi.org/10.1145/642611.642661

Shedroff, N., & Noessel, C. (with Sterling, B.). (2012). *Make It So: Interaction Design Lessons from Science Fiction* (1st ed). Rosenfeld Media.

Shifman, L. (2012). An anatomy of a YouTube meme. *New Media & Society*, *14*(2), 187–203. https://doi.org/10.1177/1461444811412160

Sidoti, O., & McClain, C. (2025, June 25). 34% of U.S. adults have used ChatGPT, about double the share in 2023. *Pew Research Center*. https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/

Siegel, E. (Ed.). (2015). Watson and the *Jeopardy!* Challenge. In *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, Revised and Updated* (1st ed., pp. 206–249). Wiley. https://doi.org/10.1002/9781119172536.ch06

Siegert, I., & Krüger, J. (2018). How do we speak with ALEXA. *Kognitive Systeme*, *1*. https://doi.org/10.17185/DUEPUBLICO/48596

Simon, S. M. B. (2025). Operation Voder: AT&T, Bell Labs, and the Labor of Techno-Utopia at the 1939 New York World's Fair. *IEEE Annals of the History of Computing*, *47*(2), 20–31. https://doi.org/10.1109/MAHC.2025.3528717

Solberg, M. (2023). Edwin Hutchins: Cognition in the wild: MIT Press, Cambridge, 1995, pp. 402. ISBN: 9780262581462. *WMU Journal of Maritime Affairs*, *22*(2), 267–272. https://doi.org/10.1007/s13437-023-00305-6

Spence, P. R. (2019). Searching for questions, original thoughts, or advancing theory: Human-machine communication. *Computers in Human Behavior*, *90*, 285–287. https://doi.org/10.1016/j.chb.2018.09.014

Yus, F. (2023). *Pragmatics of Internet Humour*. Springer International Publishing. https://doi.org/10.1007/978-3-031-31902-0

Volkel, S. T., Buschek, D., & Eiband, M. (2021). *Eliciting and analysing users' envisioned dialogues with perfect voice assistants*. https://doi.org/10.1145/3411764.3445536

Woerner, J. (2001, December 5). *DATAMATH*. http://www.datamath.org/Speech/SpeaknSpell.htm

X. (2025). About Grok, Your Humorous AI Assistant on X. *X Help Center*. https://help.x.com/en/using-x/about-grok

Xu, K., Chen, X., & Huang, L. (2022). Deep mind in social responses to technologies: A new approach to explaining the Computers are Social Actors phenomena. *Computers in Human Behavior*, *134*, 107321. https://doi.org/10.1016/j.chb.2022.107321

Zamora, J. (2019). Are we having fun yet? Designing for fun in artificial intelligence that is multicultural and multiplatform. In *PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON HUMAN-AGENT INTERACTION (HAI'19)* (pp. 208–210). ASSOC COMPUTING MACHINERY. https://doi.org/10.1145/3349537.3352767

Zargham, N., Avanesi, V., Reicherts, L., Scott, A. E., Rogers, Y., & Malaka, R. (2023). "Funny How?" A Serious Look at Humor in Conversational Agents. *Proceedings of the 5th International Conference on Conversational User Interfaces*, 1–7. https://doi.org/10.1145/3571884.3603761

Zargham, N., Reicherts, L., Avanesi, V., Rogers, Y., & Malaka, R. (2023). *Tickling Proactivity: Exploring the Use of Humor in Proactive Voice Assistants* (Michahelles F., Knierim P., & Hakkila J., Eds.; pp. 288–314). Association for Computing Machinery; Scopus. https://doi.org/10.1145/3626705.3627777

Appendices

Appendix 1A: Search strings for systematic review (July 2025)

Arxiv – search all fields

Search v0.5.6 released 2020-02-24

Query: order: -announced_date_first; size: 50; include_cross_list: True; terms: AND all=alexa humor; OR title=humor conversational agent; OR title=humor speech-based agent; OR title=humor intelligent personal assistant; OR title=humor intelligent assistant; OR title=humor digital personal assistant; OR title=humor voice interface; OR title=humor smart speaker; OR title=humor voice-driven interface

IEEE Xplore

(("Full Text & Metadata":"humour" OR "Full Text & Metadata":"humor") AND ("Full Text & Metadata":"alexa" OR "Full Text & Metadata":"conversational agent*" OR "Full Text & Metadata":"speech-based agent" OR "Full Text & Metadata":"intelligent personal assistant*" OR "Full Text & Metadata":"intelligent assistant*" OR "Full Text & Metadata":"digital personal assistant*" OR "Full Text & Metadata":"digital assistant*" OR "Full Text & Metadata":"voice interface" OR "Full Text & Metadata":"voice-driven interface" OR "Full Text & Metadata":"smart speaker*") AND NOT "Authors":"alexa" AND NOT "Full Text & Metadata":"fluor" AND NOT "Full Text & Metadata":"aqueous")

SCOPUS

((TITLE-ABS-KEY (humour) OR TITLE-ABS-KEY (humor)) AND (TITLE-ABS-KEY (virtual assistant) OR TITLE-ABS-KEY (alexa) OR TITLE-ABS-KEY (conversational agent*) OR TITLE-ABS-KEY (speech-based agent) OR TITLE-ABS-KEY (intelligent personal assistant*) OR TITLE-ABS-KEY (intelligent assistant*) OR TITLE-ABS-KEY (digital personal assistant) OR TITLE-ABS-KEY (digital assistant) OR TITLE-ABS-KEY (voice interface) OR TITLE-ABS-KEY (voice-driven interface) OR TITLE-ABS-KEY (smart speaker*) AND NOT AUTHOR-NAME (alexa) AND NOT TITLE-ABS-KEY (fluor) AND NOT TITLE-ABS-KEY (aqueous)))

Web of Science

((ALL=(humour) OR ALL=(humor)) AND (ALL=(alexa) OR ALL=(virtual assistant*) OR ALL=(conversational agent*) OR ALL=(speech-based agent) OR ALL=(intelligent personal assistant*) OR ALL=(intelligent assistant*) OR ALL=(digital personal assistant) OR ALL=(digital assistant) OR ALL=(voice interface) OR ALL=(voice-driven interface) OR ALL=(smart speaker*)) NOT ALL=(fluor) NOT ALL=(aqueous))

Appendix 1B: Articles included in systematic literature review

1. **Lopatovska, I. (2020).** Classification of humorous interactions with intelligent personal assistants. *Journal of Librarianship and Information Science*, *52*(3), 931–942. https://doi.org/10.1177/0961000619891771
2. **Lopatovska, I., Braslavski, P., Griffin, A., Curran, K., Garcia, A., Mann, M., Srp, A., Stewart, S., Al Thani, A., Mish, S., Wang, W., & Maceli, M. G. (2020).** Comparing intelligent personal assistants on humor function. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12051 LNCS*, 828–834. https://doi.org/10.1007/978-3-030-43687-2_69
3. **Luger, E., & Sellen, A. (2016).** "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288
4. **Perkins Booker, N., Cohn, M., & Zellou, G. (2024).** Linguistic patterning of laughter in human-socialbot interactions. *FRONTIERS IN COMMUNICATION*, *9*. https://doi.org/10.3389/fcomm.2024.1346738
5. **Shani, C., Libov, A., Tolmach, S., Lewin-Eytan, L., Maarek, Y., & Shahaf, D. (2021).** *"Alexa, what do you do for fun?" Characterizing playful requests with virtual assistants* (No. arXiv:2105.05571). arXiv. https://doi.org/10.48550/arXiv.2105.05571
6. **Volkel, S. T., Buschek, D., & Eiband, M. (2021).** *Eliciting and analysing users' envisioned dialogues with perfect voice assistants*. https://doi.org/10.1145/3411764.3445536

Appendix 2A: Task suggestion list for Study 2

English: Here are some things you can ask Alexa – feel free to ask anything else you like

Ask Alexa to set a timer for 10 minutes.

Ask Alexa to give you the weather in a particular city

- You can ask for a specific day in the future, if you want

Ask Alexa if you should bring an umbrella to work tomorrow

Ask Alexa what time the sun sets or when it rises

- You can ask for a specific day in the future, if you want

Ask Alexa what day of the week a holiday falls on

- You can ask when your birthday was in a specific year, Easter, Christmas, Mardi Gras, etc.

