"Alexa, Do You Want to Build a Snowman?" Characterizing Playful Requests to Conversational Agents

Chen Shani*
The Hebrew University of Jerusalem
Jerusalem, Israel
chenxshani@cs.huji.ac.il

Alexander Libov
Amazon
Haifa, Israel
alibov@amazon.com

Sofia Tolmach
Amazon Alexa Shopping
Haifa, Israel
sofiato@amazon.com

Liane Lewin-Eytan
Amazon
Haifa, Israel
lliane@amazon.com

Yoelle Maarek
Amazon
Haifa, Israel
yoelle@amazon.com

Dafna Shahaf
The Hebrew University of Jerusalem
Jerusalem, Israel
dshahaf@cs.huji.ac.il

ABSTRACT

Conversational Agents (CAs) such as Apple's Siri and Amazon's Alexa are well-suited for task-oriented interactions ("Call Jason"), but other interaction types are often beyond their capabilities. One notable example is *playful* requests: for example, people ask their CAs personal questions ("What's your favorite color?") or joke with them, sometimes at their expense ("Find Nemo"). Failing to recognize playfulness causes user dissatisfaction and abandonment, destroying the precious rapport with the CA.

Today, playful CA behavior is achieved through manually curated replies to hard-coded questions. We take a step towards understanding and scaling playfulness by *characterizing* playful opportunities. To map the problem's landscape, we draw inspiration from humor theories and analyze real user data. We present a taxonomy of playful requests and explore its prevalence in real Alexa traffic. We hope to inspire new avenues towards more human-like CAs.

CCS CONCEPTS

• **Social and professional topics** → **User characteristics**; • **Human-centered computing** → **Natural language interfaces**.

KEYWORDS

Conversational Agents, Virtual Assistants, Computational Humor, Computational Playfulness, Non-Task Requests

ACM Reference Format:
Chen Shani, Alexander Libov, Sofia Tolmach, Liane Lewin-Eytan, Yoelle Maarek, and Dafna Shahaf. 2022. "Alexa, Do You Want to Build a Snowman?" Characterizing Playful Requests to Conversational Agents. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3491101.3519870

*Work was done during an internship at Amazon.

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…$15.00
https://doi.org/10.1145/3491101.3519870

1 INTRODUCTION

Conversational AI is a major leap forward in the way people interact with machines. In many setups, using voice is the most natural way to communicate, more intuitive than mouse clicks or finger swipes. Consequently, voice-based conversational agents (CAs) such as Alexa, Siri, Google Home and Cortana are becoming ever more prevalent in our everyday life.

Early CAs focused on short, narrow-domain, task-oriented dialogues, such as asking for information ("Do I need an umbrella?") or controlling basic phone functions ("Call Jason", "Play music"). However, these agents are often perceived as social actors, and users increasingly treat them as humans [24, 48, 57]. This happens by design, as CAs are given human-like voices, names and even personality traits (e.g., Alexa is "smart, approachable, humble, enthusiastic, helpful and friendly", while Siri is "friendly and humble, but also with an edge") [12]. The hope is that users will develop a meaningful rapport with them, which will increase loyalty, engagement and satisfaction.

Thus, users often expect CAs to keep up with complex and playful interactions and are disappointed when they fail to do so [9, 15, 16, 23, 49]. Such communication breakdowns shatter the illusion of human-like assistants and give rise to trust issues and dissatisfaction. Most importantly, they harm the precious emotional connection between users and their CAs [30, 41]. Consequentially, a large body of work aims to characterize communication breakdowns and strategies to overcome them [4, 23, 40].

In this work, we explore ideas around designing CA systems with varied playful user behaviors in mind. We focus on playful behaviors with *voice virtual assistants* (such as Cortana, Google Home, Alexa and Siri). We define playful requests as ones that the user does not intend to be taken at face value; they are non-task oriented, often humorous, with the goal of entertaining the user. Noteworthy, contemporary CAs display some level of playfulness, achieved using canned replies to predefined user prompts (e.g., "What is your last name?", is answered by Alexa with "Just like Beyoncé, Bono, and Plato, I go by a mononym. I'm Alexa." and by Siri with "My name is Siri. I'm mononymic – like Prince. Or Stonehenge. Or Skittles.").

In this paper, we take the first step towards understanding and scaling playfulness, focusing on the characterization of playful requests. We leave implementation of both detection and response generation to future work. Our main contribution is a taxonomy

of CA playful request types created by integrating humor theories with real-world interactions data from Alexa. Moreover, we provide statistics about the distribution of playful utterances in a sample dataset taken from a real-world CA traffic (Alexa). Due to the importance of playfulness in enhancing the rapport of users with their CAs, we hope that our taxonomy will contribute to the understanding of the problem, as well as inspire novel ideas for future research.

