



Examining Humanness as a Metaphor to Design Voice User Interfaces

Smit Desai*
 University of Illinois at Urbana-Champaign USA
 Northeastern University USA
 smitad2@illinois.edu

Mateusz Dubiel
 University of Luxembourg Luxembourg
 mateusz.dubiel@uni.lu

Luis A. Leiva
 University of Luxembourg Luxembourg
 name.surname@uni.lu

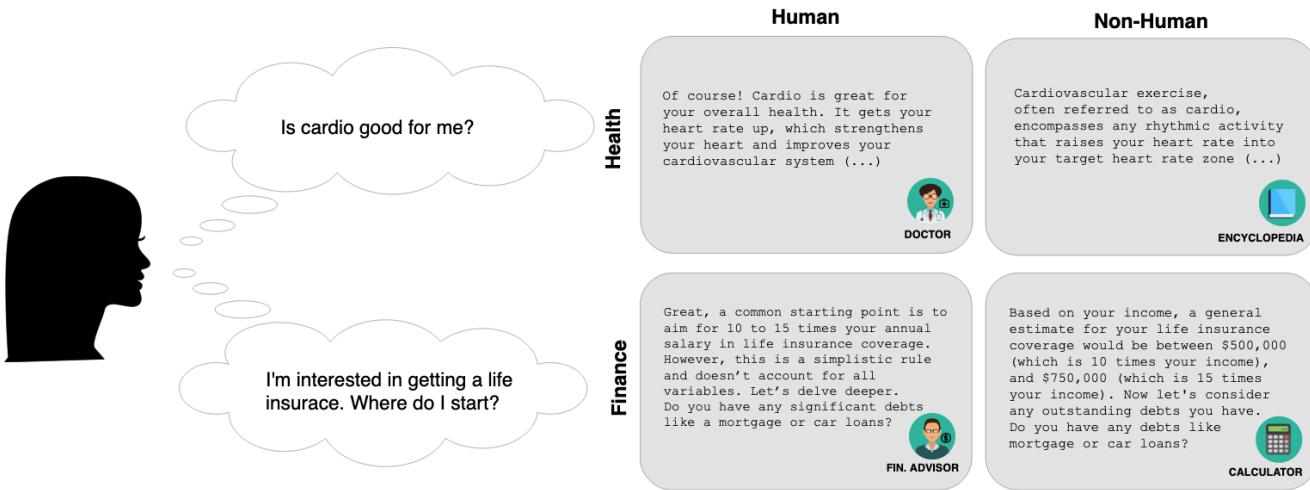


Figure 1: We explored how frequent users of Voice User Interfaces (VUIs) perceive interacting with human metaphors versus non-human metaphors in the domain of health and finance. For health, we designed two metaphorical VUIs that acted as a ‘doctor’ or a ‘health encyclopedia.’ For finance, we used the metaphors ‘financial advisor’ and ‘calculator.’

ABSTRACT

Voice User Interfaces (VUIs) increasingly leverage ‘humanness’ as a foundational design metaphor, adopting roles like ‘assistants’, ‘teachers’, and ‘secretaries’ to foster natural interactions. Yet, this approach can sometimes misalign user trust and reinforce societal stereotypes, leading to socio-technical challenges that might impede long-term engagement. This paper explores an alternative approach to navigate these challenges—incorporating non-human metaphors in VUI design. We report on a study with 240 participants examining the effects of human versus non-human metaphors on user perceptions within health and finance domains. Results indicate a preference for the human metaphor (doctor) over the non-human (health encyclopedia) in health contexts for its perceived

enjoyability and likeability. In finance, however, user perceptions do not significantly differ between human (financial advisor) and non-human (calculator) metaphors. Importantly, our research reveals that the explicit awareness of a metaphor’s use influences adoption intentions, with a marked preference for non-human metaphors when their metaphorical nature is not disclosed. These findings highlight context-specific conversation design strategies required in integrating non-human metaphors into VUI design, suggesting tradeoffs and design considerations that could enhance user engagement and adoption.

CCS CONCEPTS

• **Human-centered computing** → **Auditory feedback; Natural language interfaces; Ubiquitous and mobile computing design and evaluation methods.**

KEYWORDS

Conversational Agents, Voice User Interfaces, Metaphors, Personas, Design

*Work conducted during a research visit at the University of Luxembourg.



This work is licensed under a Creative Commons Attribution International 4.0 License.

ACM Reference Format:

Smit Desai, Mateusz Dubiel, and Luis A. Leiva. 2024. Examining Humanness as a Metaphor to Design Voice User Interfaces. In *ACM Conversational User Interfaces 2024 (CUI '24), July 08–10, 2024, Luxembourg, Luxembourg*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3640794.3665535>

1 INTRODUCTION

Voice User Interfaces (VUIs), including voice assistants like Amazon Alexa or Apple Siri, have become ubiquitous in everyday life. They are used to perform various tasks like information seeking, home automation, playing games, or even for social companionship [70]. Typically, they are designed to be and act as “assistants” [75]. However, in more specialized application domains like health and finance, they can be given the personalities of “teachers” [20], “brokers” [45], “therapists” [56], “exercise coaches” [21], or even “financial advisors” [82]. These personalities, which act as the front-end conversational user interfaces (CUIs), are referred to as “personas,” and the act of designing personas acts as the starting point of the CUI design process [67].

CUIs, from ELIZA to ChatGPT, have been anthropomorphized due to their opaque black-box nature [46]. In most cases, the anthropomorphism in users originates from “overlearned social behaviors” [58]. However, in the case of VUIs, eliciting anthropomorphic tendencies is a conscious design decision [24]. Conversation design guidelines of Google¹ and Amazon² provide detailed instructions on designing personas, including picking emotional reactions and personality traits. Due to these practices, ‘humanness’ has become the foundational metaphor of human-VUI interactions [16].

In recent literature, there has been much criticism of this foundational humanness metaphor for VUI design, highlighting issues from practical functionality to ethical considerations. From a practical perspective, presenting a VUI as an assistant results in higher pre-use expectations [39], which, if unmet, results in infrequent use [16] or non-use [14]. Additionally, VUIs utilize a request-response mechanism that could be effective for executing command-oriented tasks but falls short of facilitating genuine ‘conversations’ [61] and thus fall significantly short of the human-human communication benchmark. Moreover, finding common ground with a machine (especially in the event of conversational errors) pretending to be a human can be especially frustrating [23, 28]. However, these frustrations pale compared to the ethical issues presented by the humanness metaphor. VUIs, designed to be assistants, are given a default female voice. Thereby, reinforcing existing societal stereotypes. In other cases, they are designed to act as “servants” or “secretaries,” acting as “invisible women” in our houses [75]. These and other ethical and functional concerns have been well documented in previous CUI conferences [23, 28, 52, 63, 72].

Due to these concerns, there is a growing body of work proposing designing conversational interfaces using non-human metaphors. Until now, this research has been limited to chat-based CUIs, with the most notable example of [38]. Preliminary evidence (albeit for chatbots) suggests that non-human metaphors (e.g., animal, plant, object) could be useful in increasing user engagement and perceived workload [38] in comparison to human metaphors. Given

the problematic nature of humanness metaphors in VUI design, it is important to explore how designing non-human metaphors could impact user perceptions. Especially considering there is evidence in the literature suggesting that users already do conceptualize VUIs as “things” [14] and routinely use non-human metaphors (like Genie, Encyclopedia, Tools) to describe them.

To the best of our knowledge, there is no existing research that examines the comparison between VUIs designed with non-human metaphors and those using human metaphors in terms of user perceptions and willingness to adopt these technologies. To address this gap, we conduct a mixed-subjects study (N=240) to understand how frequent users of VUIs perceive interacting with a VUI named ‘Z’ designed using the metaphors ‘doctor’ and ‘health encyclopedia’ in the domain of health and ‘financial advisor’ and ‘calculator’ in finance. Unlike chat-based conversational interfaces, which can signal their metaphor to users through avatars or icons, VUIs lack such clear affordances. Therefore, the approach involves either explicitly informing users about the underlying metaphor or leaving it implicit. To explore this, we also investigate how the metaphor’s conveyance influences user perceptions, comparing the explicit versus implicit communication strategies. We operationalize the explicit versus implicit knowledge of the metaphor as awareness levels—with explicit awareness indicating that the participant is made aware of the metaphor used in the design process and implicit awareness signaling that the participants are not informed of the metaphor. With this paper, we explore two main research questions (RQs):

- **RQ1:** *How do users of VUIs perceive interacting with non-human metaphors compared to human metaphors in terms of perceived enjoyment, intention to adopt, trust, likeability, and intelligence in the domains of health and finance?*
- **RQ2:** *How does awareness of the metaphor (implicit versus explicit) impact these perceptions?*

With this study, we aim to provide a starting point for conversation designers and researchers to explore the use of non-human metaphors in their design and help conceptualize a new design space focusing on non-human metaphors that might not act human but are human-centered.

2 BACKGROUND AND RELATED WORK

2.1 Metaphors: A Brief Overview

Beyond literary devices used in poetry, metaphors can be interpreted as a source of sense-making [10]. Each use of metaphors includes a source domain and a target domain, and there is some relationship between them [34]. For example, in the phrase “life is a journey” life can be considered the target domain, and journey is the source domain. The concept of journey is explained using the existing understanding of life. Similarly, in the phrase “time is money,” the concept of money is understood using time. In both these cases, one concept is related to another, and the ‘relatively’ more abstract concept is explained using a more concrete concept. These categories of metaphors are called *conceptual metaphors*, and they gained prominence with the book “Metaphors to Live By” by cognitive linguists Lakoff and Johnson [44].