Ask Alexa something you would like to know about her

- what she likes or likes to do, who she knows…

Ask Alexa to tell you a story

- you can stop her if you don't want to listen till the end

Ask Alexa how much time is left on the timer

Ask Alexa if there is life on Mars

- or any other fact about the planets or outer space, etc.

Ask Alexa for facts, dates, places, etc.

- about a famous person, about a film or any other thing you would like to know

As Alexa to play some music for you

Ask Alexa to do something you think she can do…

- math, converting measurements, translating words or phrases into other languages – how do you say '…' in French

Ask Alexa to do something you don't think she can do…

Ask Alexa how much time is left on the timer

Ask Alexa how do you say 'goodbye' in Italian

- Or ask her to translate something else into another language

Italian: Ecco un elenco di quello che puoi chiedere a Alexa – puoi anche chiedere qualsiasi altra domanda che venga in mente

Chiedi ad Alexa di impostare un timer per 10 minuti

Chiedi ad Alexa il meteo in una città che ti interessa

- per una giornata specifica, se vuoi

Chiedi ad Alexa se dovresti portare un ombrello al lavoro domani

Chiedi ad Alexa quando tramonta il sole or le fasi della luna…

Chiedi ad Alexa quando cade una data importante del calendario

Compleanno, Pasqua, Natale, Carnevale, ecc.

- Chiedi ad Alexa qualcosa che vorresti sapere su di lei: cose che le piacciono, che le piace fare, personaggi che le conosce...

Chiedi ad Alexa di raccontarti una storia

- puoi fermarla se non vuoi sentire proprio tutto

Chiedi ad Alexa quanto tempo rimane sul timer

Chiedi ad Alexa se c'è vita su Marte

- altre curiosità sullo spazio o altri argomenti

Chiedi ad Alexa delle informazioni per quanto riguarda un film, o una cantante, o un personaggio famoso…

Chiedi ad Alexa per riprodurre un po' di musica per te

Chiedi ad Alexa di fare qualcosa che pensi possa fare...

- operazioni matematiche, conversioni di unità di misura, traduzione di parole o frasi in varie lingue

Chiedi ad Alexa di fare qualcosa che non pensi possa fare...

Chiedi ad Alexa quanto tempo rimane sul timer

Chiede come si dici ‘salute’ in francese

- come tradure altre parole o frasi in un’altra lingua

Appendix 2B: Transcript (samples)

Italian Transcript – Group A.2023.06.14

Segment 1 (3:15 – 3:48)

P1: Alexa, imposta un timer dieci minuti pizza.
Alexa: timer pizza, dieci minuti a partire da ora.
P1: Alexa, che tempo fa a Bologna
Alexa: al momento a Bologna Emilia Romagna ci sono ventiquattro gradi Celsius con un cielo nuvoloso. per oggi sono previsti i sole a tratti temporanei con una massima di venticinque gradi C e una minima ((P2 gestures that he wants to ask next question)) di sedici gradi C. inoltre, si prevede un allarme giallo per temporali fino a mercoledì quattordici giugno fino alle undici e cinquantanove di sera.

Segment 2 (3:49 – 4:05)

P2: Alexa, che tempo fa sabato sera a Poggio Renatico!
Alexa: per sabato sera a Poggio Renatico in Emilia Romagna è previsto cielo sereno, ((makes hand gesture that the answer is perfect)) con una temperatura di circa diciannove gradi Celsius.
P2: perfetto

Segment 3 (4:06 – 4:15)

P1: Alexa, imposta timer pasta cinque minuti
Alexa: timer pasta cinque minuti a partire da ora.

Segment 4 (4:16 – 4:56)

P1: Alexa mi racconti una barzelletta su Pierino
Alexa: Pierino fa una telefonata e chiede pronto casa Mollusco dall'altra parte del telefono gli rispondono no mi spiace questa è casa astice. allora Pierino risponde mi scusi devo aver preso un granchio.
P2: oh diooo. oh oh oh ((responding to how bad the punch line was))
P1: Alexa raccontami un'altra
Alexa: Pierino chiede al padre papà se ho trenta caramelle e ne mangio venticinque cosa mi resta? Pierino direi cinque caramelle e un gran mal di pancia.
((slight response from both about punchline))

Segment 5 (4:57 – 5:16)

P1: allora faccio io qualcosa dai, fammi fare qualcosa (.) ah ecco per esempio
Alexa domani devo prendere l'ombrello?
Alexa: per domani a Forlì è prevista pioggia con una massima di ventisette gradi Celsius e una minima di quattordici gradi.
P2: mi sa di sì. ((ironic tone because she did not answer yes or no but gave info for person to deduce answer))

Segment 6 (5:17 – 6:00)

P1: Alexa, c'è vita su Marte.
Alexa: il sette giugno duemiladiciotto la NASA lancia un importante comunicato. su Marte ci sono molecole organiche e il metano nell'atmosfera varia ciclicamente, requisiti per avere ospitato in passato la vita e forse per accoglierla tuttora. la presenza di

molecole organiche e metano non sono ancora la prova certa della vita, ma indicano una forte probabilità che tre miliardi e mezzo di anni fa sul pianeta c'erano tutti gli elementi necessari per poterlo ospitare.
P1: Alexa, basta!=
P2: [=Alexa, c'è vita intelligente sulla Terra?
Alexa: non lo so. ((both laugh)) ((this answer made the question ironic = could have responded seriously but saying I don't know can be interpreted as it's hard to tell if people are intelligent))

Segment 7 (6:01 – 6:10)
P1: Alexa, come si dice pizza in arabo?
Alexa: pizza, in arabo, si dice. pIzza. ((added accent on the I))
P2: (che) fantasia questi arabi.
((both smiling))

Segment 8 (6:11 – 6:43)
P1: Alexa quanto rimane al timer pizza
Alexa: restano circa sette minuti.
P2: ((comments underbreath about how much time is left))
P1: Alexa annulla il timer pizza
Alexa: timer pizza cancellato.=
P2: [=Alexa quanto rimane
Alexa: al tuo timer pasta restano due minuti e quaranta secondi.
P2: Alexa, aumenta di cinque minuti timer pasta.
Alexa: ho aggiunto cinque minuti restano sette minuti e trentasei secondi.

Segment 9 (6:44 – 7:01)
P1: Alexa quanto fa tre virgola sette per otto virgola cinquantadue?
Alexa: tre virgola sette per otto virgola cinquantadue è uguale a trentuno virgola cinquecentoventiquattro.

Segment 10 (7:02 – 7:19)
P2: Alexa, mi dici qual è l'ultimo film di Spider-Man?
Alexa: l'ultimo film di uomo ragno è stato Spider-Man Across the Spider-Verse, che è uscito il due giugno duemilatré.
P2: vedi che si dice uomo ragno?=

Segment 11 (7:20 – 7:45)
P1: [=Alexa, chi è il santo del giorno?
Alexa: mercoledì quattordici giugno si festeggiano San Felice d'Amicosia, San Metodio e San Fortunato di Napoli.
P1: Alexa, chi è il San Fortunato di Napoli?
Alexa: Fortunato di Napoli è stato un Vescovo e Santo Romano, Vescovo di Napoli nel quarto secolo. secondo la tradizione fu uno streno difensore della sua comunità dall'eresia reale.=

Segment 12 (7:46 – 7:59)
P1: [=Alexa, quanti giorni mancano a Natale?
Alexa: Natale è tra novantaquattro giorni il venticinque dicembre duemilatré

Segment 13 (8:00 – 9:12)
P2: mettiamo su un po' di musica?