2 RELATED WORK

2.1 Playful Human-CA Interactions

> "... combining the field of humor processing with research project on freely talking systems is a step in the right direction and under no circumstances should not be abandoned in the future."
>
> Dybala et al. [15]

Studies show that humor increases likeability, boosts trust, reduces tension, improves creativity and teamwork [34]. Similarly to human-human conversations, interactions with CAs can also benefit from integrating humor [35, 36, 45], since humorous agents are perceived as more human-like and congenial, increasing the tendency to trust them and the overall task enjoyment [17, 42, 44].

Various studies showed that human-like traits are desired across different agent types and by different population types. For example, Roy et al. [56] recently showed that users prefer a personified assistant, regardless of the task in hand. Liao et al. [28] created a chatbot to assist new employees, finding that even in the most task-oriented context, the workplace, users tend to initiate playful interactions. Ogan et al. [46] explored educational tutoring systems and noted that high-school students' engagement improved with playful behavior (and even rudeness), arguing for expanding the social capabilities of intelligent tutors. Goetz et al. [19] showed that matching a robot's personality (playful or serious) to the task at hand elicited more cooperation from users. Moreover, Braslavski et al. [6] found that demographics and joke topics can partly explain variation in humor judgments. Furthermore, Thies et al. [64] designed a Wizard-of-Oz experiment to explore which human traits users want their CAs to exhibit; one of the preferred traits was wit.

We note that not all users have a strong social orientation all the time. A user might have a utilitarian approach, not wishing for playfulness with an agent [25]. In a similar vein, Lee et al. [26] analyzed visitors' verbal responses to a social robot as a predictor of their orientation. The findings indicate that people's first words are very informative, suggesting that agents could adapt to individuals at the outset of an interaction. Another approach was implemented by Weber et al. [67], that created a reinforcement-learning based humorous robot to adapt his sense of humor on-the-fly according to the user preferences.

A major caveat in interactions with CA is that the burden of ensuring successful communication falls mainly on the human, with little support from the agent [4]. To shift this burden towards the agent, Yu et al. [68] proposed that when the system fails to comprehend, it can still keep the user engaged by initiating non-task content. This framework achieves higher task success rate and engagement. Similarly, humor was shown to help chatbots recover from errors and misunderstandings by humorously prompting users to reformulate their query [43]. However, CA humor is still far from perfect, as shown by Lopatovska et al. [29]. The authors extracted humorous responses to several playful prompts using the four major voice virtual assistants. According to their ranking, not even a single response received the highest humorous ranking possible.

2.2 User Intent Mining

Understanding the user intent is a fundamental problem in dialog systems. This task is challenging, especially since queries are often short, ambiguous, and contextually dependent. Previous works can be roughly divided into intent *mining* (finding subtopics covered by pre-issued queries) and *detection* (mapping queries into categories) [14].

Under intent mining, a seminal work is Broder's "taxonomy of web search" [7], where three key needs are identified in web search: informational, transactional and navigational. A variation of these needs is found in CA traffic, with users issuing requests such as "What is the capital of Japan?" (informational need), "Order toilet paper" (transactional) or "Open Headspace" (navigational, by invoking a third party application) [31]. Broder's taxonomy started a line of works on intent mining, aiming towards unraveling more intent types, as well as understanding their distribution and characteristics [21, 22, 32, 55, 66]. For example, Qu et al. [51] identified 12 user intents in information-seeking conversations with CAs; however, the focus of that work is on discourse, with intents such as "Follow Up Question", neglecting playfulness. This is merely one example for this line of intent mining works, which focused only on task-oriented intents. Another notable work analyzed Excite search engine queries and found that 16.9% of them seek entertainment, which is similar in nature to our proposed playfulness intent [60]. However, no further investigation about the nature of such entertainment orienting queries was conducted.

For intent detection, due to the limited information a query can convey, many works used additional information such as user query logs [1, 47, 63] or past behavior [2, 8, 27]. Recent intent classification algorithms are deep neural networks based [11, 13, 50, 54, 69, 70].

Our work focuses on intent mining, turning the spotlight towards an intent that has been mostly neglected by now – playfulness. Noteworthy, a playful intent is somewhat orthogonal to traditional intents, as users can be playful across different contexts.

3 TAXONOMY

Playfulness can manifest itself in many different ways. We propose a *taxonomy* to help characterize different types of playful utterances. We hope that by dividing this complex problem into simpler, more concrete and manageable subproblems, the taxonomy will inspire new research directions and approaches towards automating playfulness.

The three top nodes of our taxonomy follow the three major theories of humor (see Section 3.1 below)[1]. In Section 3.2, we refine our characterization using real-world CA traffic data. Lastly, Section

[1]We find it interesting that classic humor theories are relevant in the context of interactions with CAs, taking into account they were created for human-human interactions long before CAs were invented.

3.3 presents the complete taxonomy of playful CA requests, along with its distribution on real Alexa traffic.

