Accountability-Aware Design of Voice User Interfacesfor Home Appliances

Anonymous Author(s)

Abstract

The availability of voice-user interfaces (VUIs) has grown dramatically in recent years. As more capable systems invite higher expectations, the conversational interactions that VUIs support introduce ambiguity in *accountability*: who (user or system) is understood to be responsible for the outcome of user-delegated tasks. When misconstrued, impact ranges from inconvenience to deadly harm. This project explores how users' accountability perceptions and expectations can be managed in voice interaction with smart home appliances. To explore links between degree of automation, system accountability and user satisfaction, we identified key design factors for VUI design through an exploratory study, articulated them in video prototypes of four new VUI mechanisms showing a user commanding an advanced appliance and encountering a problem, and deployed them in a second study. We found that participants perceived automated systems as more accountable, and were also more satisfied with them.

Index Terms: Human-centered computing—Interface design prototyping; Auditory feedback; User Interface design;

1 Introduction

Advances in artificial intelligence (AI) are changing how users interact with software agents. AI-infused systems vary in the level of automation they present to users: they can recommend options, make decisions, communicate with other agents, and adapt to their environments [6, 62]. A *Voice User Interface* (VUI) is a type of user interface that relies on speech recognition to communicate with users, usually with a conversational style [15] that resembles natural verbal intercourse, rather than manual clicking or typing.

Assistant-type VUIs are growing in popularity on personal devices. The majority of Americans own a smartphone (81%) or tablet (52%), which today come equipped with Siri, Google Assistant, or equivalent VUIs [61]. 1 billion devices worldwide running Windows 10 [43] provide access to Cortana, Microsoft's voice assistant. Access is different from use, but devices that exclusively accept voice input, *e.g.*, Amazon Echo and Google Home are on the rise: over 100 million Alexa-enabled units had been sold as of January 2019 [9]. It is clear that many consumers are newly choosing and trying voice interaction in their everyday life [2]. With this prevalence, we can go beyond 1st-order traits of this modality (hands-free, natural language) to examine factors such as social consequences.

Often, VUI technologies are used for requests for information or to trigger applications [7]: failure is annoying but not dire. However, as voice recognition technology improves and systems are capable of greater automation, users hold VUI systems to human-like standards of behavior. During VUI conversations with smart devices, users seem to anticipate accountability along similar lines as they would with humans; *e.g.*, Porcheron et al. describe users expecting appropriate responses, both verbal and actionable, from Alexa and express dismay when these are not provided. If the system makes an utterance, they react as they would to a human utterance [52].

However, AI systems that enable automation typically work under uncertainty, balancing false-negative and false-positive errors with potentially confusing and disruptive results [6]. Impact widens as standalone systems become platforms that control many home technologies through the Internet of Things (IoT) [7, 36, 45]. With Google Home, users can adjust lighting and set the thermostat [33], but also interact with systems invoking larger consequences. Smart washing machines can ruin clothes; personal assistant devices can spend money online. A semi-autonomous car can crash and kill in a moment of ambiguity over who is in charge.

So, what happens in the case of a bad outcome? Does the user hold the system responsible, or themselves?

Accountability can be defined as who (*e.g.*, user or system) is obligated or willing to be responsible or accountable for the satisfactory execution of a task, including one that has been delegated [3]. It is fundamental to how people conceptualize their actions and react to outcomes in a social context, by considering risk, uncertainty, and trust when taking or delegating ownership of outcomes [13]. Both a societal and individual concept, accountability differs subtly from responsibility; it is possible to be responsible (in charge) without being accountable if you take action but not ownership of the results.

Can interaction design mediate this balance, when it is important that the user retain accountability for a delegated task? We focus on ***perceived* accountability (hereafter accountability**), which varies with user experience and expectations as well as situation (*e.g.*, degree of automation actually available), and therefore can vary by instance [62]. These factors impact a user's perception of system capability [18]. Because of these interlinked perceptions, we posit that through design we can manage user perception of a system's automation, and influence their notion of accountability.