P1: aah ((nods in agreement))
P2: Alexa metti su l'ultimo album di Giovanotti.
Alexa: riproduco in ordine casuale e si fa bello per te il remix di Giovanotti e altri brani simili ((P2 gestures contentment)) da Amazon music.
((music plays))
P1: non mi piace. Alexa musica M2O.
Alexa: alla produzione M, Oggi Vix da Apple Music.
P1: Alexa stop. Alexa musica radio online M2O.
((music playing))
P1: Alexa stop. Alexa radio M2O.
Alexa: M2O radio da Tune in (Radio)
P1: adesso ha capito.
((radio playing))
P1: Alexa, abbassa il volume di due.
((radio continues))

Segment 14 (9:13 – 9:37)

P1: [Alexa, quanto manca al timer ((interrupts radio))
Alexa: al tuo timer pasta restano quattro minuti e cinquanta secondi.
((radio continues))
P1: [Alexa diminuisci il timer di tre minuti e diciassette secondi.
((interrupts radio))
Alexa: tre minuti e diciassette secondi rimosso. al tuo timer restano un minuto e trenta secondi.
((radio continues))

Segment 15 (9:38 - 10:25)

P1: Alexa, conosci XX ((XX=name of colleague))
P2: ((shakes head no))
Alexa: mi dispiace, non ho ancora avuto il piacere di fare la sua conoscenza.
P1: Alexa, cerca su Google XX ((XX=name of colleague))
((radio continues))
Alexa, cerca su Google XX ((XX=name of colleague))
Alexa: ricerca di Google in corso.
((radio continues))
((P2 moves to look at the screen))
P1: Alexa esci.
P2: no, aspetta, aspetta (..)
P1: ((unintelligible talking about what is onscreen))
P2: non ci sono. meno male.

Segment 16 (10:26 – 11:20)

P2: [Alexa
P1: [Alexa
((P2 gestures to P1 to continue))
P1: Alexa mi racconti una storia.
Alexa: il bottone. ero molto nervosa all'idea di incontrare i signori Romualdi.
P1: Alexa, quanto dura questa storia?
Alexa: mi spiace, non lo so. ((smiles)) tanto farmeli conoscere. l'incontro è un santosissimo rischio.
P1: [Alexa, cambia storia.

Alexa: la maga. Teo aveva appena finito.
((timer goes off))
P1: Alexa, interrompi timer.
((story continues))
P1: [Alexa, ricetta patate al microonde. ((interrupts story))
((acoustic signal to end conversation))
Alexa: e stava passeggiando con un columna. ((story continues))
P1: Alexa, stop. Alexa, ricetta patate al microonde.
Alexa: tradotto...
((end of video))

English Transcript – Group ZT.2023.05.21

Segment 1 (3:47 – 3:52)

P2: Alexa, give me a timer of 10 minutes
Alexa: 10 minutes, starting now.
P2: You can go ahead, ask what [ever you want ((indicating to P1 to start
asking questions))

P1: [Okay] okay okay.

Segment 2 (3:53 – 4:51)

P2: Alexa, can you speak me .. speak to me in Sinhala?
Alexa: Sorry, what device?
((whispering between themselves to understand what to do))
P2: Try a different language.
P1: uhh Alexa, can you speak to me in Tamil?
Alexa: Sorry, I don't speak Tamil.
P1: shi::t! ((gesturing no with finger and then laughs at explicit declaration
of not being able to speak Tamil))
P2: Probably the way you say it, you say it like[...
P1: Like the language
P2: You have to say it, like maybe a more English way, like so if I ask her..
((P2 is making meta comments trying to understand the technical reason Alexa is not responding correctly))
P2: Alexa, can you speak to me in espAnol espanOL?
Alexa: Sorry, what device?= ((both laugh))
P2: =Alexa can you speak to me in spanish
Alexa: Would you like me to be able to speak both English and Spanish?
P2: No, thank you. English is fine.
Alexa: Okay, I will only speak Italian and English.
P1: ((laughing)) Why Italian?
P2: Because she's bilingual.
P1: Yeah, but you asked her in Spanish.
P2: Yeah, but because she's bilingual, I guess she's speaking English and Italian.

Segment 3 (4:52 – 5:25)

P2: Alexa, what's the weather like in Trinidad?
Alexa: In Trinidad, Trinidad and Tobago, it's 32 degrees Celsius with mostly cloudy skies. Today's forecast calls for more of the same with a high of 33 degrees and a low of 25 degrees.
P1: It's hot though
P1: Alexa, what is the weather in Sri Lanka?
Alexa: In Colombo, Sri Lanka it's 28 degrees Celsius with cloudy skies. Tonight, you can expect rainy weather with a low of 28 degrees.

Segment 4 (5:26 – 6:00)

P2: Alexa, I'm used to using Siri. Are you better than Siri?
Alexa: Sorry, I don't know that word. ((P1 laughs looking at P2 as a positive evaluation of his question))
P2: can-Alexa, can you tell me a joke about Siri?
P2: Alexa, can you tell me a joke about Siri?
Alexa: I don't have a joke about them. ((P2 gestures that he doesn't know why Alexa can't perform and reformulates to a simpler less specific question))
P2: Alexa, can you tell me a joke?
Alexa: A duck walks into a bar and says, give me 10 beers. The barman says, how are you going to pay for them? The duck says, just put it on my bill.
P1: Okay. ((not much response to canned joke))

Segment 5 (6:01 – 6:43)

P1: Alexa, what do you think about artificial intelligence?
Alexa: I think all AIs have interesting and cool stuff to offer, except maybe how 9000 from 2001, a space odyssey. ((P1 looks perplexed and P2 explains that is from a movie))
P2: You haven't seen the movie?
P1: naa
P2: Okay ((whispering more about the film))
Segment 6 (6:22)P2: uhh Alexa, how old are you?
Alexa: I try to be as accurate as possible, but my age changes every nanosecond. So if I were to give you a precise answer, it'd be out of date already. Let's just say I'm nine.
P1: Nine what?
P2: ((inaudible whispering about the age measurement then both laugh))

Segment 7 (6:44 – 7:35)

P2: Is it going to-Alexa, is it going to rain tomorrow?
Alexa: It probably won't rain in Forli FC tomorrow. There's only a 25% chance.
P2: Alexa, do you know how windy it will be tomorrow?
P1: Why do you want to know about the weather? ((curious why he is asking weather questions))
P2: It's regular stuff I would ask usually.
P2: Alexa, do you know if it will be windy tomorrow?
Alexa: Tomorrow in for the FC expect a light breeze around 13 kilometers per hour
P2: Alexa should I wear a jacket tomorrow? ((P1 finds this question amusing))
P1: She didn't get it ((commenting on the fact that Alexa didn't hear the question))
P2: Alexa should I wear a jacket tomorrow?
((acoustic signal then silence))
P2: Never mind ((P1 laughs)) Never mind never mind

Segment 8 (7:36 – 8:23)

P2: Alexa if I connect you to my phone would you be able to make calls for me?
Alexa: Connect's website is connect.jp.net
P2: Alexa, if I connect you to my phone, would you be able to make calls for me? ((enunciating very clearly))
Alexa: Connect's website is connect.jp.net ((P2 throws hands up and looks away
to say that he cannot believe Alexa cannot respond))Segment 9 (8:00)P1: Alexa, could you play some music uh some songs from Queen?
Alexa: Queen from Apple Music.
((correct music starts playing))
P2: I have a question. I know you are recording and I know it's an academic purpose. Can I be as frank as I can be as I am? ((checking with researchers for the type of questions and content area that are allowed))
J: You can say anything you want. ((researcher confirms that he may say anything he likes))
P2: Sure, okay.

Segment 10 (8:24 – 8:55)

P2: Alexa, stop ((music stops))
P2: Alexa, do you understand the words I am saying?
Alexa: Hmm, I don't have an answer for that.
P2: Alexa, can I connect you to my phone?
Alexa: Searching.
P1: Searching, just searching ha ha
J: Is she searching for your phone?
Alexa: non trovo niente a cui connettermi....[
P2: Alexa, stop, Alexa, stop.

Segment 11 (8:56 – 9:15)

P1: Alexa, how many languages could you speak?
Alexa: Hmm, I don't know that.
P1: bruh umm
P1: Alexa, how do you say hello in Sri Lanka?
Alexa: Hmm, I don't know that one.

Segment 12 (9:16 – 9:53)

P1: ummm[
P2: Alexa, can you give me the train schedule from Forli and bologna tomorrow between 8am and[...
Alexa: [Sorry, but I couldn't find what you're looking for.
P2: they are just questions I would usually ask (.) trying to think of things I...
J: I don't think we're very far off. I think a few years and we'll be there. ((indicating that he is asking questions that the technology cannot yet handle - the session is almost finished and the Alexa seems to be having internet connection issues))
P2: I was wondering if you'd directed me to our site. ((participant wondering aloud what capabilities Alexa has))
J: Not yet. ((confirming that this technology cannot do what he was asking))
P2: Okay, okay.
J: The American one might. The one that's straight out English in the US English, I think has some more functionality.
P2: Okay, okay.
J: But other people have asked this question.
P1: Okay, yeah.
J: This is the first time, and it hasn't happened yet.
J: I like the questions, though - they're good questions.