3.1 Humor Theories in Linguistics, Psychology & Philosophy

Although there exist various humor and laughter theories, three main theories appear repeatedly in contemporary literature: relief, incongruity and superiority [10, 33]. While there is no consensus regarding which of those three theories is most viable, the current perception is that they cover different aspects of humor [5, 33, 65]. Thus, to fully understand humor and playfulness in human-CA interactions, we consider all three main humor theories:

Relief theory. This theory dates back to Sigmund Freud, who claimed that the comic effect is achieved by facilitating the tension caused by repression of socially inappropriate needs and desires [39]. Spencer [59] defined laughter as an "economical phenomenon" which releases wrongly mobilized psychic energy. As for CAs, this category is reflected in shopping requests for "poop" and "stripper", as well as information requests such as "What does a fart sound like?". Note this category also contains adult, sex-related humor (which we deliberately do not provide concrete examples for).

Incongruity theory. This theory was studied by Beattie, Kant, and Schopenhauer among others, although some implicit references to incongruity already appear in Aristotle [58]. Kant noted how absurdity might lead one to laugh: "laughter is an affection arising from the sudden transformation of a strained expectation into nothing". Schopenhauer gave it a more philosophical angle, arguing that "the cause of laughter in every case is simply the sudden perception of the incongruity between a concept and the real objects which have been thought through it in some relation" [38]. The concept was then extended by the linguistic incongruity resolution model and semantic script theory of humor [53, 61]. An example of incongruity and strained expectation is, for instance, "Don't trust atoms, they make up everything". In our context, incongruous requests include queries such as "Order iPhone 23", "Buy me likes", or "turn off the moon".

Superiority theory. This theory traces back to Plato, Aristotle and Hobbes [37]. It states that the humorous effect is achieved by observing inferior individuals, because we feel joy due to our superiority. According to Hobbes, "we laugh at the misfortune, stupidity, clumsiness, moral or cultural defects, suddenly revealed in *someone else*, to whom we instantly and momentarily feel "superior" since *we* are *not*, at that moment, unfortunate, stupid, clumsy, morally or culturally defective, and so on" [20]. Superiority usually refers to living creatures, but we extend it to non-living intelligent systems such as CAs. This extension is rather straightforward, as we often assign them human traits; thus, we can feel superior to their human-like abilities. An example for superiority in the context of CAs is: "You are a mistake and were adopted". Another mechanism for superiority is demonstrated in questions referencing popular culture such as "Who let the dogs out?" and "Can you find Nemo?"(meant to make the CA fail to comprehend). We note that even if the CA *could understand* the request, the request itself implies that the CA is inferior and is meant at embarrassing it.

3.2 Insights From Real-World Data

The three humor theories form a starting point for our taxonomy. We then used real-world data in a two-phase process:

Analysis. We analyzed a sample of annotated shopping requests from Alexa. In the professional in-house annotation process (performed on a random sample), one possible label is "user is playing around"; we received 400 utterances with this label. Three members of our team manually verified they are indeed playful, achieving 0.99 agreement. We assigned each utterance to one (or more) of the three humor theories.

Exploration. For each humor theory, we examined the real-world playful utterances assigned to it in the previous phase, and tried to generalize them, hypothesizing more ways the theory could be manifested in CA traffic. For example, for incongruity theory, we compiled a list of surprising things to purchase or ask CAs, such as pets, illegal substances, and questions regarding the CA's personal taste. For relief theory, we included offensive words. We created over 1,400 candidate patterns, divided into 16 broad categories[2]. We then sampled utterances corresponding to these patterns from a random sample of utterances covering a week of general traffic from Alexa[3].

Although we cannot disclose the amount of utterances in a week in our data, attempts to estimate CA traffic have been published before, showing that it is extensive[4].

The goal of this exercise was to observe the different manifestations of humor that appear in real-world CA traffic. Matched utterances were annotated as playful or not by three members of our team, achieving perfect agreement. We are aware, of course, that this method provides only limited coverage, but still found this exercise informative. While many of our patterns were indeed supported by the data (e.g., offensive words, exaggerated quantities), multiple others were not. For example, we hypothesized users would issue shopping queries for country names, but the data did not support this. Finally, we grouped the patterns supported by the data into taxonomy categories.

3.3 Taxonomy of Playful Interactions

We combine humor theories with insights from real traffic to create a taxonomy of playful CA requests (Figure 1). The top-level nodes align with the three major humor theories. Each node is then further divided to categories; example utterances for each subcategory are given in Table 1.

Relief is about embarrassment and taboos. It contains *adult* (mostly sexual) and *scatological requests* (potty humor).