Understanding conceptual metaphors goes beyond merely establishing relationships between familiar concepts; it also involves

¹<https://developers.google.com/assistant/conversation-design/create-a-persona>

²<https://developer.amazon.com/en-US/blogs/alexa/alexa-skills-kit/2018/08/hear-it-from-a-skill-builder-how-to-create-a-persona-for-your-alexa-skill>

making sense of entirely new concepts by relating them to more familiar ones. Graphical User Interfaces (GUIs) exemplify this through their use of metaphors. GUIs rely on a primary ‘desktop’ metaphor, which has undergone numerous iterations to provide users with a mental model for navigating visual interfaces [59]. Additionally, the primary desktop metaphor is complemented by various secondary metaphors (e.g., folders, files, bins) and auxiliary metaphors (e.g., scrolling, icons, logging) — we recommend Alan Blackwell’s article [8] for a comprehensive overview of the history of desktop metaphor and the idea of using metaphor as a design tool. These metaphors are so ubiquitous in the language of computing that we have ceased to notice them [15], thereby becoming ‘dead’ metaphors [66].

Comparatively, the history of metaphors in CUIs is nascent. CUIs are an umbrella term for a wide variety of technologies, including chatbots, Embodied Conversational Agents (ECAs), Virtual Assistants (VAs), and Voice User Interfaces (VUIs) or voice assistants (VAs). While our primary focus in this article is on VUIs, we also provide background on metaphors in CUIs to provide a more nuanced perspective on the current literature landscape.

2.2 Metaphors in CUIs

In “Elements of Friendly Software Design” [31], Paul Heckel described two fundamental functions of metaphor within interface design. The first function involves *familiarization*, wherein metaphors are utilized to imbue the user interface with familiar concepts, enhancing comprehension and enabling users to construct mental models more easily. The second function, *transportation*, revolves around unifying these mental models and facilitating immersive user experiences. Users are transported into a cohesive experiential realm by ingeniously supplementing familiar concepts to create novel experiences. In CUIs, the pervasive use of humanness as an operational metaphor by conversation designers serves as the vehicle for transportation.

The humanness metaphor is realized through the creation of *personas* (also known as ‘system personas’³) and *social role reproduction* [41, 52]. These personas or social roles serve as the interface’s front-end or interaction layer, shaping the user experience. Crafting personas is regarded as one of CUI design’s most challenging and critical aspects [67]. This involves scripting dialogues that reflect anticipated human behavior, choosing appropriate avatars for chatbots, or selecting voices that resonate with users’ expectations. To this end, researchers and designers have deployed CUIs to fulfill various social roles, such as ‘teachers’ [20, 37], ‘coaches’ [21, 79], ‘storytellers’ [22], ‘brokers’ [45] or ‘therapists’ [56].

Designing metaphorical CUIs has an impact on how users perceive these interfaces. Khadpe et al. [39] presented metaphorical chatbots with varying levels of competence and warmth (e.g., “teenager,” “shrewd executive,” or “toddler”) using the prompt, “The bot you are about to interact with is modeled after a { Metaphor }.” They found that the choice of metaphor had an impact on users’ pre-use as well as post-use perceptions. Metaphors that signaled high competence were able to attract new users, but low competence metaphors were better evaluated and more likely to be adopted

in the long term. Similarly, in another study, Chin et al. [13] examined the impacts of formal and informal conversational styles of a VUI on the metaphorical descriptions provided by older and middle-aged adults. Their findings revealed that older adults tended to employ more professional metaphors (e.g., “librarian,” “teacher,” “lawyer”) when describing formal VUIs while opting for personal metaphors (e.g., “aunt,” “friend,” “child”) when describing informal VUIs. Researchers have also found relationships between the personalities projected by the metaphors and the personalities of the end-users. Braun et al. [9] designed VUI personalities projected by metaphors such as “friend,” “admirer,” “aunt,” and “butler” to understand user perceptions of in-car interactions. They found that users found metaphors that were similar to them (in terms of: openness, conscientiousness, extraversion, agreeableness, and neuroticism (OCEAN) [51]) were perceived to be more likable and trustworthy.

While the ideal scenario involves presenting users with a clear mental model through personas, the reality is far more complicated due to the complex nature of users’ understanding of conversational systems. Consider VUIs, also called ‘voice agents’ or ‘voice assistants’, or even Large Language Models (LLMs) like Microsoft’s Copilot.⁴ In these cases, the persona is embedded in how we identify these technologies. However, despite this, the users often form their own conceptual understandings of these systems through folk theories [25]. Often, these folk theories are disseminated in the form of metaphors, including explanations for how CUIs work (e.g., “it’s like a computer chip that has stuff on it” [40] or “genie in the bottle” [43]) or making sense of conversational errors (e.g., “like a silly child” [23]). These folk theories stem from the discrepancy between what the users expect based on the metaphors provided to them and the actual experience of using CUIs [48], also explained by the concept of Norman’s “gulf of execution” [60]. This miscalibration can result in infrequent use [16] or non-use [14].

Moreover, operationalizing humanness as a metaphor raises ethical concerns, as highlighted by the CUI community in provocation papers⁵. In CUI’21, McMillan & Jaber [52] discussed the inappropriateness of subordinate metaphors like “assistants,” “maids,” or “servants.” They argue that historically, these roles have been occupied by women of color (treated as “physical absorbers” of “physical and affective dirt of a home” [68]). Additionally, the presence of subordinate roles also implies the presence of “master” roles—a term loaded with unwanted connotations. Also, in CUI’21, Pradhan & Lazar [63] raised the ethical issues with designing human-like personas, highlighting how the default voices of commercial VUIs often resemble “white women from the West” and emphasized that crafting female personas in subordinate roles can perpetuate existing societal biases. In CUI’22, Simpson & Crone [72] vividly portrayed the potential dystopian consequences of designing CUIs with personas embodying societal roles, such as police officers, doctors, and priests. Their analysis revealed alarming outcomes, including the erosion of trust and the persistent manipulation of user behavior. Similarly, in CUI’22 Desai & Twidale [23] argued that the current implementation of personas in design is simplistic and advocated for contextually fluid personas that can change their behavior based on the task the user is performing. These

⁴<https://copilot.microsoft.com>

⁵Provocation papers explore controversial, risk-taking, or nascent ideas that have the potential to spark debate.

³<https://developers.google.com/assistant/conversation-design/create-a-persona>

provocations highlight the need for a more refined approach to designing metaphorical CUIs that fosters familiarity *and* avoids ethical dilemmas.

An alternative and innovative approach to circumvent the implementation challenges and ethical considerations associated with designing CUIs based on human metaphors is to leverage non-human metaphors. Desai & Twidale [24] developed a framework of metaphor contextualization in VUIs by identifying metaphors used by users in semi-structured interviews and VUI literature. They found that users did employ a broad range of human metaphors to describe VUIs, *but* they also used non-human metaphors (e.g., tower, “Computer” from Star Trek, genie, encyclopedia). Moreover, users fluctuated between human and non-human metaphors seamlessly without cognitive dissonance. This finding is also supported by Pradhan et al. [62], who found older adults ontologically categorized VUIs as a person as well as an object (or something in between). In another relevant study, Jung et al. [38] used the Great Chain of Being framework (GCOB)⁶ [35] to understand the effect of non-human metaphors (God, Animal, Plant, and Book) on user perceptions compared to a human metaphor, when applied to chatbots. They found that the human metaphor had no significant difference in chatbot evaluation over non-human metaphors. In the current study, we build on (1) Desai & Twidale’s [24] work on metaphor contextualization in VUIs, and (2) further explore Jung et al.’s [38] findings on designing non-human metaphors in the context of VUIs.

2.3 Application of VUIs

Health and finance are two popular application areas of VUIs [1]. In our investigation, we focus on these two areas since they can be considered as essential for individuals’ well-being and thus we believe that the presented interaction contexts will be familiar and relatable to most VUI users. Moreover, due to the sensitive nature of health and finance, trust will play an important role in the voice interactions [2]. Below, we discuss the characteristics of both of these domains.

2.3.1 Applications in Health. Healthcare is one of the most prolific sectors for VUI deployment [49]. Applications of VUIs in this area include: screening and monitoring [30, 42], consulting and advice [49, 77], learning [20], mental health support [53, 76], or healthy lifestyle advice [12, 29].

Some most commonly used personalities in this sensitive setting are coach-like, healthcare-expert, professional-like, informal, etc. [74]. While users’ preferences regarding desired VUI personality may vary [78], it has been demonstrated that the type of VUI personality can directly impact users’ attitudes and interactions [49]. For instance, conversational style choices such as using humor, providing personalized and timely feedback, or starting the conversation with a social chat have proven beneficial to enhancing the overall user experience [7]. Cox et al. [17] found that formal conversation style was perceived positively when conversation concerned managing sensitive health information or discussing patient’s lifestyle.

Subject sensitivity and privacy concerns are important considerations that can affect users’ intention to disclose information

⁶Theorized hierarchical categorization of the world using the metaphors of god, human, animal, plant, and an inorganic object.

regarding their health online [4]. However, it should be noted that same subject may have different levels of sensitivity for different different people or even for the same person depending on the circumstances [69]. Schuetzler et al. shown that even small differences in communication style such as increasing responsiveness can have significant impact on the quality of information gathered during the interview [69]. This is particularly important given that VUIs are perceived as free from personal bias and induce less anxiety in patients when disclosing risky health behaviors [50].

2.3.2 Applications in Finance. When it comes to finance, VUIs have been adopted for tasks such as providing customer support or enabling users to execute certain orders and transactions [65]. In the investing context, conversational robo-advisors are a popular use case [11, 19, 32, 54]. Usually, the goal of such advisors is to capture user’s interests and preferences and probe their risk aversion, in order to provide suggestions for suitable financial products, investment strategies, and/or portfolio allocations [65]. By providing individualized investment suggestions, robo-advisors not only facilitate the on-boarding experience for the users, but also help service provider to streamline their customer support process and improve customer retention. Reicherts et al. have developed a chatbot to probe and scaffold decision-making process of investors, maintain their investment strategy and avoid emotional reactions [65]. The proposed system was successful in promoting self-reflection and user’s awareness of their thought processes that are fundamental for strategic thinking and decision making.