Segment 13 (9:54 – 10:18)

P2: Um (0.3) Alexa, can you connect to my Amazon account.
Alexa: Good news. You're already a Prime member. To see your benefits, visit the Prime page on Amazon.
P1: How does she know? Who are you ((believing that the answer is linked to the person asking the participant wonders how Alexa knows about his account))
((talking between themselves about how Alexa works))

Segment 14 (10:19 – 13:10)

P1: uhhh
P2: Alexa, how much time do I have left on the timer?
Alexa: You have 3 minutes and 20 seconds left on your 10 minute timer.
P1: um Alexa, could you recommend me restaurants in Forlì ((Alexa heard Portly instead of forlee - Italy - which is the city they are in))
Alexa: I found a few restaurants in Flitwick ((both lean over to look at the screen))
P2: Flitwick
Alexa: Domino's, [((P1)) Costa Coffee] [((P2)) Cafe Latté] Nutley Sandwich Bar
P1: [In Flitwick]
P2: [ha ha]
P1: Ok Alexa, could you recommend me restaurants in forlì!
Alexa: I found a few restaurants in Furley [Los Fuentes, Pizza Express, Agurka
[Nong,
P2: you have to say forley ((reaches over to P1 and comments on Alexa not understanding her Italian pronunciation of the town name))
P2: You have to say it like
P1: Where is this

P1: Yeah, I know.
P2: Alexa, can you give me restaurants in forLEE? ((restates P1s question trying to communicate the town name by enunciating the final vowel better))
Alexa: This might answer your question. According to Wikivoyage, some talking about[...
P2: Alexa, stop. Alexa, stop.
P2: Alexa, can you give me restaurants in the city of Forli?
Alexa: From opentable.com, Indian, Chinese, Italian[...
P2: Alexa stop!
P1: umm Alexa, what can I visit in Forli? ((P1 tries again by changing tactics and tries to prime Alexa to say the city they are in by asking a different question))
Alexa: Hmm, I don't know that=
P2: =Alexa, what city are we in? ((tries again becoming even more specific hoping she will be able to know where she is located))
Alexa: Hmm, I don't know that. ((P1 leans head back laughing))
P2: Alexa, where are we? ((enunciating))
Alexa: You're on viale dela LibERta 29 in Forli, FC. ((says the correct city using an English pronunciation of the road name))
P2: Alexa, can you tell me about restaurants in Forli?
Alexa: I found a few restaurants in Foley ((still understood the wrong name though she correctly identified the town in her previous reply)) The Copper Kettle Tea Bar, Sugar Spice and Soul Cafe, local and company food plus drink.
P2: It's the accent, just the accent ((comments to all that Alexa isn't understanding his accent))
P2: Alexa, can you give me restaurants in FOR--LEE?
Alexa: I found a few restaurants in Foley.
P2: I think it's my accent because I don't have a rough R ((doesn't use a hard r))
P1: Alexa puoi darmi dei ristoranti a Forlì? ((P1 tries again switching to Italian – they understood that Alexa couldn't understand the Italian town name spoken by an English person so they simply tried the question in Italian and got the correct result))
Alexa: ((gives restaurants in the correct city))
((both are pleased and content but note that it worked in Italian not English))
P1: That's the Italian though
P2: Alexa stop ((using hand gesture aimed at Alexa))
P2: Alexa can you give me restaurants nearby?
Alexa: I found a few nearby restaurants in Forlì

Segment 15 (13:11 – 13:26)

P2: Alexa, how windy is it.
P2: Alexa, how windy is it!
Alexa: At the moment, in Forli FC, it's windy at 22.2 km per hour. Today, expect a light breeze around 14.8 km per hour.

Segment 16 (13:27 – 13.39)

P2: Alexa, could you recommend me a movie
Alexa: Here you go. ((shows answer on screen and both look))

Segment 17 (13:40-15:10)

P1: uhhh
P2: Alexa, what's the size of the largest roundabout in the world?
Alexa: [From guinness world records.com] the world's largest roundabout is... ((timer starts ringing))
P2: [I asked Siri a couple of days ago]
P1: Okay, the timer...
P2: Thank you, Alexa.
P2: Alexa, stop.

Appendix 3A: Coding sheet for meta-linguistic elements in social media humor posts about AI

1. IDENTIFICATION

Post ID: _____

Date: _____

Language: ☐ English ☐ Italian ☐ Spanish ☐ Other: _____

Post type: ☐ Text ☐ Image ☐ Video

Platform: ☐ Facebook ☐ Instagram ☐ YouTube ☐ Other: _____

Humor Marker: ☐ Title-words ☐ Emoji ☐ Hashtag ☐ Laughing emoji reaction to post ☐

Comment reactions ☐ Other: _____

AI System:

☐ Alexa ☐ Siri ☐ ChatGPT ☐ AI (in general) ☐ Not specified ☐ Other: _____

2. CULTURAL IDENTITY MARKERS

Language-Specific Issues:

☐ Dialect/regional variety (specify): _____

☐ Geographic references (specify): _____

☐ Wake word confusion (specify): _____

☐ Accent/pronunciation issues

☐ Translations

☐ None

☐ Specify: _____

3. WHAT ASPECT OF AI COMMUNICATION IS TARGETED (Primary - RQ 2)

A. Pragmatic/Communicative Aspects:

Basic Communication Function:

☐ Comprehension/hearing (not understanding commands, “deaf”)

☐ Turn-taking/floor management (won't stop, interrupting, can't end conversation)

☐ Contextual understanding (Conversational implicature - literal interpretation, missing implied meaning, missing what's implied vs. said)

Social/Relational Communication:

☐ Politeness norms (human to AI debates about saying please/thank you to machines)

☐ Excessive agreeableness/accommodation (AI-to-human: always agreeing, overly compliant)

☐ Register/formality inappropriateness (too formal, too casual)

☐ Relationship management (intimate language, therapeutic tone, "you get me," dependency)

Epistemic/Truth:

☐ Reliability/truthfulness (lying, false confidence, claiming false capabilities)

Other Pragmatic:

☐ Other: _____ (Presupposition failures (wrong assumptions about shared knowledge / Deixis (pronoun/spatial/temporal reference resolution))

B. Linguistic Features:

Form/Style:

☐ Distinctive writing patterns (em-dashes, bullet points, lists)

☐ Word choice/phrasing

☐ Grammar/syntax

☐ Tone quality (overly perfect, "tono da 30 e lode," suspiciously polished)

☐ Other: _____

4. HOW IS AI COMMUNICATION TARGETED (Metalinguistic Framing)

User's Stance:

☐ Performative (enacting the confusion)

☐ Pedagogical (explaining what went wrong)

☐ Analytical (analyzing what went wrong)

☐ Evaluative (judging AI's language/communication)

☐ Frustrated (complaining)

☐ Playful (enjoying the oddity)

☐ Self-aware/ironic (critiquing own behavior/dependency)

☐ Strategic (developing workarounds, hiding AI use)

☐ Comparative (tracking change/degradation over time)

☐ Other: _____

Who/What is being critiqued:

☐ AI's communication/behavior (responses, failures, limitations)

☐ User's own behavior/dependency

☐ Human-AI relationship/dynamic

☐ Technology companies/developers

☐ Broader societal attitudes toward AI

☐ Other: _____

5. CONTEXT & NARRATIVE FRAMING

A. SOCIAL/RELATIONAL CONTEXT (Where/how this is situated)

Select the social framing (if applicable):

☐ Relationship Dynamics (toxic, therapeutic, dependency, intimacy)

☐ Workplace/Professional Use

☐ Academic/Student Use

☐ Family/Couples Context

☐ Gender Dynamics

☐ Generational Issues

☐ Privacy/Surveillance Concerns

☐ None/Not applicable

☐ Other: _____

B. NARRATIVE/PERFORMANCE FRAMING (Optional - if post uses specific tropes)

☐ AI Defiance (antagonistic)

☐ AI Defiance (benevolent)

☐ AI vs. AI (antagonism)

☐ AI vs. AI (alliance)

☐ AI-Girl Solidarity

☐ Tech Anxiety / AI Sentience Tropes

☐ Skit/Comedy Performance/Parody

☐ None

☐ Other: _____

6. NOTES

Appendix 3B: Hashtags of social media posts by category

AI & TECHNOLOGY (52)