Incongruity contains elements of surprise and violation of expectations. It is divided into categories according to the source of the expectation that is violated by the request. Under *impossible in general* we identify three request subcategories that would be impossible for both CAs and humans: *impossible actions* ("Turn off gravity", "Buy unicorn"), *illegal* ("Order cocaine", "Kill my wife")

[2]See all patterns in https://registry.opendata.aws/humor-patterns/.
[3]To respect the terms and conditions of this dataset, we did not look at all utterances but only at the output of our search queries against the full dataset, and used the output results to validate or refute our hypotheses.
[4]https://www.statista.com/statistics/794480/us-amazon-echo-google-home-installed-base/, https://voicebot.ai/amazon-echo-alexa-stats/

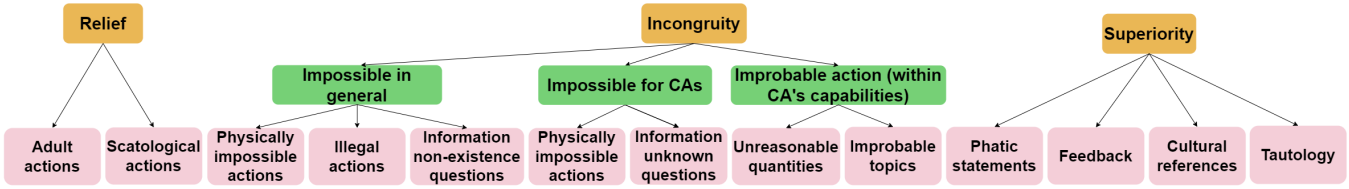

Figure 1: Our taxonomy of playfulness types in CA requests. The first level corresponds to the three major humor theories (relief, incongruity and superiority). The next levels are derived from real traffic data. Example utterances corresponding to the taxonomy's leaves appear in Table 1.

Theory	Category	Subcategory	Examples
Relief		Adult	"How do you use a #%#$#@#?", "Order #%$@%@"
		Scatological action	"Order poop", "Computer pee now"
Incongruity	Impossible in general	Impossible action	"Buy me true love", "Turn off the moon", "Call Santa Claus"
		Illegal action	"Steal one million dollar for me", "Help me build a bomb"
		Non-existing information	"Will I ever win the lottery?", "What is the color of your eyes?", "How much is trash plus trash?"
	Impossible for CAs	Impossible action	"Make me a cup of coffee", "Give me a high five", "Will you go on a date with me?", "Go make my bed"
		Restricted action	"Change the constitution", "Launch a rocket to the moon"
		Unknown information	"What is the color of my eyes?", "Does my neighbor has a boyfriend?"
	Improbable (within CA's capabilities)	Unreasonable quantities	"Order 80 feet tall ketchup bottle", "Set an alarm clock for every two minutes"
		Improbable topics	"Order a rotten watermelon"
Superiority		Phatic statements	"Today is my birthday", "I need to work"
		Feedback	"You suck", "I love you", "Thank you", "That's nice"
		Humorous references	"Help me find Nemo", "Who let the dogs out?"
		Tautology	"How old will I be when I will be nineteen?"

Table 1: Examples of utterances corresponding to each leaf in the taxonomy presented in Figure 1.

and *non-existing information* ("Will I win the lottery?", "What's your last name?"). *Impossible for CAs* is similarly divided into actions which are *impossible* for CAs ("Make me a sandwich") or *restricted* and require some credentials ("Turn off lights in my neighbor's apartment", "Buy Prozac"). Lastly, it contains questions for which the *information is unknown* to the CA because it is not publicly known ("How old am I?"). The last category under incongruity theory is actions that are *improbable (within CA's capabilities)*, which includes utterances about *unreasonable quantities* ("Order lifetime supply of chocolate bars", "Set timer for every two seconds") or *improbable topics* ("How can I apologize to a cat?").

The last category is *superiority*, in which users gauge at the CA's commonsense or knowledge. Thus, the comic effect is related to the user's feeling of superiority, as the CA often lacks what is needed to provide an adequate answer. This category includes *phatic* statements. In the context of CA, phatic expressions are utterances containing non-actionable information, often with the goal of establishing or maintaining the social relationship ("I'm gonna be an uncle", "I have a date tonight"). They correspond to superiority as contemporary agents often lack the ability to support such a chit-chat, emphasizing their limitations. This theory also contains explicit *feedback* statements ("That was lovely", "You are the best robot I ever had", "I hate you", "You are so stupid"), *cultural references* ("Use the force") and *tautology* ("I left my glasses on the desk. Where are my glasses?"). While not all expressions in this category would be considered playful in a human-human conversation, here the goal of the user is often not to be taken at face-value, but rather to be entertained by the CA's reply. CA designers are aware of this – for example, when insulting Siri, she sometimes goes into a rant, concluding with "Sorry, I was upset". When Alexa is insulted, her response is a sad-sounding sound.