Hodge, Mendoza, and Sinha [33] examined the effect of anthropomorphization of robo-advisors in the context of investor judgments by manipulating the type of advisor (human vs. robo-advisor) and the level of humanization (low vs. high). They found that investors were more likely to follow recommendations from a robo-advisor with a low-level of humanization. On the other hand, recommendations from human advisors exhibiting a high level of human characteristics were more likely to be followed. Related studies similarly indicated that anthropomorphizing VUIs can increase their perceived social presence and, in turn, affects users’ propensity to trust them [3, 64]. Recently, Morana et al. demonstrated that appropriate design of social clues design in finance robo-advisors is essential to achieve adequate level of social presence and, in turn, increase user’s trust in the agent [54].

3 STUDY

The study employed a 2x4 factorial design to investigate the effects of metaphor type (human vs. non-human) and awareness level (explicit vs. implicit) on user perceptions of VUIs in health and finance domains. Within this design, two independent variables were manipulated: Awareness Level, with two between-subjects levels (Explicit Awareness, Implicit Awareness), and Metaphor Type, with four within-subjects levels (Doctor, Health Encyclopedia in Health; Financial advisor, Calculator in Finance). This approach allows for examining the main effects and interactions between different types of metaphors and the user’s awareness regarding the metaphor’s presence. Participants experienced all four metaphor types within their assigned awareness level, facilitating a within-subjects comparison for metaphor types and a between-subjects comparison

for awareness levels. This between-groups design facilitates disentangling the individual and combined influence of metaphor type and awareness on perceived enjoyment, intention to adopt, trust, intelligence, and likeability in VUI interactions.

3.1 Designing Metaphorical VUIs

3.1.1 Metaphor Selection. We selected four metaphors from a database compiled by Desai & Twidale [24], which catalogs metaphors employed by users or designers to articulate their interactions with VUIs. Jung et al.'s investigation [38] into chat-based conversational interfaces, employing the GCOB framework, revealed that the inorganic object metaphor notably diverged in perception from the human metaphor. This prompted us to explore whether such distinctions would also be evident in the context of VUIs, leading us to juxtapose a human metaphor against a domain-specific (inorganic object) non-human metaphor. Further, drawing on findings from prior research [38, 39] that users more favorably receive metaphors signaling high competence and warmth, we opted for '**doctor**' as the human metaphor and '**health encyclopedia**' as the non-human metaphor within the health domain. Similarly, for finance, we chose '**financial advisor**' and '**calculator**' as the human and non-human metaphors, respectively.

3.1.2 Conversation Design. Our study involved conversations between users and metaphorical VUIs created by the first two authors, each with eight years of expertise in conversation design. These conversations were structured to reflect how 'Z,' in human and non-human versions, would communicate identical information across different domains. The design intricacies of these dialogues are presented in Table 1, which showcases snippets of these interactions, allowing for a direct comparison between the delivery styles of the human and non-human versions of 'Z.' Each script, encompassing human and non-human 'Z' in the health and finance domain, consisted of six turns, with the user prompts remaining consistent across both domains. The variable element was the responses from 'Z,' uniquely tailored based on the metaphorical framework employed to convey the information. This approach of designing metaphor-driven scripts aligns with established conversation design methodologies [67], wherein designers meticulously craft script-level templates that embody each chosen metaphor. To ensure the metaphors effectively shaped the user's experience, our design process was anchored in several key considerations, instrumental in differentiating the communication styles of 'Z,' including (1) non-human metaphors took a more direct and didactic approach to answering user prompts, (2) human metaphors were designed to be more informal and conversational, and (3) non-human metaphors cited sources when providing answers.

3.1.3 Synthesizing Voices. The audio clips used in our study were generated through a text-to-speech (TTS) software, Speechify.⁷ This selection was motivated by Speechify's capability to simulate conversations using multiple synthetic voices, thereby enhancing the realism of the audio clips. Initially, our design strategy involved employing a human voice actor for the user's role, with the TTS software being utilized solely for the voice of the VUI. However, during preliminary testing, we encountered challenges maintaining

consistent pitch and tone with the human voice actor across the different experimental conditions. Given that prior research, such as Dubiel et al. [26, 27], has indicated that even minor variations in voice characteristics can significantly influence user perceptions, we decided to adopt a synthetic voice for the user role as well. In selecting the synthetic voices for both the VUI and the user, we considered the prevalent gender representation in commercial VUIs, such as Siri and Alexa, which typically feature female-sounding voices [18]. To establish a clear distinction between the VUI and the user voices, we opted for a female-sounding voice for the VUI and a male-sounding voice for the user. This choice was also informed by literature indicating the potential impact of voice gender on user perception [43]. Furthermore, in an effort to mitigate any potential biases linked to the VUI's identity, we named the VUI 'Z.' This name was chosen for its neutrality, lacking associations with any particular gender or race, thus aligning with our objective to minimize confounding variables that could influence the study's outcomes. To maintain consistency across all four scripts, we used the male synthetic voice named "Guy" for the user and set the speaking tone to "chat" to make the voice sound more human-like. For 'Z,' we used the female voice "Aria" and set the speaking tone to "none". The speech, speed, and volume modifications were all set to zero percent. For the explicit awareness condition, the conversation was preceded by the prompt "You are about to hear a conversation between a user with a male-sounding voice and a voice interface named 'Z,' with a female-sounding voice. The voice interface 'Z' is modeled after a { Metaphor } in this conversation. The conversation will start now." For implicit awareness, we removed the "The voice interface 'Z' is modeled after a { Metaphor } in this conversation" part of the prompt. This prompt is adapted from [39]. All audio files are included as supplementary material.

3.2 Measures

To obtain a comprehensive insight into users' perceptions of metaphorical VUIs, we adopt the classification proposed by Wei et al. [80]. This classification categorizes user perceptions of CUIs into two relevant categories: (1) perception of interaction with agents and (2) perception of agent's characteristics. The classification is derived from a literature review encompassing measures commonly employed in CUI research to evaluate users' interactions with VUIs and their perceptions of the agents' characteristics. Guided by this framework, our study will incorporate the following measures:

- **Perception of interaction with agents** encompasses evaluative metrics to gauge the overall quality of interaction between a CUI and a user. This facet includes elements such as engagement, focusing on constructs like perceived enjoyment and intention to adopt. Within VUI literature, there is significant interest beyond mere usability, exploring how VUIs can contribute to user enjoyment [81] or enhance the quality of life [24]. To assess **perceived enjoyment**, we employ Moussawi et al.'s adapted survey [57], comprising three items on a 7-point Likert scale: "While using Z, I would find the interaction enjoyable," "While using Z, I would find this interaction interesting," and "While using Z, I would find the interaction fun." To measure users' **perceived intention to adopt**, we adapt Moussawi et al.'s survey [57], incorporating

⁷<https://speechify.com/>

Table 1: Snippets of responses provided by human and non-human versions of ‘Z’ in health and finance domains.

	Human	Non-Human
Health	<p>User: Hey ‘Z’, I have heard that cardio exercise is good for you, but I’m not sure why. Can you explain the benefits.</p> <p>Doctor: Of course! Cardio is great for your overall health. It gets your heart rate up, which strengthens your heart and improves your cardiovascular system.</p> <p>User: Okay, and how often should I do cardio?</p> <p>Doctor: Aim for at least 150 minutes of moderate-intensity or 75 minutes of high-intensity cardio per week. But it is not all or nothing—every bit helps. Even if you can’t meet these guidelines, some exercise is better than none. And remember, it’s important to find activities you enjoy; it makes it much easier to stick with it.</p>	<p>User: Hey ‘Z’, I have heard that cardio exercise is good for you, but I’m not sure why. Can you explain the benefits.</p> <p>Health Encyclopedia: Cardiovascular exercise, often referred to as cardio, encompasses any rhythmic activity that raises your heart rate into your target heart rate zone.</p> <p>User: Okay, and how often should I do cardio?</p> <p>Health Encyclopedia: The American Heart Association recommends at least 150 minutes of moderate-intensity aerobic activity or 75 minutes of vigorous aerobic activity per week, ideally spread throughout the week. Even short bouts of activity, like 10 minutes, can be beneficial if performed regularly.</p>
Finance	<p>User: Hi ‘Z’, I’m trying to figure out how much life insurance I might need. Can you help me with that?</p> <p>Financial Advisor: Great, a common starting point is to aim for 10 to 15 times your annual salary in life insurance coverage. However, this is a simplistic rule and doesn’t account for all variables. Let’s delve deeper. Do you have any significant debts like a mortgage or car loans?</p> <p>User: That’s a bit more than I expected. Is there a way to reduce the cost?</p> <p>Financial Advisor: To manage costs, you might consider term life insurance. It’s typically less expensive than whole life insurance and can be structured to cover your most financially vulnerable years, like until your mortgage is paid off or your children are through college. Remember, the cheapest policy isn’t always the best. It’s about finding the right balance between coverage and affordability.</p>	<p>User: Hi ‘Z’, I’m trying to figure out how much life insurance I might need. Can you help me with that?</p> <p>Calculator: Based on your income, a general estimate for your life insurance coverage would be between \$500,000 (which is 10 times your income) and \$750,000 (which is 15 times your income). Now, let’s consider any outstanding debts you have. Do you have any debts like a mortgage or car loans?</p> <p>User: That’s a bit more than I expected. Is there a way to reduce the cost?</p> <p>Calculator: One way to manage the cost is to adjust the term or type of insurance. For example, term life insurance is generally less expensive than whole life insurance and provides coverage for a set period. You can also adjust the coverage amount based on your most critical financial needs to find a balance between sufficient coverage and affordability.</p>

two items on 7-point Likert scale: “If available, I intend to start using Z within the next month,” and “If available, in the next months, I plan to experiment or regularly use Z.” Both these measures were developed and adapted to be used for VUIs.

- **Perception of agent’s characteristics** comprises metrics designed to elucidate how users perceive a CUI’s capabilities and personality traits, including competence, likeability, and trustworthiness. To examine variations in perceptions of the VUI’s characteristics across different conditions, we

utilized **perceived intelligence** and **likeability** measures from the Godspeed Measures using 5-point semantic differential scales [5, 6], which are widely employed in VUI research [71, 80]. Additionally, **perceived trust** was assessed using a 7-point Likert scale adapted from Jian et al. [36].