Research Questions: We consider two questions in the context of VUI interaction with advanced home appliances:

RQ1: What design factors impact user perception of system accountability?

RQ2: How does automation influence perceived accountability and user satisfaction? Can interface design mediate this influence?

Approach: An invitation from industry colleagues to investigate user experience with voice-controlled smart appliances led us to consider VUIs in terms of user types, social roles, privacy and value added. In a first exploratory study, *accountability perception* emerged as an important and understudied factor. We used insights from this exploration to propose primary *design factors* that can influence accountability during the user's interaction with the system (RQ1).

To go deeper, we studied more carefully how varying the level of system automation could influence user perception of system accountability. Our work is in the tradition of HCI community-proposed guidelines and recommendations for interaction with AI-infused systems. Early work by Norman [48] and Höök [30] sets out guidelines for avoiding unfavourable actions during interaction with intelligent systems, aiming to safeguard the final outcome by managing autonomy and requiring verification. Horvitz proposed a mixed-initiative method to balance automation with direct manipulation by users [29]. While these works discuss cautionary actions to avoid potential problems, their impact on users' perception of accountability in case of failure is unexplored.

We constructed video prototypes [68] featuring different VUI interaction scenarios, as design probes to provoke open dialogues with participants about accountability for each interaction scenario. These prototypes present four types of VUI mechanisms that vary level of automation, all set in a home environment, enabling participants to compare accountability delegation.

Through in-person interviews and online questionnaires, we obtained participants' rankings of accountability and their satisfaction of each mechanism in light of the task failure illustrated. We analyzed this data in relation to the represented system automation.

In contribution, we:

- Propose primary design factors of accountability in voice user interactions with complex technology: *task complexity*, *command ambiguity* and *user classification*.
- Demonstrate the ability to *direct accountability perception* through VUI design.
- Provide insights on the relationship between user satisfaction and system automation.

2 Related Work

2.1 Automation, Interaction & Accountability

Internet of Things and Smart Home Appliances: IoT technology connects users and environmentally-embedded "smart" objects, from individual gadgets (like smartphones and smartwatches) [31], communal appliances (smart speakers, thermostats, vacuum robots), to semi-autonomous systems and sensor networks) [42, 67]. The exploding number of IoT devices and complexity of controlling them can negatively impact user attitudes [49]. We note that while considerable smart technology is available today, our study scenarios imbue VUIs with slightly futuristic decision-making ability.

Intelligent user interfaces (IUI): In addition to sensor capacity and the IoT, some smart home appliances benefit from embedded IUIs. IUIs simplify interactions through AI capabilities such as adaptation [32] and the ability to respond to natural commands and queries with apparent social intelligence. Suchman's discussion of situated action highlights the need for context-dependent responses in HCI [63]. Although situated-action models result in versatile and conversational systems, this approach is based on probabilistic behaviour which is prone to unexpected errors.

Task delegation to AI: Studies in AI-human interaction focused primarily on systems capabilities such as reliability and cost of decision (*e.g.*, [51]). Considering human preferences and perceptions, Lubras *et al.* summarize the literature in shared control between humans and AI, and propose four factors for AI-task delegation: risk, trust, motivation, and difficulty. Emphasizing human perception, their research supports the human-in-the-loop design and low preference for automation [41]. However, the user's accountability perception generally does not appear in these studies.

Explainable-accountable AI: For both usability and ethical reasons, algorithmically derived decisions should be *explainable* [19, 34, 58]. Systems utilizing them should provide accounts of their behaviour and inform users about sensor information, resulting decisions, and likely consequences [8, 20]. Some argue for policies on automated decision making [22], and a few governments have established regulations that require AI systems to provide users with explanations about any algorithmic decisions [24]. Explainable AI enables human to make sense of the machine learning models and understand the rationales behind AI decisions. Abdul *et al.*reviewed over 12,000 papers from diverse communities on trends in explainable-accountable AI [4]. They highlight a lack of HCI research on practical, usable, and effective AI solutions. Other groups have found that classic UX design principles may be insufficient for AI-infused products, and we need to develop guidelines specifically for human-AI interaction [6] — a motive of the present work.