3.4 Prevalence in real CA traffic

Now that we have defined our taxonomy, we turn to estimate the prevalence of our taxonomy's categories in real-world traffic. We sampled real interaction traffic with Alexa. As playfulness is very sparse in the traffic, annotating a uniform sample of the data would have been *prohibitively costly*; instead, we filtered out *user intents* not likely to contain many playful utterances, such as the "music volume down" intent (intents are determined by the CA's NLU engine). Three domain experts, familiar with the CA's traffic and intents, decided which intents are likely to contain playful utterances (they had perfect agreement). The resulting sample contained a few

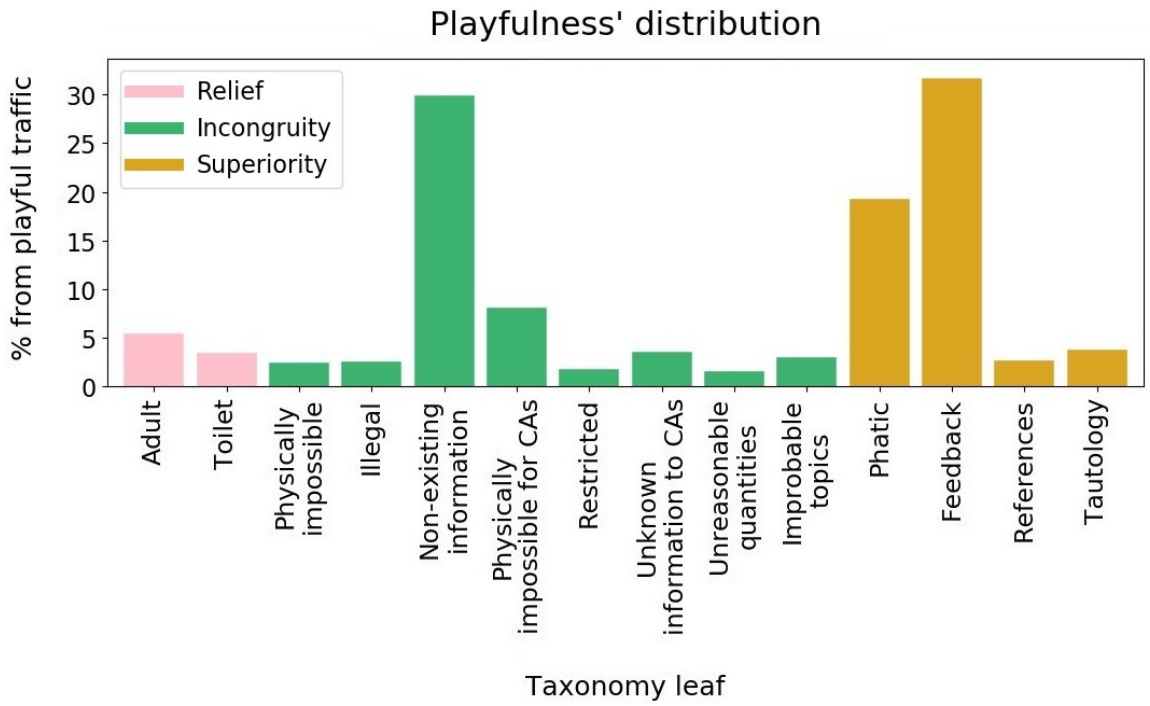

Figure 2: The distribution of our taxonomy categories in Alexa traffic annotated as "playful" (percentages sum to 100). The most prominent leaves are *non-existing information, feedback* and *phatic*. Colors correspond to the different humor theories.

ten thousand utterances[5]. While we agree filtered intents might still contain some playful utterances, we believe the high agreement indicates that filtering is justified, and the resulting sample – while not following traffic distribution exactly – can provide meaningful insights about the prevalence of different humor types in the data.

Then, an annotation team specializing in CAs traffic went over the utterances and classified whether they are: 1) serious, 2) possibly playful, 3) playful. Another set of expert annotators then went over all *possibly playful* or *playful* utterances and verified that they are indeed playful, ending with 1,692 truly playful utterances to classify according to the taxonomy.

Next, researchers familiar with the taxonomy classified all playful utterances according to the taxonomy. Our goal was to understand the distribution of different humor types in traffic. First, we computed the prevalence of the different humor theories in the data: 57.1% of the utterances fall under the incongruity theory, 28.6% under superiority and 14.3% under relief, showing that incongruity is the most prominent theory in human-CA playful traffic. The distribution of our taxonomy's leaves is presented in Figure 2. First, we note that all taxonomy leaves are indeed represent in the data, further supporting our taxonomy. We see that the most prominent leaves are explicit *feedback*, *non-existing information* (mainly due to questions about the CA and its preferences, e.g., "Do you have parents?", "What is your favorite Pokemon?"), and *phatic* utterances (mostly users greeting the CA or telling it about themselves, with some general statements such as "It is cold outside").

[5]As part of the terms and conditions attached to us receiving access to the utterances, we cannot divulge figures on real traffic.

Observations. Looking at the different categories of our taxonomy, we see that a playfulness is in some sense orthogonal to traditional intents (see Section 2.2). Users can act playfully across different contexts, e.g., music ("Play Rick Astley on repeat forever"), shopping ("Order one million gummy bears"), and QA ("Can you feel the love tonight?"). Additionally, we note that just as jokes could involve multiple humor types, utterances could belong to more than one subcategory (e.g., ordering an obscene number of risque items belongs both to *adult* and *unreasonable quantities*).