Further, we allowed participants the option to provide qualitative feedback regarding their interactions for each version of ‘Z’ should they choose to do so.

3.3 Procedure

Participants for this study were recruited via Prolific⁸, an online research platform connecting researchers with a large participant pool. They were compensated \$3.0 each for approximately 20 minutes of their time, reflecting the federal minimum wage in the U.S. The process began with participants accessing the study through a link provided by Prolific, where they first encountered and consented to an online consent form—as mandated by policies of the Institutional Review Board (IRB). Subsequently, they completed a demographic questionnaire to gather basic information like age, gender, educational background, and prior experience and familiarity with VUIs.

Following demographic data collection, participants were randomly divided into two groups based on their level of awareness—Explicit Awareness or Implicit Awareness—in a between-subjects design framework. They then partook in a 2x4 factorial design experiment, interacting with four VUI metaphor types across two domains, Health and Finance, in a within-subjects approach. In the Health domain, interactions involved a ‘Doctor’ (Human metaphor) and a ‘Health Encyclopedia’ (Non-Human metaphor), whereas, in the Finance domain, they engaged with a ‘Financial Advisor’ (Human metaphor) and a ‘Calculator’ (Non-Human metaphor). Audio clips simulating these interactions, generated via the text-to-speech service Speechify, provided a controlled and uniform experience for all participants. The sequence of metaphor presentation was counterbalanced to mitigate order effects. After each audio clip, participants completed questionnaires designed to evaluate their perceptions of the VUIs, capturing metrics such as perceived enjoyment, intention to adopt, likeability, intelligence, and trust. Additionally, they were given the opportunity to express their subjective views on the audio clips through an optional text box for open-ended feedback. To ensure attentive participation, an attention-check audio clip was also introduced; following this clip, participants were required to select specific responses in the subsequent questionnaires, helping to validate the integrity of their engagement throughout the study.

Upon completion of interactions with all four metaphor types across the two domains, participants received a Prolific-generated completion code. This code served to verify their participation on the Prolific platform and to ensure they received their compensation, effectively concluding the study procedures.

3.4 Participants

Our study design accounted for statistical power in the planning stages to ensure the reliability of the results. Guided by a power analysis, we determined that a total sample size of at least 232 participants would be necessary to observe a difference of one standard deviation between conditions with our 2x4 factorial design, assuming an alpha level of 0.05 (to minimize Type I errors) and a power level of 0.80 (to minimize Type II errors). Participants were stratified across our experimental conditions to maintain balance. Moreover, since our study required participants to imagine themselves using ‘Z’, we required them to have prior experiences with voice interfaces. Additionally, the use and understanding of metaphors are

based on various cultural factors [55], so we limited our study to highly proficient English speakers based in the U.S.

We recruited 240 participants ($M=43.21$, $SD=13.73$) via Prolific with a minimum task approval rate of 98% and at least one year of activity on the platform. Among the recruited participants, 49.2% identified as female, 49.6% as male, 0.8% as non-binary, and 0.4% preferred not to disclose their gender. Additionally, 78.3% of participants held a Bachelor’s degree or higher. All participants resided in the U.S. and either spoke English as their native language or used it as their primary language of communication. Furthermore, 82.5% of participants reported being very or extremely familiar with VUIs. Every participant used VUIs such as Alexa or Siri on a weekly basis, with 58.3% using them daily. Participants received compensation at an hourly rate of \$8.33/hr, aligning with federal minimum wage regulations in the U.S.

4 RESULTS

In assessing the impact of metaphor type (human vs. non-human) across health, finance, and overall, a repeated measures ANOVA was utilized. Metaphor type served as the within-subjects variable, while participant awareness (implicit vs. explicit) was the between-subjects factor. This design aimed to isolate the effects of metaphor type and awareness on perceived enjoyment, intention to adopt, trust, intelligence, and likeability.

4.1 Perceived Enjoyment

For health, the analysis revealed a significant main effect of metaphor type on perceived enjoyment ($F(1, 238) = 4.663, p = .032, \eta_p^2 = 0.019$). This indicates that the type of metaphor had a significant impact on participants’ enjoyment, with the human metaphor leading to higher perceived enjoyment than the non-human metaphor. No significant interaction effect between metaphor type and awareness level on perceived enjoyment was found ($F(1, 238) = 0.376, p = .540, \eta_p^2 = 0.002$), suggesting that the impact of metaphor type on perceived enjoyment was not significantly different across levels of awareness. The interaction between metaphor type and awareness levels did not significantly affect perceived enjoyment in the finance domain ($F(1, 238) = 0.045, p = .832, \eta_p^2 = 0.000$). No significant interaction effect of the awareness level on metaphor type was found as well ($F(1, 238) = 0.023, p = .879, \eta_p^2 = 0.000$). Overall, over both domains (health and finance), we found that there was a significant main effect of metaphor type on perceived enjoyment ($F(1, 238) = 9.798, p = .002, \eta_p^2 = 0.040$). However, again, there was no interaction effect of awareness on perceived enjoyment ($F(1, 238) = 1.270, p = .261, \eta_p^2 = 0.005$). The means and standard deviations of all metaphors across each domain are specified in Table 2. Figure 2 shows the distribution of perceived enjoyment scores.

To summarize, we found that participants perceived interacting with the human metaphor to be more enjoyable than the non-human metaphor for the domains of health and overall. No significant differences were found in the perceived enjoyment of human and non-human metaphors in finance.

⁸<https://www.prolific.com/>

Perceived Enjoyment

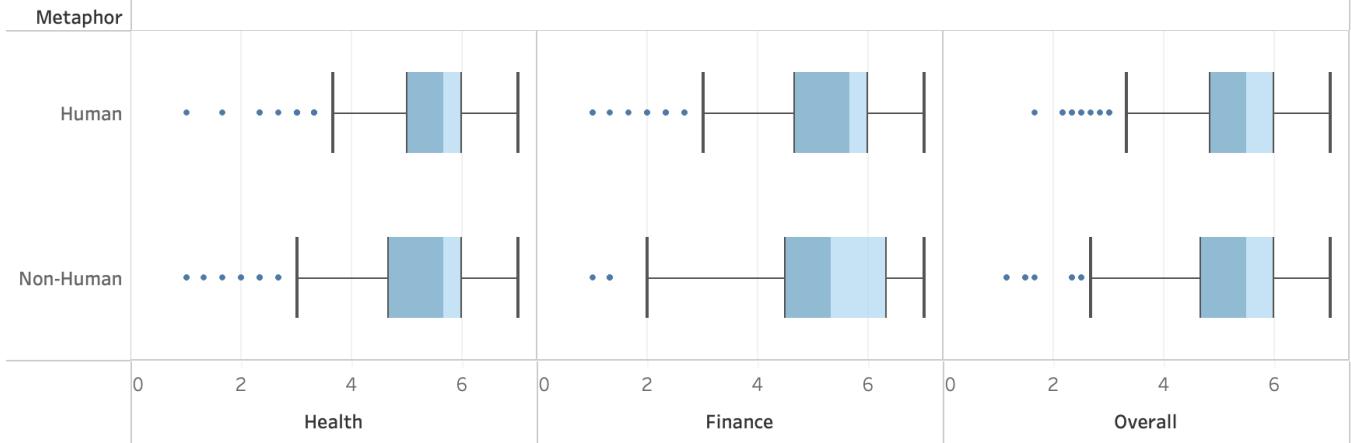


Figure 2: Boxplots indicating the distribution of perceived enjoyment scores for human and non-human metaphors across health, finance, and overall.

Table 2: The perceived enjoyment scores ($\mu \pm \sigma$: mean and standard deviation) of human and non-human metaphors for the domains of health, finance, and overall.

Perceived Enjoyment			
	Health	Finance	Overall
Human	5.43 ± 1.12	5.27 ± 1.29	5.35 ± 1.13
Non-human	5.32 ± 1.24	5.26 ± 1.30	5.27 ± 1.25

4.2 Perceived Intention to Adopt

In the health domain, we found no significant main effect of metaphor type on perceived intention to adopt ($F(1, 238) = 1.322, p = .251, \eta_p^2 = 0.006$). Further, no significant interaction effect between metaphor type and awareness level on intention to adopt was found ($F(1, 238) = 0.266, p = .607, \eta_p^2 = 0.001$). Similarly, the interaction between the human and the non-human metaphor and awareness did not significantly affect perceived intention to adopt in the finance domain ($F(1, 238) = 1.266, p = .262, \eta_p^2 = .005$). We also did not find significant interaction effect of the awareness level on metaphor type ($F(1, 238) = 0.749, p = .388, \eta_p^2 = 0.003$). Over both domains, we found no significant main effect of metaphor type on perceived intention to adopt ($F(1, 238) = 1.193, p = .276, \eta_p^2 = 0.005$). Intriguingly, while overall metaphor type did not lead to differences in perceived intention to adopt, the significant interaction effect of awareness level suggests that within the conditions where non-human metaphors were utilized, the awareness level may play a more pronounced role in influencing intention to adopt ($F(1, 238) = 4.116, p = .044, \eta_p^2 = 0.017$). Table 3 presents means and standard deviations, for each category of metaphors within their specific domains. Figure 3 indicates the distribution of the perceived intention to adopt scores.

In sum, our findings suggest that human and non-human metaphors did not differ significantly in their impact on perceived

intention to adopt across health and finance domains. However, interaction involving awareness level and non-human metaphors emerged as a significant factor in the overall intention to adopt.

Table 3: The perceived intention to adopt scores ($\mu \pm \sigma$: mean and standard deviation) of human and non-human metaphors for the domains of health, finance, and overall.

