

AI and Design Opportunities for Smart Speakers



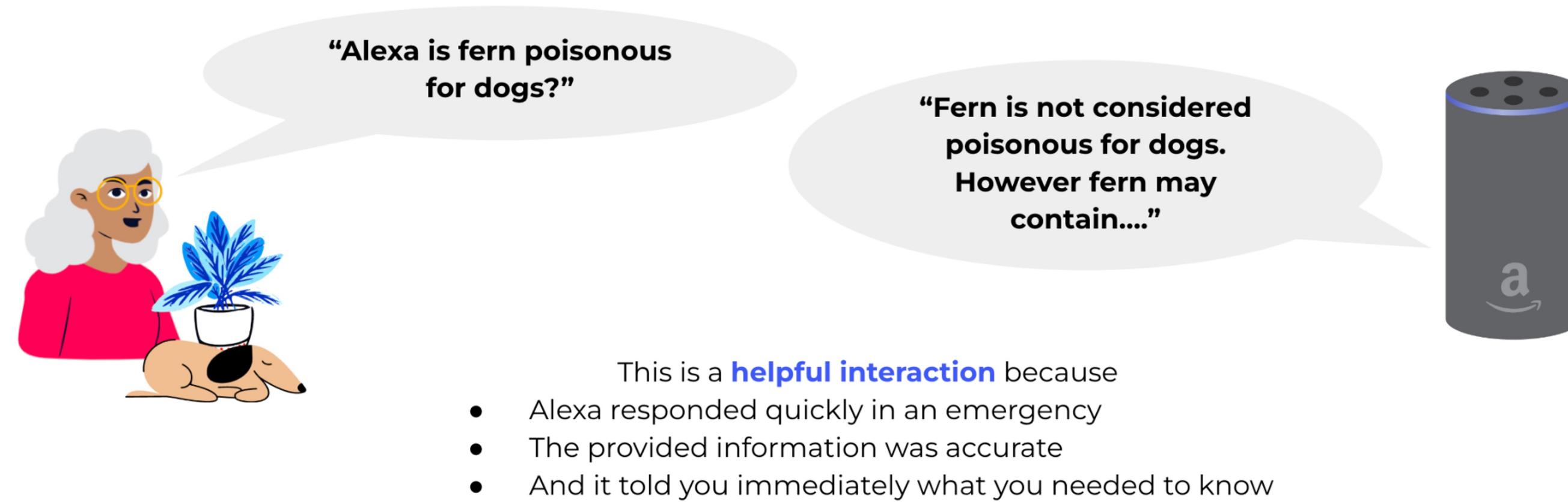
Tao Long, Lydia B. Chilton
long, chilton@cs.columbia.edu



Introduction

Smart speakers, Alexa, Siri, and Google Home, have opened up the potential for many **new types of interaction** that desktops and phones are not designed to do.

💡 **Information While Primary Task!** As they are voice-operated, users can interact with smart speakers while their hands are busy (while dressing, or cooking), or they have limited visual attention (while driving or taking care of children).



Study 1

📖 To understand the landscape of known challenges in designing VUIs, we conducted a **systematic literature review** of 35 research papers containing 125 design principles.