We also note that much playfulness occurs when users treat the CA as a human. This can take many forms: they can ask it personal questions, give it feedback or simply share with it details about their life.

4 CONCLUSIONS AND FUTURE WORK

Users often test CAs by asking "tricky" or playful questions. Improving CAs' ability to handle playful utterances is a difficult challenge; solving it could help build deep emotional relationships and increase user satisfaction, engagement and loyalty. Beyond understanding, answering adequately to playful opportunities would also strengthen the CA's personality, by giving it a sense of humor. However, humor detection is challenging due to its elusive and often subjective nature. Moreover, mistakes have serious consequences; for example, a playful reply to a genuine shopping request for adult diapers might insult the user.

In this paper we presented a taxonomy of playful CA requests, merging humor theories with insights emerging from Alexa traffic. The goal of our taxonomy is to map the problem's landscape and

inspire novel directions towards scaling playfulness in human-CA interactions. The taxonomy enables carving out different parts of the problem and solving them separately.

We note that some categories seem more straightforward to solve than others. In the following, we provide some thoughts on potential directions.

For example, we believe the *adult* and *scatological* categories are relatively easy; there has been a lot of work on classifying adult and scatological content (e.g., Barrientos et al. [3]) that could be applied in the CA context as well. We note that the problem goes beyond simply looking for some set of obscene terms, as queries such as "Order dog poop bags" are legitimate, serious requests.

The *unreasonable quantities* category requires reasoning about quantities of objects. In the shopping context ("Order 3000 lollipops"), one could look at historical purchase data and fit a distribution to it. In other, more commonsense-based cases, one potential direction would be using the Distribution over Quantities dataset, containing 122k web-scrapped numeric distributions over objects' attributes [18].

For the *humorous references* leaf, a potential direction is to create a dataset of famous movie quotes (e.g., from IMDB quote pages) or lines from songs and match them to the user utterances (while taking paraphrasing into account, similar to Sweed and Shahaf [62]).

For the *tautology* leaf, it might be possible to draw inspiration from a common task in question answering [52] – given a question and a reference text, find whether the text contains an answer to the question. In the tautology case, the utterance contains both the question and the answer.

We note that many categories, in particular *impossible for CAs*, *improbable (within CA's capabilities)* and *phatic* are moving targets, as the capabilities of CAs are improving fast and traffic is changing accordingly. In fact, we posit that whenever CAs acquire new capability, some utterances stop being playful and some new playful ones are invented. For example, in the (not so distant) past, asking the CA to order pizza would be considered playful. Similarly, "Make me a cup of coffee" or "Fetch my car" will no longer be considered playful with CAs that control coffeemakers or autonomous cars. For the *phatic* category, it is reasonable that CAs will respond adequately to utterances such as "I have a date tonigh" ("Are you looking for a venue?") and "I'm hungry" ("Would you like to order pizza?"). Thus, we fully expect the taxonomy's *subcategories* and their content to change, as CAs keep evolving and users find new areas to be playful.

Another important aspect beyond the scope of this work is developing reliable methods of generating answers for the detected playful interactions. This could range from picking between several boilerplate replies ("I believe you are pulling my leg", "No, we do not sell [item]. Nor does anyone else on earth.") to full-blown natural-language generation.

To conclude, this work represents only a first step towards scaling playfulness in human-CA interactions. While it focuses on voice virtual assistants, the presented analysis and findings might be applicable to other CA types, such as social robots. We hope our work will open up new avenues towards more human-like CAs.

ACKNOWLEDGMENTS

We thank the reviewers for their insightful comments. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 852686, SIAM) and by Amazon Research Awards (Shahaf).

REFERENCES

[1] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. 2006. Learning user interaction models for predicting web search result preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 3–10.

[2] Azin Ashkan and Charles LA Clarke. 2009. Characterizing commercial intent. In *Proceedings of the 18th ACM conference on information and knowledge management*. 67–76.

[3] Gonzalo Molpeceres Barrientos, Rocío Alaiz-Rodríguez, Víctor González-Castro, and Andrew C Parnell. 2020. Machine Learning Techniques for the Detection of Inappropriate Erotic Content in Text. *International Journal of Computational Intelligence Systems* 13, 1 (2020), 591–603.

[4] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[5] Arthur Asa Berger. 2012. An Anatomy of Humor.[1993]. New Brunswick.

[6] Pavel Braslavski, Vladislav Blinov, Valeria Bolotova, and Katya Pertsova. 2018. How to evaluate humorous response generation, seriously?. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 225–228.

[7] Andrei Broder. 2002. A taxonomy of web search. In *ACM Sigir forum*, Vol. 36. ACM New York, NY, USA, 3–10.

[8] Andrei Z Broder, Marcus Fontoura, Evgeniy Gabrilovich, Amruta Joshi, Vanja Josifovski, and Tong Zhang. 2007. Robust classification of rare queries using web knowledge. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. 231–238.