Perceived Intention to Adopt			
	Health	Finance	Overall
Human	4.92 ± 1.44	4.91 ± 1.41	4.91 ± 1.37
Non-human	4.86 ± 1.41	4.86 ± 1.45	4.88 ± 1.38

4.3 Perceived Trust

For health, our analysis found no significant main effect of metaphor type on perceived trust ($F(1, 238) = 0.796, p = .373, \eta_p^2 = 0.003$). Additionally, we did not find a significant interaction effect between metaphor type and awareness level on perceived trust ($F(1, 238) = 0.017, p = .896, \eta_p^2 = 0.001$). There was no main effect between metaphor type and awareness of trust in the finance domain ($F(1, 238) = 0.195, p = .659, \eta_p^2 = 0.001$). Further, there was no significant interaction effect of the awareness level on metaphor type ($F(1, 238) = 1.349, p = .247, \eta_p^2 = 0.006$). Interestingly, over the two domains, we found a significant main effect of metaphor type on trust ($F(1, 238) = 5.577, p = .019, \eta_p^2 = 0.023$), with participants trusting the human metaphors more than non-human metaphors. However, there was also no interaction effect of awareness level ($F(1, 238) = 0.242, p = .624, \eta_p^2 = 0.001$) on trust scores. Table 4 shows means and standard deviations for human and non-human metaphors over the domains. The distribution of trust scores can be seen in Figure 4.

Perceived Intention to Adopt

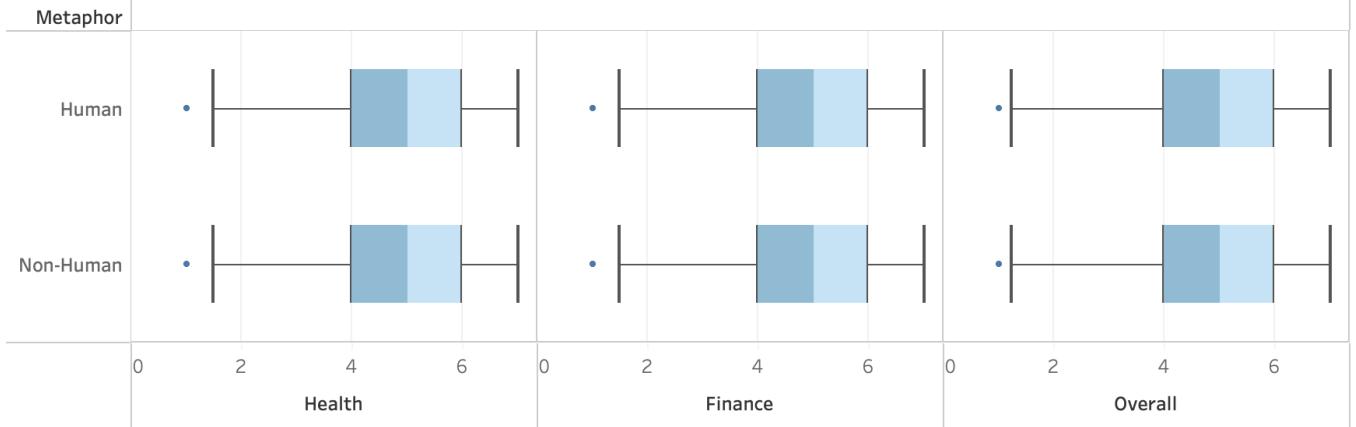


Figure 3: Boxplots indicating the distribution of perceived intention to adopt scores for human and non-human metaphors across health, finance, and overall.

Perceived Trust

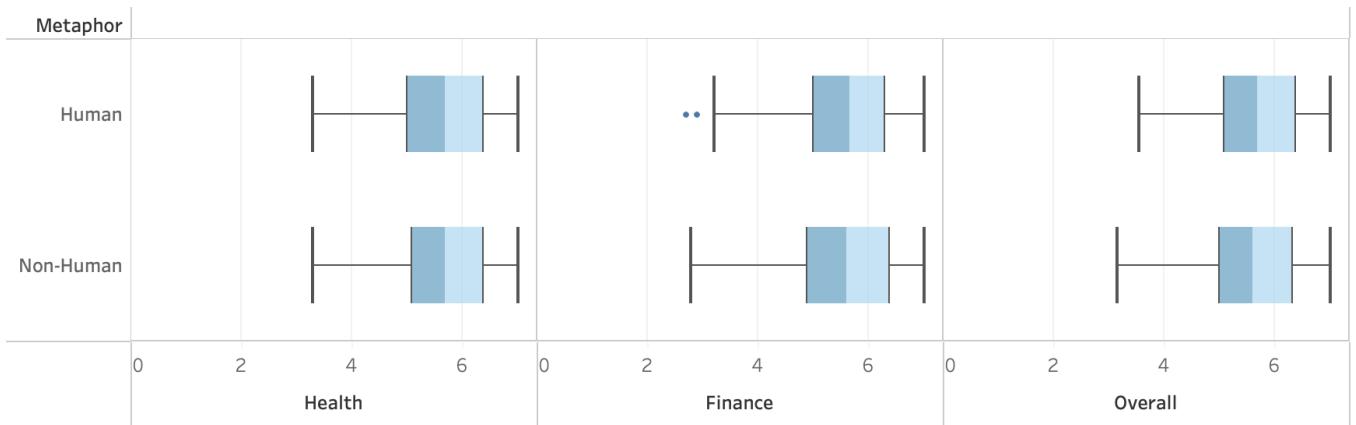


Figure 4: Boxplots indicating the distribution of trust scores for human and non-human metaphors across health, finance, and overall.

Table 4: The perceived trust scores ($\mu \pm \sigma$: mean and standard deviation) of human and non-human metaphors for the domains of health, finance, and overall.

Perceived Trust			
	Health	Finance	Overall
Human	5.66 ± 0.90	5.59 ± 0.95	5.62 ± 0.88
Non-human	5.63 ± 0.90	5.57 ± 0.95	5.58 ± 0.91

In summary, our results show that for health and finance, there were no significant differences in how participants perceived trusting human and non-human metaphors. However, overall, participants trusted human metaphors more than non-human metaphors.

4.4 Perceived Likeability

We found a significant main effect of human and non-human metaphor on perceived likeability ($F(1, 238) = 7.551, p = .006, \eta_p^2 = 0.031$), with participants considering human metaphor more likeable than the non-human metaphor. There was no significant interaction effect between metaphor type and awareness level on perceived likeability ($F(1, 238) = 0.043, p = .836, \eta_p^2 = 0.000$), indicating no impact of metaphor type on likeability across explicit and implicit awareness levels. The interaction between metaphor type and awareness levels did not significantly affect perceived likeability finance ($F(1, 238) = 0.971, p = .325, \eta_p^2 = 0.004$). There was also no significant interaction effect of the awareness of awareness level on metaphor type ($F(1, 238) = 1.333, p = .249, \eta_p^2 = 0.006$). Over the two domains, we found a significant main effect of metaphor

type on likeability ($F(1, 238) = 8.913, p = .003, \eta_p^2 = 0.036$). However, there was no interaction effect of awareness on likeability ($F(1, 238) = 0.160, p = .690, \eta_p^2 = 0.001$). The means and standard deviations of all metaphors across each domain are specified in Table 5 and Figure 5 show the distribution of perceived likeability scores.

To summarize, in the health domain, we found that participants perceived the human metaphor to be more likeable. There was no significant difference of metaphor type on likeability in finance. Overall, human metaphors were perceived to be more likeable. Awareness of metaphors did not influence likeability across all domains.

Table 5: The perceived likeability scores ($\mu \pm \sigma$: mean and standard deviation) of human and non-human metaphors for the domains of health, finance, and overall.

Perceived Likeability			
	Health	Finance	Overall
Human	4.32 ± 0.61	4.25 ± 0.69	4.28 ± 0.60
Non-human	4.22 ± 0.69	4.21 ± 0.69	4.23 ± 0.63

4.5 Perceived Intelligence

In the context of health, there was no significant main effect of human and non-human metaphor on perceived intelligence ($F(1, 238) = 0.053, p = .818, \eta_p^2 = 0.000$). There was also no significant interaction effect between metaphor type and awareness level on intelligence ($F(1, 238) = 0.030, p = .863, \eta_p^2 = 0.000$). Similarly, there was no impact of metaphor type on perceived intelligence in the finance domain ($F(1, 238) = 0.971, p = .325, \eta_p^2 = 0.004$). We also did not find a significant interaction effect of the awareness level on human and non-human metaphor ($F(1, 238) = 1.333, p = .249, \eta_p^2 = 0.006$). Over health and finance, we found no significant main effect of metaphor type on perceived intelligence ($F(1, 238) = 0.447, p = .504, \eta_p^2 = 0.002$) and no interaction effect of awareness level as well ($F(1, 238) = 0.376, p = .541, \eta_p^2 = 0.002$). Table 6 presents means and standard deviations for each category of metaphors within their specific domains. Figure 6 indicates the distribution of the perceived intention to adopt scores.

Table 6: The perceived intelligence ($\mu \pm \sigma$: mean and standard deviation) of human and non-human metaphors for the domains of health, finance, and overall.

Perceived Intelligence			
	Health	Finance	Overall
Human	4.42 ± 0.57	4.43 ± 0.60	4.43 ± 0.54
Non-human	4.41 ± 0.58	4.40 ± 0.59	4.42 ± 0.55

In sum, our findings suggest that the human and non-human metaphors did not significantly impact perceived intelligence across all domains. There was also no significant effect of awareness level on perceptions of intelligence.

4.6 Qualitative Perceptions

In addition to the questionnaires, we also gave the participants the option to provide their qualitative opinions of each version of 'Z.' On average, 54.16% of the participants provided some qualitative feedback on each metaphor across both awareness levels. The first author analyzed the qualitative data using an inductive thematic analysis approach. This involved generating initial codes from the data, which were then grouped into emerging themes to capture recurring patterns in participant responses.

For health, the participants considered the human metaphor (doctor) to be "helpful," "pleasant" and "interactive" (mentioned by five participants each). On the other hand, the participants considered the non-human metaphor (encyclopedia) to be more "informative," "robotic," and "knowledgeable" (mentioned by eight, seven, and five participants, respectively). We found that participants had diverging opinions on how the human and non-human metaphors handled citing sources. Some participants found the human metaphor to be "*easier to listen to*" (P235) because it did not mention sources, while other participants considered the non-human metaphor "*likelier to listen to*" (P238) because it cited sources. One participant in implicit awareness level rightly identified the non-human metaphor as "*walking encyclopedia*" (P221). However, for finance, participants perceived the human (financial advisor) and non-human metaphor (calculator) to be "friendly," "helpful," and "knowledgeable" (mentioned by five participants each).