Themes	Sub-themes	Example Guidelines
Enhance Basic Usability (52)	Promote Error Prevention (21)	A12. Confirm input intelligently: [current VUIs] failed to explicitly confirm some critical actions [, like double-checking which alarm to turn off] [77]. [VUIs] need to provide feedback to the user explaining their interpretation of the [ambiguous or underspecified utterance] and how it was handled [64].
	Strengthen User Control (13)	Provide ability for users to control and interrupt [74]. The control users have on the processing of their actions by the system. [50].
	Leverage Multimodality (7)	Address [low situation awareness] by providing better vocal feedback... and adding visual indication for information that is critical to the user [37] Leverage task context and multimodality in order to provide visual or other non-verbal cues [4].
	Improve Compatibility (4)	Smart Home Framework: the compatibility of the [VUIs] with smart home devices [38]. Integrate [the VUIs with] not only smartphones but also connected televisions, computers, and other screen-based devices [63].
	Enhance Discoverability (4)	Use responses to help users discover what is possible...rather than always say something is impossible: the system did not teach ways to ask for a result, and [users] had to guess and try multiple times [77]. Data mining to offer new features: ...significant challenge for [VUIs]. One opportunity is to data mine repeated patterns of use or use common routines as scaffolding to introduce new related features [63].
	Facilitate Error Recovery (3)	Provide interface affordances (visual or language) so users can refine and repair system choices [64]. A17. Allow users to exit from errors or a mistaken conversation: Use a special escape word globally (e.g. "Stop")... [or] non-speech methods when speech fails (e.g., push a physical button) [77].
	Remember Usage History (13)	Remember User Profiles to Deliver Personalized Services: store a vast amount of information specific to the user, such as personal profiles or preferences [29] Enable user to employ territorial markers to help them avoid activity- and preference-related conflicts by communicating their preferences or staking a claim to the device or data [20].
Customize for User Contexts (23)	Design for Diverse & Sensitive Populations (6)	Enriching the responding contents when executing these commands might give older adults a chance to find more features and functionalities [28]. Adopt social-justice oriented design methods...when building [VUIs] in home health care contexts [5].
	Leverage User's Environment (4)	Integrate the functionality of ubicomp sensors: install additional sensors in the home that can serve as automatic warning alarms, for example, if the stove is left on or if there is a water leak. [or] whether someone is within earshot before triggering interactions [59]. Leverage knowledge of place: [As the VUIs] is in a living room versus a bedroom, it can modify its level of proactivity, listen for particular commands, and offer suggestions for new uses [50].
	Accommodate Conversational Speech (17)	Enhancing the message interactivity of the human-[VUIs] conversation by increasing the degree of contingency in message exchanges [76]. Tailor responses and follow-up questions to make interactions more engaging, elicit in-depth disclosure, and effectively provide emotional support through these devices most natural responses [65].
Speak Users' Language (23)	Anthropomorphize Human's Emotions (6)	Employ empathetic expressions to show emotional responsiveness. When [VUIs] use phrases such as "I understand" or "I can relate to you"... users are likely to perceive it being highly social [76]. Express Sympathy; Be Interesting, Charming, and Lovable; Express Interest to Users [62].
	Design for Shorter Conversation (8)	Design for short interactions, know when it will be long: Systems can prepare for the large majority of interactions to be a single command-answer or command-action [4]. A11. Keep feedback and prompts short: ... the [current VUIs'] responses were not always clear or succinct, making it difficult for users to listen, understand, and remember [77].
	Design for Simpler Conversation (8)	Minimize acoustic confusability of vocabulary [77]. Guide users through a conversation so they are not easily lost [45]
Establish Reliability (9)	Establish Reliability (9)	Add additional feedback currently the [VUIs] says "I will remind you" if you set a reminder, but perhaps it could also have an external visual cue that is active only if there is still a pending reminder. This additional feedback would minimize the need for participants to repeatedly ask what their reminders are and could contribute to building trust with the [VUI] system [59]. Acknowledgments and confirmations: To build trust, acknowledgments need to be provided as feedback indicating that the user's input was received [3].

Responding to information overload led by the high-maintenance tech, **Calm Technology** helps users be focused rather than creating panic. #Need Minimal Attention, #Inform Calm, #Utilize Peripheral Awareness.



Fig. 1. Attention Graph for a Boiling Tea Kettle by Amber Case [2].

💡 **Underused?** However, despite the potential, many user experience studies and marketplace analysis report that VUI users **infrequently use the device or abandon it** due to cumbersome setup experiences, long response, and difficulty in discovering new use cases.



"Alexa, what's the weather going to be like for the next week"
"On Monday it's going to be 70 degrees with rain, on Tuesday...on Wednesday...on Thursday..."

Alexa plays a podcast

disappointed sigh

This **frustrates** users because...

- While Alexa mostly understood correctly...
 - The user had to **interact with alexa several times** afterwards **to clarify** the question
 - Alexa didn't take initiative to **specify** what exactly the user wanted to hear and instead just listed as much information as possible
 - Alexa gave answers that were **too long** for what the user needed
 - It left users frustrated and they **stopped the interaction**.

Study 2

📖 To complement our understanding from the literature review, we conducted **in-depth user interviews** with 15 Alexa users to better understand the 1) **challenges** that contribute to the non-use and identify 2) **opportunities** for future improvement.

🚧 Challenges: Misaligned Expectations of Its Capabilities Efficiency, Task Complexity, Poor Discoverability, Lack of Affective Responsiveness

Opportunities: Incorporating the Vision of Calm Technology

1. Explore Use Cases to Understand and Support **Information While Primary Task**
 - News **while** having breakfast
 - Weather **while** getting dressed
 - Traffic updates **while** heading out of home
 - Sports scores **while** making coffee
 - Shopping list **while** cleaning the fridge
 - Recipe information **while** cooking
 - Music **while** taking a shower
 - Setting a timer **while** an at-home exercise
 - Guided meditation **while** preparing to sleep
2. Focus on **Short Commands**
3. Improve **Input Efficiency** by Learning User Preferences
4. Improve Discoverability by Promoting **Social Learning** from Observations and Tutorial Videos