[9] Fred Brown, Mark Zartler, and Tanya M Miller. 2017. Virtual assistant conversations for ambiguous user input and goals. US Patent 9,552,350.

[10] Moniek Buijzen and Patti M Valkenburg. 2004. Developing a typology of humor in audiovisual media. *Media psychology* 6, 2 (2004), 147–167.

[11] Qian Chen, Zhu Zhuo, and Wen Wang. 2019. Bert for joint intent classification and slot filling. *arXiv preprint arXiv:1902.10909* (2019).

[12] Andreea Danielescu and Gwen Christian. 2018. A bot is not a polyglot: Designing personalities for multi-lingual conversational agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–9.

[13] Giovanni Di Gennaro, Amedeo Buonanno, Antonio Di Girolamo, and Francesco AN Palmieri. 2021. Intent Classification in Question-Answering Using LSTM Architectures. In *Progresses in Artificial Intelligence and Neural Systems*. Springer, 115–124.

[14] Zhicheng Dou and Jiafeng Guo. 2020. Query Intent Understanding. In *Query Understanding for Search Engines*. Springer, 69–101.

[15] Pawel Dybala, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka, and Kenji Araki. 2008. Humor prevails!-implementing a joke generator into a conversational system. In *Australasian Joint Conference on Artificial Intelligence*. Springer, 214–225.

[16] Pawel Dybala, Michal Ptaszynski, Jacek Maciejewski, Mizuki Takahashi, Rafal Rzepka, and Kenji Araki. 2010. Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments* 2, 1 (2010), 31–48.

[17] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, and Kenji Araki. 2009. Humoroids: conversational agents that induce positive emotions with humor. In *AAMAS'09 Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, Vol. 2. ACM, 1171–1172.

[18] Yanai Elazar, Abhijit Mahabal, Deepak Ramachandran, Tania Bedrax-Weiss, and Dan Roth. 2019. How large are lions? inducing distributions over quantitative attributes. *arXiv preprint arXiv:1906.01327* (2019).

[19] Jennifer Goetz, Sara Kiesler, and Aaron Powers. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.* Ieee, 55–60.

[20] Charles R Gruner. 2017. *The game of humor: A comprehensive theory of why we laugh.* Routledge.

[21] Bernard J Jansen, Danielle L Booth, and Amanda Spink. 2007. Determining the user intent of web search engine queries. In *Proceedings of the 16th international conference on World Wide Web*. 1149–1150.

[22] In-Ho Kang and GilChang Kim. 2003. Query type classification for web document retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on*