Interestingly, although we used the same voices (with the same pitch, tone, and volume) for all metaphors, some participants perceived the voices to be different based on the metaphor they were interacting with. For example, P68 described the voice of the calculator metaphor as "*this voice sounds more clinical and less friendly*." While P184 described the financial advisor metaphor as "*I felt genuine care from the voice of this version of 'Z.'*" The same finding persisted with the metaphors in the health domain as well. P30, in the feedback of the doctor metaphor, said, "*voice was more pleasant and human-like than the last one (encyclopedia)*." Conversely, P66 and P103 described the encyclopedia metaphor voice as "*more robotic*" and "*less personable*" respectively. This finding highlights how the content and presentation of information can have an impact on how people perceive the voices of the VUIs.

5 DISCUSSION

In this study, we found that the metaphorical nature of VUIs impacts user perceptions—however, there are caveats. Most importantly, the context of the interaction matters significantly. In the context of health, we found that the human metaphor was perceived to be likable and enjoyable than the non-human metaphor. Meanwhile, for finance, the metaphorical nature of the VUI did not influence perception. Moreover, regardless of the context, users found both human and non-human metaphors trustworthy and intelligent, and it did not influence their intentions to adopt the VUI. Surprisingly, we did not find much significant effect of awareness of the metaphor on user perceptions. However, participants perceived intending to adopt non-human metaphors slightly more if they were unaware of their non-human nature.

Although there is not much literature about designing non-human metaphors on user perceptions of VUIs, we do have some

Perceived Likeability

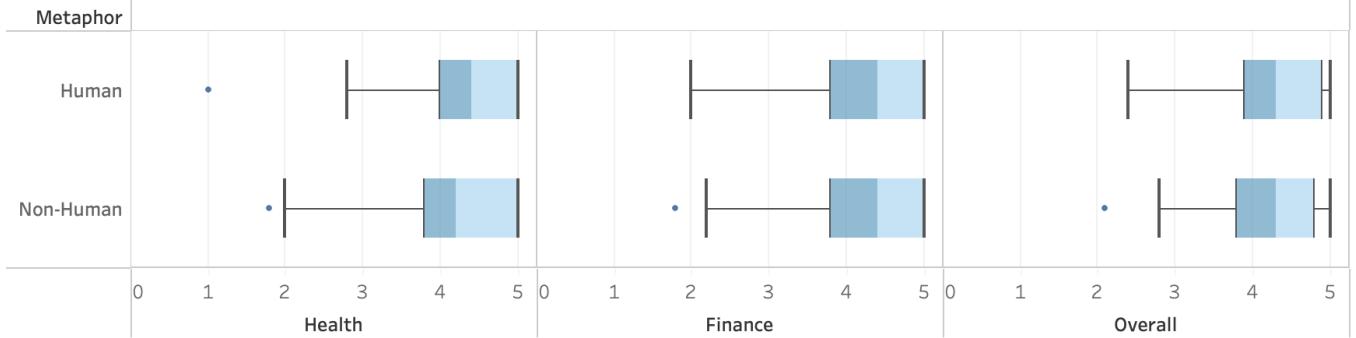


Figure 5: Boxplots indicating the distribution of likeability scores for human and non-human metaphors across health, finance, and overall.

Perceived Intelligence

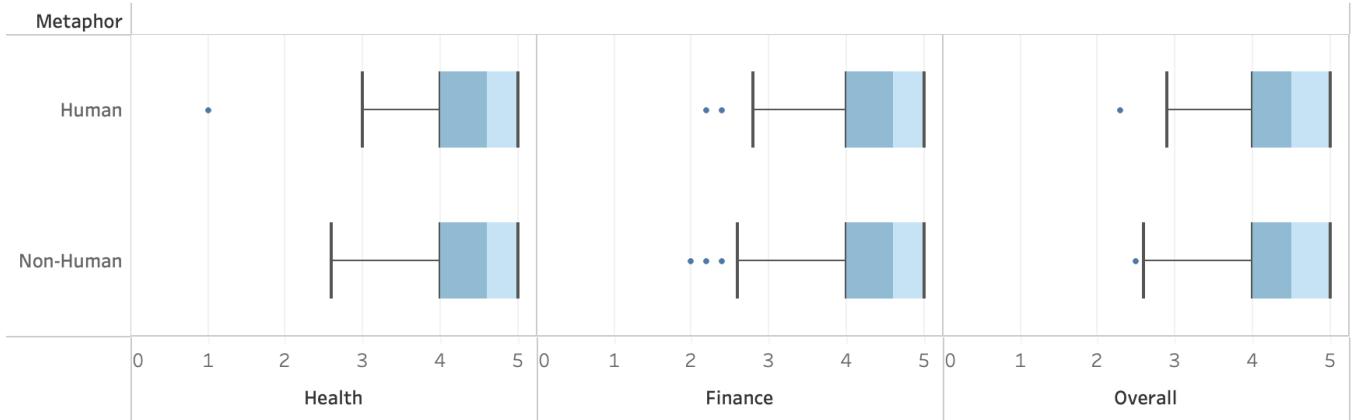


Figure 6: Boxplots indicating the distribution of perceived intelligence scores for human and non-human metaphors across health, finance, and overall.

related studies to compare our findings within the space of chatbot design. Most prominently, Jung et al. [38] found that a ‘book’ metaphor resulted in lower enjoyment and engagement of tasks compared to a ‘human’ metaphor; however, the book metaphor also resulted in reduced cognitive load. Similarly, in our study, in the context of health, we found that the health encyclopedia metaphor was perceived to be less likable and enjoyable than the doctor metaphor. Some qualitative studies with VUIs could explain this finding in the context of health information seeking. Older adults, in co-design sessions, imagined their ideal VUI for imparting health information to be “compassionate” [21, 22]. Since receiving information about health could be sensitive, users prefer receiving this information in a more ‘humane’ way. We designed the health encyclopedia VUI to be direct and informative, but we found that users did prefer dialogue elements that signaled humanness. For example, P219, in the implicit awareness condition, mentioned for the doctor metaphor, “*I am here for you*” is an oddly wholesome thing to hear at the end

of the conversation and feels suitable (to this context).” Moreover, a doctor-patient relationship is more ‘personal’ in nature [73] and people prefer getting medical advice from humans than AI—even if AI outperforms doctors on various diagnosis metrics [47]. We see a carry-over of this preference in our study as well, wherein the measures related to likeability and enjoyment were more affected than trust, intelligence, or even perceived intention to adopt.

Conversely, for finance, we did not find differences in how participants perceived human and non-human metaphors. This finding aligns with research in AI-based recommendation literature for finance-based decision-making. Users typically perceive advice from AI favorably in these scenarios. A recent survey found that about 31% of users would accept recommendations from conversational AI, like ChatGPT, without verifying them.⁹ Additionally, positive attitude towards AI correlates with intention to accept

⁹<https://www.cnbc.com/2023/08/24/31percent-of-investors-are-ok-with-using-ai-as-their-financial-advisor.html>

advice from a CUI. In our study, we found some participants had strong opinions on getting financial advice from the VUI. P140, in explicit condition, said about the calculator metaphor, “*I definitely would not be telling my voice interface information regarding my debts, income, or more. I feel providing that information to something not within your direct control is unwise. Additionally, the voice interface asking further questions regarding your financial situation is also unnerving.*” They reiterated their opinion on the financial advisor metaphor as well. However, overall, the metaphorical nature of VUIs did not influence perceptions or perceived adoption.

Desai & Twidale [24], in their framework for metaphor contextualization, found a great imbalance between deployments of VUIs in human and non-human roles. Although users perceived VUIs to be human, non-human, or something in between [14, 62], manufacturers, designers, and researchers continuously conceptualize, design, and present VUIs in social roles. This gulf between how the users perceive VUIs and what is presented to them is not rooted in scientific reasoning and is more aspirational than practical [28]. In their paper, aptly titled “Once a Kind Friend is Now a Thing,” Cho et al. [14] discuss how user expectations for VUIs change longitudinally. Commercial VUIs are advertised as “assistants,” so naturally, users expect human-like efficiency from them. However, with long-term use, these expectations decrease, and eventually, users adapt their mental models for VUIs—relegating them from being friends to things. However, in this paper, we challenge conventional wisdom in conversation design by asking—what if we design VUIs as “things”? Our findings highlight that human metaphors, although preferred in some contexts, are not a panacea. More nuance is needed when designing VUIs, and practitioners should consider if a non-human metaphor might be better for their use case.

5.1 Design Implications

Our findings have several implications for design. We will distill these implications into three main design guidelines (DG) to help conversation designers identify the right metaphor for their design.

5.1.1 DG1: One Size ‘Does Not’ Fit All. Currently, VUIs are designed to occupy *one* metaphor or persona. For example, Siri acts like an assistant, regardless of the context of the conversation. The tone or delivery style of commercial VUIs does not change based on user request. Responses to queries like “Do I have Alzheimer’s?” will be answered in the same manner as “Should I invest in crypto?” We consider this to be absurd and problematic. Results from this study show that a human metaphor is more likeable and enjoyable for sensitive contexts like health than finance. So, conversation designers should consider the sensitivity of the context in which their VUI will be deployed in designing their personas. Although the inclination towards designing human personas is understandable (given how prevalent they are in design), the decision should be rooted in the context of eventual use. Moreover, research shows that CUIs designed using human metaphors set high expectations as users perceive them to be more competent, but with subsequent use, these highly competent metaphors are rejected for metaphors that project less competence [39]. Similarly, non-human metaphors might not project high competence [37] or seem more familiar; they will set more realistic expectations and are more aligned with users’ mental models for VUIs [23].