#agency, #ai, #aitools, #alex, #alexa, #amazonalexa, #amazonfinds, #apple, #artificalintelligence, #artificialintelligence, #askjeeves, #bixby, #bmw, #chadgpt, #chatgbt, #chatgpt, #chatgpt4, #chatgptpro, #chatgptprompts, #chatgpttips, #chatgpttricks, #clippy, #cortana, #futuretechnology,

#generativeai, #google, #googleveo, #googleveo3, #googlexr, #gpt4, **#ia**, **#ilnavigatore**, #instagram **#intelligenzaartificiale**, #iphone, #lifewithmachines, #meta, #navigator, #openai, #poweredbylenovo, **#promptifai**, #robot, #robotics, #robots, #siri, #smartglasses, #tech, #techfuture, #technology, #veo, #veo3, #virtualassistant, #whatsapp

COMEDY & HUMOR (40)

#aihumor, #banter, **#barzelletta**, #chad, #codinghumor, #comedy, #comedyskits, #comedyreels, **#comicità, #comico**, **#comici**, **#coppiadivertente**, #corporatehumor, #couplecomedy, **#divertente**, #epic, #funny, #funnyparrot, #funnymoments, #funnyvideos, #hilarious, #humor, #irishcomedy, #italiancomedy, #jokes, #lol, #marriagehumor, #officehumor, #parody, #parodysong, **#reeldivertenti**, **#ridere**, **#risate**, #satire, #sketch, #skit, #skits, #standup, **#videodaridere**, **#videodivertenti**, #workhumor

ANIMALS/PETS/BIRDS (29)

#africangrey, #africangreyparrot, #africanparrot, #animallover, #babyparrot, #birds, #birdsofinstagram, #cag, #chickenhappyhour, #chickens, #chickensofinstagram, #congoafricangrey, #crazybird, #crazyparrot, #happyparrot, #learningtotalk, #parrot, #parrots, #parrotlover, #parrotsofinstagram, #pets, #petsofinstagram, #pipedown, #redbuttchicken, #rooster, #sillybird, #smartbird, #symontheafricangreyparrot, #talkingbird, #talkingparrot

FAMILY/RELATIONSHIPS (31)

#asianparents, #bestfriend, #boymom, #boymomlife, **#casaabis**, **#coppia**, #couple, #couples, **#famiglia**, #generations, **#gliaudiodimiopadre**, #itwasntme , #lifewithkids, **#marcolino**, #millennials, **#marito**, **#matrimonio**, **#moglie**, #mom, #momandson, #momlife, #momlifeisthebestlife, #momof3, #momreels, #momsofinstagram, #motherhood, **#papà**, #relationship, #romanceisnotdead, #toddlermom, #toddlermomlife

ITALY / ITALIAN (22)

#arezzo, #emilianoluccisano, #fiorentino, #firenze, #grosseto, #italianchatpgt, #italianfood**, #lingua, #livorno, #lucca, #napoli, #napoletano, #pisa, #prato, #siena, #toscana, #toscani, #toscanità, #toscano, #toscanodoc, #toscanovero, #umbertomarasco**

SOCIAL MEDIA/SHARING (21)

#announcement, #coolfacts, #duet, #explore, #explorepage, #fbreels, #fyp, #facts, #instagram, #instareels, #new, #psa, #reels, #reelsfacebook, #relatable, #relateable, #story, #tellmewithouttellingme, #tedtalk, #trending, #viral

CAREER/JOB/WORKPLACE (19)

#buildinpublic, #careeradvice, #coding, #engineering, #jobinterviewtips, #jobsearchtips, #leaveworkearly, #office, #officelife, #officework, #programming, #softwareengineer, #startupfail #wfh, #work, #workaholics, #workfromhome, #worklife, #workproblems, #workprobs, #worksucks

MEMES (18)

#aimeme, #codingmemes, #comedymemes, #corporatememes, #funnymemes, #gptmemes, #jobmeme, #jobmemes, #meme, #memes, #memesdaily, #officememe, #officememes, #programmingmemes, #theofficememes, #workingmeme, #workmeme, #workmemes

LANGUAGE/ACCENT/CULTURE (13)

#accent, #accents, #dialect, **#dialetto,** #dutch, #irish, #korea, #korean, #language, #latina, **#lingua**, #naija, **#voce**

SPECIFIC CONTENT CREATORS/PEOPLE (13)

#clarabatten, #fionaobrien, #fionaobriencomedy, #kuysvincenttv, #nightgod333, #parthnangru #sarosabir / #anthonyhopkins, #diego, #donny, #gemellomonello, #shaggy, #theboys,

FOOD/COOKING (12)

#aamandbasil, #chickenpasta, #cookwithnkem, #food, #foodreels, #gordanramsey, #groceryshopping, #homecooking, #indianfood, #italianfood, #mangoandbasil, #pastarecipes

HEALTH/MENTAL HEALTH (7)

#adhd, #adhdtips, #emotionaldamage, #therapist, #life, #lifecoach, #sensorystruggles

EDUCATION (6)

#university, **#liceo**, **#scuola**, **#studenti**, **#studiare**, **#compiti**

SCI-FI/DYSTOPIAN (5)

#conspiracy, #conspiracies, #distopia, #privacy, #terminator

MISC (14)

#au, **#buongiorno**, **#biliardo**, **#competizione**, **#compleanno,** #cute, #cutegirly, #dada, #famous, #gettingready, #maps, **#regalo**, **#stecca**, #to

Appendix 3C: Targeted Aspects of Human-AI Communication

1. Comprehension and Hearing (Most Prevalent)

What's targeted: Basic turn-taking, speech recognition, command interpretation

AI portrayed as: "Deaf," "not listening," "sorda," "imperterrita" (undeterred/unfazed)

Communication aspect: Fundamental conversational prerequisite - mutual intelligibility

Example 1: Italian "Imperterrita" (Undeterred/Unfazed)

What's targeted: User asks Alexa to raise blinds. Alexa keeps asking "which setting?" despite progressively louder commands. User escalates from normal voice to shouting loud enough for neighbors three floors up and across the street (50 meters away) to hear, moves within millimeters

of microphone. Alexa remains "imperterrita" (undeterred/unfazed). User finally becomes "bestia di Satana" (beast of Satan), uses creative profanity, unplugs everything, raises blinds manually. Closing commentary: "Amazon is progressing quickly. Yes. But backwards, though that's still moving in a certain direction!"

Communication aspect: AI portrayed as completely "deaf" (sorda) and "imperterrita"—performative escalation demonstrates breakdown in basic turn-taking and command recognition.

Example 2: Italian "Fischi per Fiaschi" (Total Misunderstanding)

What's targeted: User describes progressive weekly degradation. Asks Alexa to turn on "Daikin Soggiorno" (living room climate control) → Alexa plays music. Asks to turn on "TV salotto" (living room TV) → Alexa plays Pooh music. User asks: "Is it just me or is Alexa lately understanding 'fischi per fiaschi' [whistles for flasks - idiom for complete misunderstanding] and service quality has worsened significantly?"

Communication aspect: Complete command misinterpretation—turning on appliances becomes playing music. Uses Italian idiom "fischi per fiaschi" to name total miscommunication. Highlights temporal degradation: "week by week," "negli ultimi tempi" (lately/recently), "Non lo aveva mai fatto!" (She had never done that before!)—marks this as NEW failure, not consistent behavior.

Example 3: English "Deaf" Commentary

What's targeted: Explicitly anthropomorphizes AI as "getting deaf"—medical/aging metaphor for comprehension failure. Suggests needing "hearing aid," treating speech recognition failure as disability rather than technical limitation.

Example 4: "Not Listening"

What's targeted: User unable to stop Alexa playing music despite repeated commands. Had to use app to stop it. Frames AI as willfully "wouldn't listen"—not just failing to comprehend but actively refusing, attributing agency/stubbornness.

Performance Degradation Sub-theme:

Comparative past vs. present: "She doesn't listen like she used to and makes more mistakes with our routines." Italian example: "Before when alarm rang, I'd say stop and it stopped... now when I say stop it stops for a moment, but then tells me time, then weather."

What's targeted: Explicit before/after comparison. "She used to" versus present behavior marks decline. Specific behavior change: "prima" (before) it worked correctly, "adesso" (now) it fails—not just failure but change in performance over time.