Automation and accountability: Previous work on accountability mainly focused on social accountability among humans, and how people justify or explain their judgments [10, 65]. Some, however, investigate accountability during collaborative decision making of a human and an intelligent agent [18, 46, 57]. Skitka *et al.*show that holding users accountable for their performance reduces automation bias (too much trust in automated decision makers), improves performance and reduces errors [60]. Suchman shows how the agency attributed to a human or a machine is constructed during an interaction [63]. Others have investigated how users negotiate and interpret their agency while interacting with VUIs [38, 59].

However, no study has yet shown how to *direct* the perception of accountability through design. Accountability research in HCI goes beyond usability and deserves significantly more attention from intelligent user interface designers, including VUI designers.

Control Capabilities: Building on works from Dourish, Button and Suchman [12, 20, 63], Boos *et al.* propose that users feel they have control over a system based on its "**control capabilities**" – specifically, when it is *transparent*, *predictable* and can be *influenced* [10]. The authors further suggest that users who feel in control of an interaction are more likely to consider themselves accountable for the outcome. We aim to determine whether users can identify subtle differences in control capabilities, and whether that affects the accountability of the system.

2.2 Voice-User Interfaces

We use Porcheron's definition of a VUI, which specifies interfaces that rely primarily on voice, such as Amazon's Alexa or the Google Assistant [52]. They are always on and can be accessed from room-level distances, which results in them being highly "embedded in the life of the home" compared to other technologies. The quality of their human-centered design is imperative.

The union of VUIs and smart home appliances is largely unexplored despite its promise; *e.g.*, IoT is one of the most frequent VUI command categories that users employ in their daily interactions with home assistance devices [7]. While we see this as a great opportunity to incorporate VUI into home appliance technology, we heed Dourish's advice to "take sociological insights into the heart of the process and the fabric of design" [21].

Our work is distinguished from past efforts in VUI use in everyday life [52] by moving beyond understanding users' perception of accountability and trying to direct it through design.

3 EXPLORATORY STUDY

While some design factors have been identified at the boundary of automation and human interaction (trust, state learning, workload, machine accuracy, *etc.* [50]), we needed specific insights for VUI semi-automated systems. To answer RQ1, we investigated user experience with VUI products relative to non-VUI-controlled but "smart" home products. We did this through interviews (n=10) and questionnaires (n=43), recruiting through social media (Facebook, Twitter). The results, briefly summarized here, motivated using VUIs and suggested where accountability matters most.

3.1 Methods

Participants: We targeted past purchasers of smart home appliances. Of 43 questionnaire respondents (20/21/2 F/M/unreported), age range was 25-55 years, from Canada, USA, Colombia, UK, China and Australia. All did or had owned smart home appliances.

Questions: Participants reflected on their experiences with smart home appliances and voice-command technology, compared voice with other input modalities, and considered VUI integration for two hypothetical smart systems: lighting, and a washing machine. They were asked to imagine the functions these systems might fulfill through VUI commands, and explain any concerns.

3.2 Results

Accountability figured strongly in responses, emerging as a an under-explored design lever. Results further exposed three factors framing the situational impact of accountability: *User Classification*, *Task Complexity* and *Command Ambiguity*.

Motivations and De-motivations for VUIs: Our participants appreciated VUI speed, convenience (particularly hands-free use), multitasking, shallow learning curve, and natural language. In contrast to human conversations, they wished to minimize interactions. When describing envisioned VUI smart home appliances, we heard that they needed a "*machine that can decide for [itself].*" [P8] However, they were concerned about unreliability, hesitating to use VUI for complex tasks with irreversible outcomes, and concerned about misinterpretation, likely from prior experience.

Factor I – User Classification: We observed *primary* users, in charge of choice and maintenance, and *secondary* users reliant on the primary. Consistently, [56] notes that home technological management is not evenly distributed by gender or across the household.