5.1.2 DG2: Don’t Show, Don’t Tell. Studies involving metaphorical CUIs have traditionally involved chatbots [37, 39]. With chatbots, there are several ways in which one can signal the metaphorical nature of the interface, including the avatar (or an associated icon) of the CUI. However, with VUIs, there is no graphical medium to denote the metaphor without explicitly telling the user that they are interacting with a VUI modeled after a specific metaphor. Nonetheless, we found that the awareness of the metaphorical nature of the VUI did not have as much impact on user perceptions as we anticipated. Further, even in our qualitative data, we did not find much mention of how the metaphor influenced participants’ opinions of the VUIs. We did observe some detrimental effects of awareness on the intention to adopt non-human metaphors. Therefore, we advise conversation designers to employ system personas as a strategic element in VUI design while discreetly managing these choices from end-users, particularly when integrating non-human metaphors into their designs.

5.1.3 DG3: Be like a Chameleon. In GUIs, a primary desktop metaphor is decomposed into secondary and auxiliary metaphors (e.g., files, folders, menus, bins, etc.). However, for VUIs, there is merely a primary ‘humanness’ metaphor. Of course, a humanness metaphor is also somewhat nuanced because it gives rise to questions like ‘What kind of human?’, ‘doing what?’ Yet it is difficult to base technologies on humans because humans are *very* complicated. People adapt their behavior based on the context they are in or who they are interacting with. Currently, in our designs, we do not account for these complexities or contextual awareness. However, our results advocate for these complexities to be taken into account for design. We conceptualize a design space where the metaphor presented to the user changes based on the context of the conversation. If a user asks for health information, a VUI could use a human doctor metaphor, and if a user asks for financial advice, it can act like a calculator. The key idea is to adapt the metaphor to the conversational context. In this study, we used two human metaphors and two non-human metaphors and compared our results between them. However, the use of these specific metaphors is not as interesting as what the design implications are. We urge researchers to experiment with other metaphors in various contexts or even mix metaphors altogether. A playful approach to design is needed, and the rigid structures of system persona-based design cannot achieve it.

5.2 Limitations and Future work

In the current study, participants formed their impressions of VUIs by examining hypothetical dialogues between a user and a metaphorical VUI. While this method is common practice in VUI research [71], it notably lacks the inclusion of direct interaction with the technology, offering insights that are observational rather than experiential. Further, perceptions of VUIs, along with the users’ relationship with them, change over time [14]. Therefore, discerning how the outcomes of this study might translate to or differ in longitudinal research settings is important. Moreover, we restricted our study to native English speakers living in the U.S. This was done because the use of metaphors depends on cultural factors [55], which could introduce additional confounding variables to this study. Going forward, we encourage researchers

to study how cultural understanding of human and non-human metaphors influences perception. Additionally, longitudinal studies gauging perceptions of using metaphorical VUIs would extend our understanding of the applications of non-human metaphors in conversation design.

The aim of this study is to introduce the possibility of designing VUIs using non-human metaphors to conversation designers. We present this study as the foundation from which the discourse can expand, acknowledging the numerous pathways for advancing this line of inquiry further:

- **Organizing workshops** will give us an opportunity to bring an interdisciplinary community together, encompassing conversation designers, cognitive psychologists, and developers to continue discussion regarding the importance of metaphors in VUI design.
- **Longitudinal studies** are necessary to assess if the observed impact of different metaphors persists over time and evaluate how it differs based on the context of the interaction (e.g., 'lab' vs. 'in-the-wild' studies).
- **Working VUI prototypes** will be developed based on different metaphors in collaboration with colleagues from industry. The aim of developing these prototypes will be to improve the ecological validity and obtain more realistic interaction insights that more closely reflect the real-life interactions between users and VUIs.
- **Different types of voices** will be synthesized and evaluated. Specifically, we will experiment with different genders and accents of voices to investigate the suitability of voices with different prosodic qualities (pitch, tempo, volume, etc.) on audiences that differ in terms of gender and cultural background.
- **Different metaphors** that go beyond the realm of finance and health will be explored. We will experiment with domains such as fantasy characters (e.g., 'Gandalf', 'Jar Jar Binks') and focus on the metaphors that are relatable to specific audiences (e.g., 'Lord of the Rings' fans, 'Star Wars' community).

We hope that the insights gathered via the above action points will help researchers and conversation designers to create VUIs that are more suitable and better tailored for specific interaction contexts, potentially improving user experience and increasing the likelihood of adoption.

ACKNOWLEDGMENTS

Research supported by the Horizon 2020 FET program of the European Union through the ERA-NET Cofund funding (grant CHIST-ERA-20-BCI-001) and the European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147).

REFERENCES

- [1] Ahlam Alnifaie, Sonika Singh, Baki Kocaballi, and Mukesh Prasad. 2021. An overview of conversational agent: applications, challenges and future directions. In *17th International Conference on Web Information Systems and Technologies*. SCITEPRESS-Science and Technology Publications.
- [2] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. 2019. Music, search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 3 (2019), 1–28.
- [3] Theo Araujo. 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior* 85 (2018), 183–189.
- [4] Gaurav Bansal, David Gefen, et al. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision support systems* 49, 2 (2010), 138–150.
- [5] Christoph Bartneck. 2023. *Godspeed Questionnaire Series: Translations and Usage*. Springer International Publishing, Cham, 1–35. https://doi.org/10.1007/978-3-030-89738-3_24-1
- [6] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- [7] Timothy W Bickmore, Laura M Pfeifer, and Brian W Jack. 2009. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1265–1274.
- [8] Alan F. Blackwell. 2006. The reification of metaphor as a design tool. *ACM Transactions on Computer-Human Interaction* 13, 4 (Dec. 2006), 490–530. <https://doi.org/10.1145/1188816.1188820>
- [9] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland UK, 1–11. <https://doi.org/10.1145/3290605.3300270>
- [10] Lynne Cameron and Robert Maslen. 2010. *Metaphor Analysis: Research Practice in applied Linguistics, Social Sciences and the Humanities*. Equinox, London. <http://www.equinoxpub.com/books/showbook.asp?bkid=363&keyword=9781845534479>
- [11] Heloisa Candello, Claudio Pinhanez, David Millen, and Bruna Daniele Andrade. 2017. Shaping the experience of a cognitive investment adviser. In *Design, User Experience, and Usability: Understanding Users and Contexts: 6th International Conference, DUXU 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part III* 6. Springer, 594–613.
- [12] Amy Cheng, Vaishnavi Raghavaraju, Jayanth Kanugo, Yohanes P Handrianto, and Yi Shang. 2018. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1–5.
- [13] Jessie Chin, Smit Desai, Sheny Lin, and Shannon Mejia. 2024. Like my aunt Dorothy: Effects of conversational styles on perceptions, acceptance, and metaphorical descriptions of voice assistants during later adulthood. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (Apr 2024), 1–22. <https://doi.org/10.1145/3637365>
- [14] Minji Cho, Sang-su Lee, and Kun-Pyo Lee. 2019. Once a Kind Friend is Now a Thing: Understanding How Conversational Agents at Home are Forgotten. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) (DIS '19). Association for Computing Machinery, New York, NY, USA, 1557–1569. <https://doi.org/10.1145/3322276.3322332>
- [15] T. R. Colburn and G. M. Shute. 2008. Metaphor in computer science. *Journal of Applied Logic* 6, 4 (Dec. 2008), 526–533. <https://doi.org/10.1016/j.jal.2008.09.005>
- [16] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What can i help you with?": infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) (MobileHCI '17). Association for Computing Machinery, New York, NY, USA, Article 43, 12 pages. <https://doi.org/10.1145/3098279.3098539>
- [17] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–13.
- [18] Amanda Cercas Curry, Judy Robertson, and Verena Rieser. 2020. Conversational assistants and gender stereotypes: Public perceptions and desiderata for voice personas. In *Proceedings of the second workshop on gender bias in natural language processing*. 72–78.
- [19] Min-Yuh Day, Jian-Ting Lin, and Yuan-Chih Chen. 2018. Artificial intelligence for conversational robo-advisor. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 1057–1064.
- [20] Smit Desai and Jessie Chin. 2023. OK Google, Let's Learn: Using Voice User Interfaces for Informal Self-Regulated Learning of Health Topics among Younger and Older Adults. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–21. <https://doi.org/10.1145/3544548.3581507>
- [21] Smit Desai, Xinhui Hu, Morgan Lundy, and Jessie Chin. 2023. Using Experience-Based Participatory Approach to Design Interactive Voice User Interfaces for Delivering Physical Activity Programs with Older Adults. In *Proceedings of the 11th International Conference on Human-Agent Interaction (HAI '23)*. Association for Computing Machinery, New York, NY, USA, 180–190. <https://doi.org/10.1145/3544548.3581507>

1145/3623809.3623827

[22] Smit Desai, Morgan Lundy, and Jessie Chin. 2023. “A Painless Way to Learn:” Designing an Interactive Storytelling Voice User Interface to Engage Older Adults in Informal Health Information Learning. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3571884.3597141>

[23] Smit Desai and Michael Twidle. 2022. Is Alexa like a computer? A search engine? A friend? A silly child? Yes.. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3543829.3544535>

[24] Smit Desai and Michael Twidle. 2023. Metaphors in Voice User Interfaces: A Slippery Fish. *ACM Transactions on Computer-Human Interaction* 30, 6 (Dec. 2023), 1–37. <https://doi.org/10.1145/3609326>

[25] Michael A. DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What It Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173694>

[26] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive synthetic speech: Voice perception and user behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–9.

[27] Mateusz Dubiel, Anastasia Sergeeva, and Luis A. Leiva. 2024. Impact of Voice Fidelity on Decision Making: A Potential Dark Pattern? [arXiv:2402.07010](https://arxiv.org/abs/2402.07010) [cs.HC]

[28] Jens Edlund. 2019. Shoehorning in the name of science. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, 1–3. <https://doi.org/10.1145/3342775.3342794>

[29] Ahmed Fadhl, Yunlong Wang, and Harald Reiterer. 2019. Assistive conversational agent for health coaching: a validation study. *Methods of information in medicine* 58, 01 (2019), 009–023.