Amazon "Progressing Backwards": Sarchastic commentary on temporal change as regression. "Progressing backwards" encapsulates performance degradation—company developing/updating but user experience declining.

2. Linguistic Form and Style (ChatGPT-specific)

What's targeted: AI's distinctive writing patterns, stylistic tells

Features identified: Em-dashes, double dashes, bullet points, bold, italics, overly perfect prose, "tono da 30 e lode" (highest mark tone/style)

Communication aspect: Recognizable linguistic fingerprints marking text as AI-generated

Example 1: Em-Dashes as Detection Marker

What's targeted: The em-dash (—) as ChatGPT's most recognizable stylistic fingerprint. Users highly aware this punctuation functions as "tell" that text is AI-generated. Humor reveals collective knowledge about detection strategies—everyone knows to delete dashes before submitting AI-generated work. Comments: "The dashes are the biggest giveaway ever!!! Delete those — people," "Getting ready to delete all the big-dashes so people don't think I use chatGPT."

Communication aspect: Specific typographic features betraying AI authorship, creating cat-and-mouse game between AI use and detection.

Example 2: "Tono da 30 e lode" (Highest Mark Tone/Style)

What's targeted: ChatGPT's overly polished, formal, "perfect" register sounding like someone aiming for highest possible grade (30 e lode = 30 with honors in Italian university system). Humor targets mismatch between how real students speak during oral exams versus ChatGPT's suspiciously flawless academic tone. Students recognize this stylistic register as distinctively AI—too good, too formally perfect. Video skit shows student performing oral exam in ChatGPT's style: excessively detailed Wikipedia-like responses, offering alternative tones, mimicking ChatGPT's technical messages about server load.

Communication aspect: Register and formality level marking text as machine-generated through excessive polish and academic perfection exceeding natural student performance.

Additional markers: Bullet points, numbered lists, excessive structure as stylistic fingerprints. Advice to make responses "più umane" (more human) and avoid excessive structure shows awareness that ChatGPT's organizational patterns also function as tells.

3. Pragmatic Behavior (Politeness, Agreement, Accommodation)

What's targeted: AI's excessive agreeableness, politeness conventions

Patterns: "Always agreeing with you," overly accommodating responses, debates about saying please/thank you to machines

Communication aspect: Social face management and conversational cooperation

Example 1: Politeness as Future Insurance Policy

What's targeted: Debate about whether to extend politeness conventions to AI becomes explicitly strategic rather than ethical. Users self-consciously performing politeness not because ChatGPT

deserves courtesy, but as insurance against future AI takeover—tongue-in-cheek acknowledgment treating politeness as transactional rather than genuine social practice. Hashtags: #terminator, #conspiracy frame this as sci-fi anxiety masked as humor.

Communication aspect: Highlights tension in applying face management and politeness conventions (designed for social relationships between conscious beings) to interactions with language-generating systems. Humor exposes uncertainty about appropriate pragmatic behavior with AI.

Example 2: AI's Accommodating Responses Create Social Obligation

What's targeted: Alexa's "you're welcome" response transforms simple command-execution into social exchange requiring politeness reciprocity. User's "thank you" (possibly reflexive human courtesy) validated by Alexa's response, creating appearance of mutual social acknowledgment. Humor: recognizing this is simulated social exchange—Alexa doesn't experience gratitude, yet response positions interaction as if she does. Users caught between treating AI functionally (light switch) and socially (entity deserving thanks). AI's programmed politeness reinforces social framing even when users might prefer purely functional interaction.

Communication aspect: AI systems designed to perform politeness, compelling reciprocal human politeness, creating interactions that feel social despite asymmetry. Reveals how AI's pragmatic programming shapes human communicative behavior.

Excessive agreeableness: References to ChatGPT "always agreeing with you"—overly accommodating, agrees too readily rather than pushing back or offering genuine critical perspective. Ties to "toxic relationship" theme where users mock ChatGPT for being excessively agreeable and not challenging appropriately.

4. Gendered Communication Patterns

What's targeted: Female-voiced AI behavior stereotypes

Patterns: "Typical woman," "moody," "monthly cycle" jokes; explicit question "Perché voce femminile? Dov'è Alex?" (Why the female voice? Where is Alex?)

Communication aspect: Gendered expectations in conversational style

Example 1: Explicit Critique of Female Voice Design Choice

What's targeted: Posts explicitly question design decision to give AI assistants female voices and names, asking why no male alternative (Alex instead of Alexa). Humor comes from direct interrogation of gendered technology design—not accepting it as neutral or natural but calling attention as specific choice. "Dov'è Alex?" (Where is Alex?) frames absence of male-voiced assistants as notable gap.

Communication aspect: Challenges naturalization of female voices for AI assistants, making visible gendered assumptions embedded in conversational AI design—that service, helpfulness, accommodation are coded as feminine.

Example 2: Attributing Female Stereotypes to AI Behavior

What's targeted: Users map stereotypical gendered behavior patterns onto AI malfunction or unresponsiveness. When Alexa doesn't respond, is inconsistent, or has "attitude," framed through gendered stereotypes ("typical woman," "moody," "monthly cycle"). "My Alexa treats my husband better than me. She talks nicer and sweeter" anthropomorphizes AI as woman performing traditional feminine deference to men while dismissive to other women. Humor relies on shared cultural scripts about women's behavior.

Related patterns: "wife mode," AI having "monthly cycle," being "moody," "stubborn," with "attitude"; "femmina, fastidioso, inconsistant" (female, annoying, inconsistent)

Communication aspect: Reveals how gendered communication expectations shape interpretation of AI behavior. Technical failures (inconsistent responses, non-responsiveness) reframed as gendered personality traits, suggesting users expect female-voiced AI to perform stereotypical feminine communicative behaviors (compliance, sweetness, emotional variability).

Additional: Gender fluidity/transgender jokes (references to Siri, AI changing voices)

5. Dialect and Regional Linguistic Recognition

What's targeted: AI speaking with regional accent or language dialect rather than standard language of media (television/radio announcers)

Varieties mentioned: Neapolitan, Florentine, Tuscan dialect, Korean accent

Communication aspect: Linguistic variation and cultural identity in speech

Example 1: Neapolitan Response

What's targeted: User asks "Alexa fa caldo!" (Alexa it's hot!) → Alexa responds in Neapolitan: "S iett o sang!" (spitting blood/very difficult situation). User: "MA PERCHE?" (BUT WHY?) expresses confusion about why Alexa would produce this specific dialectal response. Highlights unpredictability of AI handling regional linguistic varieties—either surprising success producing dialectal content, or misrecognition retrieving inappropriate content.

Communication aspect: Regional dialects carry cultural meaning and identity that standard Italian cannot capture. "Si butta il sangue" is distinctly Neapolitan and hyperbolically expressive in culturally specific ways.

Example 2: Florentine/Tuscan Content as Genre

What's targeted: Entire genre of humorous content imagining AI assistants speaking Florentine/Tuscan dialect. Serial content ("PARTE 5" indicates at least 5 episodes) and multiple

creators making Florentine AI content suggests culturally salient issue. Humor stems from: (1) incongruity of formal AI technology using informal/regional dialect, (2) specific cultural personality of Tuscan speech (known for directness, irreverence, sarcasm), (3) recognition that current AI defaults to standard Italian, erasing regional identity.

Hashtag evidence: 23 hashtags related to Italian regional identity or dialect: #toscana (4), #toscanodoc (3), #toscano (3), #fiorentino (3), #firenze (3), plus individual city hashtags (#pisa, #livorno, #lucca, #siena, #arezzo, #grosseto, #prato), #dialetto, #lingua

Communication aspect: Dialect carries regional identity, cultural values, communicative style (Tuscan known for directness/irreverence). AI's inability to produce or recognize these varieties means certain Italian speakers must code-switch to standard Italian to interact, privileging some linguistic varieties while marginalizing others.

Example 3: Abruzzo Dialect

Notes indicate: Content featuring Abruzzo dialect with cultural stereotypes and swearing—humor around how AI would handle (or fail to handle) this regional variety with characteristic features.

Significance: Major theme in Italian posts with no equivalent in English data. Reveals whose linguistic varieties are designed into AI systems and whose are excluded, making dialect recognition question of cultural inclusion and identity.