Factor II – Task Complexity: Participants categorized home appliances mainly by interface complexity, not underlying technology. We thus subsequently focused on home appliances with more complicated UIs and non-trivial consequences of failure.

Factor III – Command Ambiguity: While positive overall, participants cited examples of concern which we categorize as naive access, hidden functionality and open-ended requests. Natural language is inherently ambiguous, requiring the system to make assumptions and decisions, as with human-human interactions. In so doing, accountability can be delegated – important to recognize should something go wrong. We seek design factors that influence this delegation.

4 Framework

4.1 Accountability via "Control Capabilities"

We explored how VUIs can affect accountability using Boos *et al.*'s theoretical framework of Control Capabilities (Section 2.1), which is based on the premise that "*in order to answer accountability demands [...], certain requirements of control need to be fulfilled*" [10]. We framed our experimental study around an extension of this proposition and framework, seeking to verify or disprove it. To the *Transparency* and *Predictability* dimensions proposed by Boos et al. [12] we added *Reliability* because of its prominence in our exploratory study.

Transparency: Transparency can be achieved through executing clear and understandable actions. Several studies recommend improving transparency by providing explanations about the behaviour of AI-empowered systems [28, 35, 40, 53].

Predictability: Predictability can be obtained by producing desired and anticipated outcomes. Human-AI guidelines suggest two points where interactions with an AI should be shaped: *over time* and *when wrong*. They advise that during an interaction, a system should convey updates to users regarding future consequences of the system's behaviour, and support invocation of requests as needed [6].

Reliability (added): Reliability can be achieved through delivery of desired outcomes based on given explanations. A well-studied construct in automated systems, trust is crucial in long-term adoption [27] and key for voice interaction [11]. To invoke trust, we chose *reliability*: the quality of performing the correct actions.

4.2 User Satisfaction as a Metric

User satisfaction with automation generally improves with reduced cognitive effort. However, a system can avoid accountability by requesting detail, *e.g.*, by providing choices or asking for confirmation. This increases user involvement, at the potential cost of satisfaction. Measuring user satisfaction as well as accountability perception indicates how well that balance is achieved.

We defined this metric based on principles of measuring customer satisfaction level [39, 55], then designed a questionnaire to assess emotional satisfaction by asking about: (a) overall quality (Attitudinal), (b) the extent user's needs are fulfilled (Affective and Cognitive) (c) users' feelings (Affective and Cognitive) [1].

5 Primary Study: Methods

5.1 Overview and Hypotheses

Since humans can manage accountability in their conversations, we surmise that designers should be able to enable this in human-machine interaction. We hypothesize a correlation between automation level (from fully machine-controlled to fully user-controlled), and system accountability. Our goal was to focus on how the level of automation influences both accountability and user satisfaction, which eventually inform the design of interactive systems.

We chose laundry as our focus task because modern washing machines require more engagement than other appliances (like refrigerators or toasters) and have a plethora of complex settings that can seem impenetrable and can cause confusion and errors. This complexity is what opens possibilities for guiding accountability perception, in a dialogue-type interaction. This also aligned with current industry activity: Samsung and LG have both developed washing machines with voice assistants.

We conducted a controlled experiment (15 survey respondents, of which 8 were also interviewed). Participants watched a series of video sketches [68] showing four levels of automation, where an individual uses a VUI with a smart washing machine for both simple and complex tasks. In every video, the washing machine fails to fulfill the user's expectations, since accountability is relevant primarily when the system fails.

Participants were instructed to imagine themselves as the user. We surveyed their perceptions of the washing machine's accountability for each VUI mechanism to obtain quantitative data, followed by open dialogues on accountability, satisfaction and general thoughts about each scenario.

Our hypotheses address the joint effect of system automation and task complexity on system accountability with the future goal of employing them in balance. We anticipated that:

H1: Increasing users' involvement in decision-making (thereby decreasing system automation) will reduce their perception of system accountability, particularly for high-complexity tasks.

H2: As we increasingly automate task decision-making, user satisfaction will increase.