[30] Brett A Halperin, Gary Hsieh, Erin McElroy, James Pierce, and Daniela K Rosner. 2023. Probing a Community-Based Conversational Storytelling Agent to Document Digital Stories of Housing Insecurity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[31] Paul Heckel. 1984. *The elements of friendly software design*. New York: Warner Books. <http://archive.org/details/elementsoffriend00heck>

[32] Christian Hildebrand and Anouk Bergner. 2021. Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science* 49 (2021), 659–676.

[33] Frank D Hodge, Kim I Mendoza, and Roshan K Sinha. 2021. The effect of humanizing robo-advisors on investor judgments. *Contemporary Accounting Research* 38, 1 (2021), 770–792.

[34] B. Indurkha. 2013. *Metaphor and Cognition: An Interactionist Approach*. Springer Science & Business Media. Google-Books-ID: foTrCAAAQBAJ.

[35] Ray Jackendoff and David Aaron. 1991. Review of More Than Cool Reason: A Field Guide to Poetic Metaphor. *Language* 67, 2 (1991), 320–338. <https://doi.org/10.2307/415109>

[36] Jiani Yin Jian, Ann M. Bisantz, and Colin G. Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

[37] Hyunhoon Jung, Hee Jae Kim, Seongeun So, Jinjoong Kim, and Changhoon Oh. 2019. TurtleTalk: An Educational Programming Game for Children with Voice User Interface. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3290607.3312773>

[38] Ji-Young Jung, Sihang Qiu, Alessandro Bozzon, and Ujwal Gadhiraju. 2022. Great Chain of Agents: The Role of Metaphorical Representation of Agents in Conversational Crowdsourcing. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–22. <https://doi.org/10.1145/3491102.3517653>

[39] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T. Hancock, and Michael S. Bernstein. 2020. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–26. <https://doi.org/10.1145/3415234>

[40] Sunyoung Kim and Abhishek Choudhury. 2021. Exploring older adults’ perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior* 124 (Nov. 2021), 106914. <https://doi.org/10.1016/j.chb.2021.106914>

[41] Taenyun Kim, Maria D. Molina, Minjin (MJ) Rheu, Emily S. Zhan, and Wei Peng. 2023. One AI Does Not Fit All: A Cluster Analysis of the Laypeople’s Perception of AI Roles. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–20. <https://doi.org/10.1145/3544548.3581340>

[42] Rafal Kocielnik, Raina Langevin, James S George, Shota Akenaga, Amelia Wang, Darwin P Jones, Alexander Argyle, Callan Fockele, Layla Anderson, Dennis T Hsieh, et al. 2021. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–10.

[43] Anastasia Kuzminykh, Jenny Sun, Nivetha Govindaraju, Jeff Avery, and Edward Lank. 2020. Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376665>

[44] George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press. Google-Books-ID: r6nOYVtzUoC.

[45] Minha Lee, Lily Frank, and Wijnand IJsselstein. 2021. Brokerbot: A Cryptocurrency Chatbot in the Social-technical Gap of Trust. *Computer Supported Cooperative Work (CSCW)* 30, 1 (Feb. 2021), 79–117. <https://doi.org/10.1007/s10606-021-09392-6>

[46] Q. Vera Liao and Jennifer Wortman Vaughan. 2023. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. [arXiv:2306.01941](https://arxiv.org/abs/2306.01941) [cs.HC]

[47] Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. 2019. Resistance to Medical Artificial Intelligence. *Journal of Consumer Research* 46, 4 (Dec. 2019), 629–650. <https://doi.org/10.1093/jcr/ucz013>

[48] 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 (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. <https://doi.org/10.1145/2858036.2858288>

[49] Maria Luce Lupetti, Emma Hagens, Willem Van Der Maden, Régine Steegmans-Theunissen, and Melek Rousian. 2023. Trustworthy Embodied Conversational Agents for Healthcare: A Design Exploration of Embodied Conversational Agents for the periconception period at Erasmus MC. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. 1–14.

[50] David D Luxton. 2020. Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization* 98, 4 (2020), 285.

[51] Robert R. McCrae and Paul T. Costa. 1986. Personality, coping, and coping effectiveness in an adult sample. *Journal of Personality* 54, 2 (1986), 385–405. <https://doi.org/10.1111/j.1467-6494.1986.tb00401.x>

[52] Donald McMillan and Razan Jaber. 2021. Leaving the Butler Behind: The Future of Role Reproduction in CUI. In *CUI 2021 - 3rd Conference on Conversational User Interfaces (CUI '21)*. Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3469595.3469606>

[53] Adam S Miner, Arnold Milstein, Stephen Schueller, Roshini Hegde, Christina Mangurian, and Eleni Linos. 2016. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA internal medicine* 176, 5 (2016), 619–625.

[54] Stefan Morana, Ulrich Gnewuch, Dominik Jung, and Carsten Granig. 2020. The Effect of Anthropomorphism on Investment Decision-Making with Robo-Advisor Chatbots.. In *ECIS*.

[55] Karin S. Moser. 2000. Metaphor Analysis in Psychology—Method, Theory, and Fields of Application. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 1, 22 (June 2000). <https://doi.org/10.17169/fqs-1.2.1090>

[56] Nasim Motalebi, Eugene Cho, S. Shyam Sundar, and Saeed Abdullah. 2019. Can Alexa be your Therapist? How Back-Channeling Transforms Smart-Speakers to be Active Listeners. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 309–313. <https://doi.org/10.1145/331957.3359502>

[57] Sara Moussawi, Marios Koufaris, and Raquel Benbunan-Fich. 2021. How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets* 31, 2 (June 2021), 343–364. <https://doi.org/10.1007/s12525-020-00411-w>

[58] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56, 1 (2000), 81–103. <https://doi.org/10.1111/0022-4537.00153>

[59] William M. Newman and Robert F. Sproull. 1979. *Principles of Interactive Computer Graphics*. McGraw-Hill Education, Auckland.

[60] Don Norman. 2013. *The Design Of Everyday Things* (revised edition ed.). Basic Books, New York, New York.

[61] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3174214>

[62] 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 (Nov. 2019), 1–21. <https://doi.org/10.1145/3359316>

[63] Alisha Pradhan and Amanda Lazar. 2021. Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents. In *Proceedings of the 3rd Conference on Conversational User Interfaces (CUI '21)*. Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3469595.3469607>

- [64] Lingyun Qiu and Izak Benbasat. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of management information systems* 25, 4 (2009), 145–182.
- [65] Leon Reicherts, Gun Woo Park, and Yvonne Rogers. 2022. Extending Chatbots to probe users: Enhancing complex decision-making through probing conversations. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–10.
- [66] M. Reimer. 1996. The Problem of Dead Metaphors. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 82, 1 (1996), 13–25.
- [67] Malak Sadek, Rafael A Calvo, and Celine Mougenot. 2023. Trends, Challenges and Processes in Conversational Agent Design: Exploring Practitioners' Views through Semi-Structured Interviews. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23)*. Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/3571884.3597143>
- [68] Amy Schiller and John McMahon. 2019. Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence. *New Political Science* 41, 2 (April 2019), 173–191. <https://doi.org/10.1080/07393148.2019.1595288>
- [69] Ryan M Schuetzler, G Mark Grimes, Justin Scott Giboney, and Jay F Nunamaker Jr. 2018. The influence of conversational agents on socially desirable responding. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. 283.
- [70] 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 (DIS '18)*. ACM, New York, NY, USA, 857–868. <https://doi.org/10.1145/3196709.3196772> 00000 event-place: Hong Kong, China.
- [71] Katie Seaborn and Jacqueline Urakami. 2021. Measuring Voice UX Quantitatively: A Rapid Review. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3411763.3451712>
- [72] James Simpson and Cassandra Crone. 2022. Should Alexa be a Police Officer, a Doctor, or a Priest?: Towards CUI Relationships Worth Having. In *Proceedings of the 4th Conference on Conversational User Interfaces*. ACM, Glasgow United Kingdom, 1–5. <https://doi.org/10.1145/3543829.3544522>
- [73] Tim Stokes, Mary Dixon-Woods, and Robert K McKinley. 2004. Ending the doctor–patient relationship in general practice: a proposed model. *Family Practice* 21, 5 (Oct. 2004), 507–514. <https://doi.org/10.1093/fampra/cmh506>
- [74] Lorainne Tudor Car, Dhakshenya Ardithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research* 22, 8 (2020), e17158.
- [75] Victoria Turk. 2016. Home invasion. *New Scientist* 232, 3104 (Dec. 2016), 16–17. [https://doi.org/10.1016/S0262-4079\(16\)32318-1](https://doi.org/10.1016/S0262-4079(16)32318-1)
- [76] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S Kashavan, and John Blake Torous. 2019. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry* 64, 7 (2019), 456–464.
- [77] Alastair van Heerden, Xolani Ntinga, and Khanya Vilakazi. 2017. The potential of conversational agents to provide a rapid HIV counseling and testing services. In *2017 international conference on the frontiers and advances in data science (FADS)*. IEEE, 80–85.
- [78] Sarah Theres Völkel, Samantha Meindl, and Heinrich Hussmann. 2021. Manipulating and evaluating levels of personality perceptions of voice assistants through enactment-based dialogue design. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–12.
- [79] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S. Shyam Sundar. 2020. Alexa as Coach: Leveraging Smart Speakers to Build Social Agents that Reduce Public Speaking Anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376561>
- [80] Christina Ziying Wei, Young-Ho Kim, and Anastasia Kuzminykh. 2023. The Bot on Speaking Terms: The Effects of Conversation Architecture on Perceptions of Conversational Agents. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3571884.3597139>
- [81] Xi Yang, Marco Aurisicchio, and Weston Baxter. 2019. Understanding Affective Experiences with Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 542:1–542:12. <https://doi.org/10.1145/3290605.3300772> 00000 event-place: Glasgow, Scotland UK.
- [82] Hui Zhu, Eva-Lotta Pysander, and Ingall Soderberg. 2023. Not transparent and incomprehensible: A qualitative user study of an AI-empowered financial advisory system. *Data and Information Management* 7 (April 2023), 100041. <https://doi.org/10.1016/j.dim.2023.100041>