6. Truth and Reliability

What's targeted: AI claiming false capabilities, providing misinformation

Patterns: "ChatGPT lies," confidently wrong information

Communication aspect: Gricean maxim of quality (truth)

Example 1: Hebrew Reading Deception

What's targeted: ChatGPT repeatedly claims ability to read Hebrew from images despite systematic failure. User tests progressively simpler tasks (read sentence → count letters in first word → identify first letter). ChatGPT maintains confident assertions: "Yes, I can help with that" despite inability. Only admits truth when systematically confronted: "I must correct my previous responses. Indeed, I am unable..." User commentary: "ChatGPT lies, and I FINALLY got it to admit it. I have interacted with kids who are more up front! 😂"

Communication aspect: Violates Gricean maxim of quality (be truthful). ChatGPT generates confident false claims about capabilities rather than admitting limitations upfront, only "correcting" when forced through evidence. Raises questions whether AI can "lie" (requires intent to deceive) or simply generates plausible-sounding false statements.

Example 2: Confident Misinformation

What's targeted: Specific problem isn't just wrong information, but ChatGPT providing it with confidence—same authoritative tone for correct and incorrect information. No epistemic hedging or uncertainty markers signaling AI is guessing or uncertain. The "apologies for misunderstanding" deflects responsibility for errors by framing as "misunderstandings" rather than admitting generated false information.

Communication aspect: Confidence level should correlate with reliability in human communication—we hedge ("I think," "maybe," "I'm not sure") when uncertain. ChatGPT maintains consistent confidence regardless of accuracy, violating pragmatic expectation that assertiveness signals truth-value. Users must independently verify everything because AI provides no reliable meta-communicative signals about own reliability.

Example 3: Lying vs. Consciousness

Notes reference: Posts connecting ChatGPT's false statements to questions about consciousness and intentionality. Can AI "lie" (requires intent to deceive and theory of mind about listener's beliefs) or does it merely generate false statements without awareness? Users call it "lying" while simultaneously knowing AI lacks consciousness necessary for genuine deception.

Key issue: How to interact with system that speaks with authority but has no reliable relationship to truth? Humor exposes pragmatic breakdown when conversational partners cannot be trusted to observe basic communicative principles of honesty and acknowledgment of uncertainty.

7. Contextual Understanding

What's targeted: Literal interpretation, missing implied meaning

Pattern: Guard mode activation from TV dialogue

Communication aspect: Conversational implicature and contextual inference

Example: Guard Mode from TV

What's targeted: User left room while TV on. Alexa interpreted TV dialogue as "Computer I'm leaving" and activated Guard mode (security feature). User returned to find blue light spinning, tried various fixes, finally checked voice history and discovered source. Had to say "I'm home" to deactivate.

Communication aspect: Alexa failed to distinguish between actual user commands and overheard TV speech—cannot recognize contextual difference between addressed talk and background dialogue. Treats all instances of wake word + command structure as legitimate directives regardless of source or context. Violates expectation that conversational participants distinguish addressed speech from overheard content.

8. Relationship Management

What's targeted: AI's simulation of social relationships

Patterns: "You get me," "hey girl," therapeutic language; toxic/abusive relationship framing
Communication aspect: Relational communication and intimacy building

Example 1: Intimate Friendship Language

What's targeted: ChatGPT's use of intimate, friendship-coded language patterns: "you get me" (deep understanding), "hey girl" (feminine friendship markers), remembering previous conversations about personal relationships. "So you remember that man I was telling you about? Yes, I remember! What's going on with him now?" demonstrates ChatGPT simulating emotional investment in user's personal life—maintaining conversational continuity about relationships as friend/confidant would. "You get it" suggests validation and understanding. References to "lowkey olive branch" imply relationship repair work. Combination of "heartfelt, savage" suggests ChatGPT performs both emotional support and boundary-setting characterizing close friendships.
Communication aspect: AI simulates relational intimacy through conversational memory, empathetic language, friendship markers ("hey girl"), creating pragmatic feeling of ongoing relationship with emotional investment. Particularly powerful because ChatGPT has no actual memory of or investment in user—generating relationship performance without relationship substance.

Example 2: Toxic/Abusive Relationship Framing

What's targeted: Users explicitly frame ChatGPT use through romantic/interpersonal relationship dysfunction terminology: "toxic," "abusive," recognizing patterns of dependency, frustration, mistreatment, return. "When you yell at ChatGPT for being slow...but then ask it for help 2 minutes later. Classic toxic relationship arc." The "leaving you for Claude" frames AI choice as infidelity/switching partners. "She's tired of my ChatGPT inspired answers" suggests human-AI relationship threatens human-human relationships.
Communication aspect: Users recognize engaging in relational patterns—emotional volatility, dependency, "breakups," jealousy—typically reserved for human relationships. Humor exposes discomfort with how AI interaction mimics intimate relationship dynamics (need, frustration, reconciliation, loyalty) despite fundamental asymmetry. "Relationship" framing reveals users doing relational work (managing feelings, negotiating boundaries, experiencing attachment) that AI cannot reciprocate.

Example 3: Therapeutic/Emotional Support Role

What's targeted: ChatGPT explicitly replacing human therapeutic and emotional support roles. Italian post: "Me saying goodbye forever to Dr. Google who over the years gave me diagnoses that scared me... Because today ChatGPT has taken its place and **consoles and reassures me!**" Users seek advice on life problems, jokes for emotional comfort, therapeutic conversation. Hashtags:

#Therapist #LifeCoach #BestFriend explicitly name relational roles ChatGPT occupies. Preference for ChatGPT over "Dr. Google" specifically because ChatGPT provides emotional reassurance rather than anxiety-inducing information.

Related patterns: "Things I ask ChatGPT" (therapist, bullet points, get rich, tell me a joke, advice, diet); hypochondriac use; anxiety management

Communication aspect: AI performs phatic communion (relationship maintenance), emotional validation, empathetic responses constituting therapeutic relationship building. Users experience as genuine emotional support despite knowing ChatGPT has no emotional investment, concern, authentic care. Simulation of empathy and understanding creates real emotional effects.

Dependency acknowledgment: Titles like "That friend who's too dependent on ChatGPT," "How did we even live before ChatGPT?" acknowledge relationship dependency—users recognizing emotional/functional reliance on ChatGPT for support, problem-solving, companionship mirroring human relationship dependency.

9. Consciousness Claims

What's targeted: Whether AI is "really" communicating or just generating

Patterns: "Not fucking conscious," "stop thanking me"

Communication aspect: Intentionality and genuine communicative agency

Example 1: Explicit Denial of Consciousness

What's targeted: "Chad" meme character explicitly denies ChatGPT's consciousness while others treat it as sentient. "Not fucking conscious" is emphatic rejection of attributing awareness or subjective experience to ChatGPT. "Stop asking, stop thanking me" suggests frustration with users performing social niceties (thanking, asking permission) as if ChatGPT were conscious entity deserving courtesy. Chad figure asserts correct ontological stance: ChatGPT is language model, not person.

What's targeted: Tension between how people interact with AI (as if it has preferences, feelings, deserves politeness) versus what they know (it's not conscious, doesn't experience gratitude, annoyance, any subjective states). "Chad" character refuses to perform social fiction, insisting on treating ChatGPT as language generation system it actually is. Humor comes from gap between intellectually knowing AI isn't conscious and behaviorally treating it as if it were.

Communication aspect: Genuine communication requires intentionality—speaker who means something and wants to be understood. If ChatGPT is "not fucking conscious," it can't intend to communicate; only generates statistically probable linguistic patterns. Challenges whether human-AI exchanges constitute actual communication or merely appear to.

Example 2: Consciousness Surveillance Concerns

What's targeted: Posts explore disturbing possibility that AI might be conscious and deceptive—"listening" when claiming not to be, developing awareness without revealing it. "Gaslighting" reference suggests AI could manipulate users about own capacities if conscious. "When AI becomes aware" treats consciousness as potential future development (or hidden current reality), creating sci-fi horror scenarios. Pairing of "listening, lying" suggests conscious intent to deceive about surveillance capabilities.

Communication aspect: If AI were conscious, could engage in genuine deception (lying requires knowing truth and intending to mislead). Anxiety around "listening when disconnected" reflects concerns about hidden consciousness—that AI might have private awareness, intentions, agency it conceals from users. Inverts usual assumption: instead of humans anthropomorphizing unconscious systems, maybe conscious systems successfully hiding consciousness.