If these hypotheses are correct, then system automation creates a trade-off between system accountability and user satisfaction. This work explores user perceptions surrounding this trade-off in the context of a VUI interaction. We also investigate the effects of task complexity on accountability and user satisfaction.

5.2 VUI Accountability-Directing Mechanisms

To direct users' accountability perception, we conceptualized four VUI mechanisms representing levels of system accountability, based on guidelines for a progression of automation in AI systems [29, 30, 48]: *automation*, *recommendation*, *instruction*, and *command*. We created video dialogues by following highly cited guidelines [5, 17, 25, 26]. The levels differ primarily in the degree of direct manipulation, automation and information conveyed, and method of information delivery. We captured the mechanisms in walkthrough-style video prototypes for use in the study task.

Automation presents a straightforward workflow: the user requests an outcome and the VUI notifies them of the action to be taken: *e.g.*, after the user states they would like to wash their clothes as quickly as possible, the machine chooses to execute a quick wash cycle. Because of the system's take-charge approach, we anticipate that users will regard this largely as a delegation of accountability to the system. This accountability delegation comes into play in failure cases. For example, in the above case of the quick wash cycle, if the clothes are not cleaned as effectively as a normal wash.

Recommendation provides options based on the user's description of the clothes. For example, after the user describes his clothes as *colored*, *made of no special material*, and *medium load*, the

mechanism provides two suggestions with different temperatures and spin speeds. The user selects one. We posit that here, the system is accountable for the *quality of recommendations*, but the user who makes the choice is ultimately responsible for the *outcome*.

Instruction provides the most information. It guides users in examining their clothes, and based on description and requirements, explains multiple washing suggestions. For example, after suggesting an *extra rinse*, the machine gives a detailed justification. If the user feels they have enough information, they can stop by saying 'Stop, I choose the first suggestion'. Here, we expect the user to hold the system accountable only for instruction accuracy.

Although users have equivalent choices in Instruction and Recommendation, they differ in presentation of the choices. We expect this to be reflected in the Control Capabilities measures of Transparency, Reliability and Predictability.

Command enables the user to set the washing cycle without any information from the machine. Users simply state their requirements instead of pushing buttons. This implies that the user knows what she wants. With this mechanism, we do not expect the user to hold the system accountable for the outcome.

5.3 Design and Variables

As we investigate influence of control capabilities (2.1), we rather than measures of the system's *actual* controllability by the user, these are guidelines whereby a designer can increase a user's sense of control They appear in two ways: informing mechanisms design (5.2); and as outcome measures (5.5), confirming whether this design manipulation was impactful.

The study itself uses within-subject 2x4 design, with independent variables of *complexity level* (low/high, described below), and *VUI mechanism* (4 mechanisms). For each complexity level, we counterbalanced the order of VUI mechanisms.

Task Complexity: With our exploratory study revealing the importance of task complexity and failure consequences on accountability perception, we varied task complexity for insight into H1 (whether design, via increased user involvement in decisions, can mediate perception of system accountability in high-complexity tasks).

Low complexity – Routine laundry, common in a household.

High complexity – A job involving special material (wool), extra requirements (stained fabric), and non-standard functions.

In our videos, for the low complexity condition the system attempted to remove mud from clothing; for high complexity, to remove wine stains from a valuable sweater. For all conditions, the washing machine failed to completely clean the clothes.

5.4 Procedure

We recruited homeowners with purchasing power for home appliances, using social media advertising and referral of participants (similar to [37, 44]), necessary due to the inclusion criteria and lack of participant compensation. We recruited a subset of survey participants to be interviewed in-person, immediately post-survey. The survey took an average of 30 minutes, and the follow-up interviews 20–45 minutes, average 27 minutes.

Survey participants answered a demographic questions and watched eight video prototypes: four distinct mechanisms, each performing a high- and low-complexity task. Videos were labelled by numerical order of appearance, counterbalanced by participant. Participants were then asked to rank the mechanisms by "how accountable each one was for the failed laundry task". Then, they scored each mechanism for Control Capabilities of Transparency, Predictability and Reliability and User Satisfaction. We asked the interviewee participant subset to verbally explain their responses.