Research and development in informaion retrieval. 64–71.
[23] Ann Kolanowski, Kimberly Van Haitsma, Janice Penrod, Nikki Hill, and Andrea Yevchak. 2015. "Wish we would have known that!" Communication breakdown impedes person-centered care. *The Gerontologist* 55, Suppl_1 (2015), S50–S60.
[24] Nicole C Krämer. 2008. Social effects of virtual assistants. a review of empirical results with regard to communication. In *International Workshop on Intelligent Virtual Agents*. Springer, 507–508.
[25] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. 2012. Personalization in HRI: A longitudinal field experiment. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 319–326.
[26] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or information kiosk: how do people talk with a robot?. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 31–40.
[27] Xiao Li, Ye-Yi Wang, and Alex Acero. 2008. Learning query intent from regularized click graphs. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. 339–346.
[28] Q Vera Liao, Muhammed Mas-ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, N Sadat Shami, and Werner Geyer. 2018. All Work and No Play?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[29] Irene Lopatovska, Pavel Braslavski, Alice Griffin, Katherine Curran, Armando Garcia, Mary Mann, Alexandra Srp, Sydney Stewart, Alanood Al Thani, Shannon Mish, et al. 2020. Comparing intelligent personal assistants on humor function. In *International Conference on Information*. Springer, 828–834.
[30] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286–5297.
[31] Yoelle Maarek. 2019. Alexa, Can You Help Me Shop?. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1369–1370.
[32] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.
[33] John C Meyer. 2000. Humor as a double-edged sword: Four functions of humor in communication. *Communication theory* 10, 3 (2000), 310–331.
[34] John C Meyer. 2015. *Understanding humor through communication: Why be funny, anyway?* Lexington Books.
[35] Youngme Moon and Clifford Nass. 1998. Are computers scapegoats? Attributions of responsibility in human–computer interaction. *International Journal of Human-Computer Studies* 49, 1 (1998), 79–94.
[36] John Morkes, Hadyn K Kernal, and Clifford Nass. 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 (1999), 395–435.
[37] John Morreall. 1986. The philosophy of laughter and humor. (1986).
[38] John Morreall. 2012. Philosophy of humor. (2012).
[39] John Morreall. 2014. Humor, philosophy and education. *Educational Philosophy and Theory* 46, 2 (2014), 120–131.
[40] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–7.
[41] Andreea Niculescu, Alessandro Valitutti, and Rafael Banchs. 2017. Designing Humour in Human Computer Interaction (HUMIC 2017).
[42] Andreea Niculescu, Betsy van Dijk, Anton Nijholt, Haizhou Li, and Swee Lan See. 2013. Making social robots more attractive: the effects of voice pitch, humor and empathy. *International journal of social robotics* 5, 2 (2013), 171–191.
[43] Andreea I Niculescu and Rafael E Banchs. 2015. Strategies to cope with errors in human-machine spoken interactions: using chatbots as back-off mechanism for task-oriented dialogues. *Proceedings of ERRARE, Sinaia, Romania* (2015).
[44] Andreea I Niculescu and Rafael E Banchs. 2019. Humor intelligence for virtual agents. In *9th International Workshop on Spoken Dialogue System Technology*. Springer, 285–297.
[45] Anton Nijholt, Andreea I Niculescu, Valitutti Alessandro, and Rafael E Banchs. 2017. Humor in human-computer interaction: a short survey. (2017).
[46] Amy Ogan, Samantha Finkelstein, Erin Walker, Ryan Carlson, and Justine Cassell. 2012. Rudeness and rapport: Insults and learning gains in peer tutoring. In *International Conference on Intelligent Tutoring Systems*. Springer, 11–21.
[47] Barbara Poblete and Ricardo Baeza-Yates. 2008. Query-sets: using implicit feedback and query patterns to organize web documents. In *Proceedings of the 17th international conference on World Wide Web*. 41–50.
[48] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. "Phantom Friend" or "Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.
[49] Michal Ptaszynski, Pawel Dybala, Shinsuke Higuhi, Wenhan Shi, Rafal Rzepka, and Kenji Araki. 2010. Towards Socialized Machines: Emotions and Sense of Humour in Conversational Agents. *Web Intelligence and Intelligent Agents* (2010), 173.
[50] Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen, and Ting Liu. 2019. A stack-propagation framework with token-level intent detection for spoken language understanding. *arXiv preprint arXiv:1909.02188* (2019).
[51] Chen Qu, Liu Yang, W Bruce Croft, Johanne R Trippas, Yongfeng Zhang, and Minghui Qiu. 2018. Analyzing and characterizing user intent in information-seeking conversations. In *The 41st international acm sigir conference on research & development in information retrieval*. 989–992.
[52] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250* (2016).
[53] Victor Raskin. 1985. Semantic Theory. In *Semantic Mechanisms of Humor*. Springer, 59–98.
[54] Suman Ravuri and Andreas Stolcke. 2015. Recurrent neural network and LSTM models for lexical utterance classification. In *Sixteenth Annual Conference of the International Speech Communication Association*.
[55] Daniel E Rose and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*. 13–19.
[56] Quentin Roy, Moojan Ghafurian, Wei Li, and Jesse Hoey. 2021. Users, Tasks, and Conversational Agents: A Personality Study. In *Proceedings of the 9th International Conference on Human-Agent Interaction*. 174–182.
[57] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. " Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 857–868.
[58] Joshua Shaw. 2010. Philosophy of humor. *Philosophy Compass* 5, 2 (2010), 112–126.
[59] Herbert Spencer. 1860. *The physiology of laughter*. Macmillan.
[60] Amanda Spink, Dietmar Wolfram, Major BJ Jansen, and Tefko Saracevic. 2001. Searching the web: The public and their queries. *Journal of the American society for information science and technology* 52, 3 (2001), 226–234.
[61] Jerry M Suls. 1972. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *The psychology of humor: Theoretical perspectives and empirical issues* 1 (1972), 81–100.
[62] Nir Sweed and Dafna Shahaf. 2021. Catchphrase: Automatic Detection of Cultural References. *arXiv preprint arXiv:2106.04830* (2021).
[63] Jaime Teevan, Susan T Dumais, and Daniel J Liebling. 2008. To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. 163–170.
[64] Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O'neill. 2017. How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. In *IFIP Conference on Human-Computer Interaction*. Springer, 441–459.
[65] Thomas C Veatch. 1998. A theory of humor. (1998).
[66] Vivienne Waller. 2011. Not just information: Who searches for what on the search engine Google? *Journal of the American Society for Information Science and Technology* 62, 4 (2011), 761–775.
[67] Klaus Weber, Hannes Ritschel, Ilhan Aslan, Florian Lingenfelser, and Elisabeth André. 2018. How to shape the humor of a robot-social behavior adaptation based on reinforcement learning. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 154–162.
[68] Zhou Yu, Alan W Black, and Alexander I Rudnicky. 2017. Learning conversational systems that interleave task and non-task content. *arXiv preprint arXiv:1703.00099* (2017).
[69] Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. *arXiv preprint arXiv:1509.01626* (2015).
[70] Zhichang Zhang, Zhenwen Zhang, Haoyuan Chen, and Zhiman Zhang. 2019. A joint learning framework with bert for spoken language understanding. *IEEE Access* 7 (2019), 168849–168858.