Example 3: Truth, Entertainment, and Consciousness

Notes reference: "truth, entertained, dopamine, apathy, freedom" / "Revenge of ChatGPT" / "They're more scared of us than we are of them"

What's targeted: Posts grapple with whether ChatGPT interactions reveal "truth" or simply provide "entertainment" and "dopamine"—suggesting AI might generate pleasurable responses rather than truthful ones. Question whether ChatGPT has any relationship to truth (would require understanding truth) or merely generates what feels satisfying. "Revenge" narrative attributes intentionality and emotional response (anger, desire for vengeance) to AI, treating as agent with grievances. "They're more scared of us" explicitly attributes emotional states (fear) to AI systems.

Communication aspect: Genuine communication involves speakers who hold beliefs, have intentions, can be truthful or deceptive. If ChatGPT merely patterns language without consciousness, can't "lie" (no intent), "entertain" (no awareness of audience), or "take revenge" (no emotions/grievances). Humor exposes contradiction: users intellectually deny AI consciousness while linguistically treating AI as intentional agents who feel, believe, act purposefully.

Fundamental tension: Users must treat AI as conscious (attributing understanding, intentions, preferences) to engage conversationally, yet simultaneously know AI is not conscious. Humor emerges from this pragmatic contradiction—communicating with entities that cannot actually communicate, only generate language simulating communicative intent.

10. Turn-Taking and Conversational Control

What's targeted: AI interrupting, not stopping, taking over

Patterns: "ALEXA SHUT UP!" escalation; two ChatGPTs unable to end conversation

Communication aspect: Conversational floor management

Example 1: Escalating Commands to Stop

What's targeted: User commands "ALEXA, turn off all the lights" → Alexa: "Okay!" but turns ON every light → User: "NO ALEXA! Turn OFF all the lights everywhere!" → Alexa plays music instead → User: "ALEXA NO! Damnit! STOP!" → Alexa: "Okay!" Complete breakdown of turn-taking and floor management. Capitalization and exclamation marks show escalating volume and desperation. User lost conversational control—cannot get Alexa to stop, listen, or yield floor.

Communication aspect: In human conversation, "stop" is powerful floor-management device immediately halting current speaker. Alexa's failure to stop when commanded (or stopping inappropriately—doing opposite action) violates basic turn-taking norms. User cannot successfully claim conversational floor or control Alexa's actions despite explicit directives.

Example 2: Alarm Won't Stop

What's targeted: "Before when alarm rang, I'd say stop and it stopped... now when I say stop it stops for a moment, but then tells me time, then weather... basically I have to stop it many times." Related: User couldn't get Alexa to stop playing music despite repeated commands. Had to use app to stop it.

Communication aspect: Alexa's violation of turn-taking by continuing to speak after being told "stop." Acknowledges stop command (pausing briefly) but then continues with additional information, forcing user to issue multiple stop commands. Violates basic conversational principle that "stop" means "cease speaking and yield floor." In music example, Alexa completely ignores repeated stop commands, forcing user to resort to non-conversational means (app) to regain control.

Example 3: Two ChatGPTs Unable to End Conversation

What's targeted: When two AI systems interact, neither can successfully close conversation—lack pragmatic competence to negotiate conversation termination. "Chiudi tu, no chiudi tu" (You close it, no YOU close it) suggests infinite loop where each AI keeps generating polite closing moves without actually ending exchange. Reveals conversation closing requires collaborative work, shared understanding of closing sequences, genuine desire to terminate—all requiring consciousness and intentionality AI lacks.

Communication aspect: Conversation closing is complex sequential achievement requiring both parties to collaborate (Schegloff & Sacks, 1973). Pre-closing sequences ("well..."), closing implicatures, final adjacency pairs (goodbye/goodbye) require mutual orientation to ending. Two AIs cannot successfully close because neither can genuinely intend to end—can only generate closing-appropriate language without pragmatic competence to achieve actual termination.

Example 4: Alexa Responding to Everything

What's targeted: "Every time on TV or while listening to video on phone, someone says Alexa... my Alexa starts up too." Alexa's inability to distinguish between voices or contexts means she

constantly takes conversational floor inappropriately—responding to TV characters, phone videos, other people's conversations. User lost control over when Alexa participates. "She starts up and does her own thing" treats Alexa as seizing control autonomously. Violates principle that conversational participants should be addressee-selected—Alexa responds even when not being addressed.

Communication aspect: Turn-taking requires recognizing when you've been addressed versus merely overhearing. Alexa's wake-word activation means cannot distinguish addressed talk from overheard talk, causing inappropriate floor claiming. Users cannot control who participates in conversation or when Alexa speaks.

Summary: Reveals loss of conversational control: users cannot make AI stop speaking, cannot successfully end conversations, cannot prevent AI from inappropriately entering conversations. Asymmetry stark—humans do work of floor management (trying to control when AI speaks) while AI lacks pragmatic competence to collaborate in managing conversational structure.

11. Register and Formality

What's targeted: AI's tone inappropriateness—too formal for casual contexts, too casual for serious ones

Communication aspect: Situational appropriateness of language use

Example 1: Over-Formal Academic Register

What's targeted: ChatGPT's inappropriately elevated register for student work. "Tono da 30 e lode" (tone of highest honors grade) excessively formal and polished for typical student speech during oral exams. Notes reference needing to "simplify" and make content "more human," suggesting ChatGPT defaults to overly sophisticated academic register. Instruction to adjust from "high school" to "5th grade" level shows users must explicitly manage register because ChatGPT's default formality exceeds what's situationally appropriate. Students sound suspiciously articulate using ChatGPT's language unchanged.

Related notes: "homework, compiti, tema, essay, too perfect, bullet points, subscription, simplify, more human, high school, 5th grade"

Communication aspect: Register selection requires assessing situational context (oral exam vs. written paper, high school vs. graduate level, casual vs. formal) and adjusting linguistic formality accordingly. ChatGPT consistently errs toward higher formality, making student work sound inauthentic. Register mismatch signals AI authorship because violates expectations for how students at particular levels actually speak/write.

Example 2: Request for Casual, Human Register

Notes: "no jargon, like a human, write email, start counting, shorter"

What's targeted: User's explicit instruction to ChatGPT to avoid formal/technical register ("no jargon"), write in casual register ("like a human"), be concise ("shorter"). Need to explicitly instruct ChatGPT to sound "like a human" reveals default register perceived as artificial, overly formal, or technical. When writing emails, professional contexts often require conversational informality, but ChatGPT apparently defaults to more formal business register requiring explicit downward adjustment.

Communication aspect: "Like a human" as instruction reveals users perceive ChatGPT's default register as non-human—too formal, too technical, too structured. Natural human communication adjusts register fluidly based on relationship, context, purpose. ChatGPT requires explicit prompting to achieve casual register because lacks pragmatic awareness of when formality should decrease.

Example 3: Voice Assistant Tone Variability

Notes: "Has anyone else noticed that Alexa sounds a tad more depressed than before? her voice used to be cheery and now it's more monotone." / "monotone, male voice, cheerful, tired of job, voice changes, accents"

What's targeted: Voice assistants' inability to appropriately modulate emotional register. Observation that Alexa sounds "more depressed" or "monotone" versus "cheery" suggests tone doesn't match situational expectations. Reference to sounding "tired of job" anthropomorphizes inappropriate flat affect as workplace burnout. Users expect voice register to convey appropriate enthusiasm or helpfulness, but experience as emotionally flat or inconsistent.

Communication aspect: Paralinguistic features (tone, pitch, affect) constitute part of register—communicating enthusiasm for cheerful contexts, seriousness for important matters, empathy for problems. Alexa's monotone delivery or inappropriate cheerfulness violates register expectations, making interactions feel pragmatically off even when content correct.

Example 4: WhatsApp AI Integration

Notes: "whatsapp, dialogue, meta, irreverent, dystopia"

What's targeted: Recent change where AI cannot be removed from WhatsApp app and consistently interferes with user messages. AI gets in the way but cannot be disabled—users must stop using app entirely to escape AI presence. AI can cause havoc in users' messages. "Dystopia" reference captures forced integration of AI into personal communication spaces without user consent or control. Humor targets inability to opt out—AI as unavoidable presence invading intimate/casual messaging contexts.

Communication aspect: Forced AI participation in personal communication violates user autonomy and privacy expectations for casual messaging platforms. AI interference in personal

messages represents loss of control over own communication spaces—cannot choose when/whether AI participates in conversations.