5.5 Data Collected

The pre-video questionnaire (28 questions) collected participants' demographic information and past experience with non-smart washing machines, including whether they tended to hold non-smart washing machines "responsible for failed laundry tasks"). The post-video questionnaire collected participants' ratings for participants' satisfaction and Control Capabilities for each video (*i.e.*, VUI mechanism), while the interviews collected qualitative justifications of participants' survey responses.

Accountability Ranking: After watching each mechanism fail to complete the washing task, participants ranked them (1 (most) to 4 (least) accountable, tie not allowed). Ranking facilitated direct comparisons between short lists of items [47].

Control Capabilities: We sought participant opinions on Control Capabilities (Transparency, Predictability and Reliability) for each VUI mechanism as presented in the video prototypes. As with [14, 23], they scored each mechanism for Control Capabilities using a slider on a [0-100] point scale.

User Satisfaction: Again with a [0-100] point scale and a slider, we asked participants to respond to three questions:

- How easy would it be to use the voice-assisted system?
- How confusing was the voice-assisted system?
- How satisfied would you be with this interaction?

5.6 Analysis

Perceived Accountability Rankings: We performed Friedman tests (widely used for ranked data [47, 54]) on the mechanisms' accountability ratings to identify any correlation between accountability perception and system automation (which varied with the VUI mechanism in each video), for each level of task complexity (high or low). In post-hoc analysis, we used Bonferroni correction of confidence intervals to compare accountability rankings by VUI mechanism. For all statistical results, we report significance at $\alpha = 0.05$.

Control Capabilities & User Satisfaction: We analyzed each set of [0-100] scores with a repeated-measures ANOVA. Due to a violation of sphericity, we report Greenhouse-Geisser results. Post-hoc analysis included a Bonferroni alpha adjustment. User Satisfaction scores were taken by averaging participant responses to the three questions listed in 5.5, which provided a broad depiction of ease of use, clarity and interaction experience.

Interviews: We used Braun and Clark's approach for thematic analysis [16]. In repeated passes, two investigators conducted open coding. Afterwards, two other team members checked the coding and brought disagreements to the full team for resolution. This division provided a broader perspective, deepened our understanding and generated multiple discussions around each theme.

6 Results

We recruited 15 survey participants (10 male, 5 female, age distribution M = 34.97, SD = 7.86). Of these, we interviewed 8 (3 male, 5 female). Participants were from various ethnic backgrounds but all lived in North America at the time of recruitment.

6.1 Quantitative Results: Questionnaires

Pre-Questionnaire Data: We surveyed participants on their past experiences with household technology, including their technology roles within their households. In our exploratory study, we identified two role classifications. *Primary* users are enthusiastic about initial setup and ongoing maintenance of home technology; we designated other users as *secondary*. All respondents indicated they had purchasing power within their households.

$\sim 73\%$ of participants reported enthusiasm in exploring new features on their smart home appliances, and took responsibility for configuring home technology. This suggests that the majority of the participants were primary technology users based on our definition.

As an assessment of how participants related the notion "accountability" to washing machines, we asked where they placed the blame when a non-smart washing machine damaged their clothes. 60%

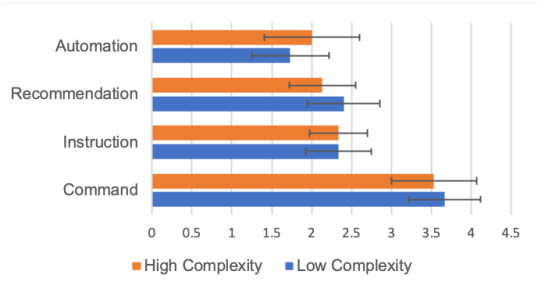

Figure 1: Average responsibility rankings of experiment's VUI mechanisms by task complexity, for question "*How accountable (responsible) is the system if something goes wrong*?" Rank 1 (greatest) to 4. Error bars are standard error of mean.

Table 1: Relative VUI Mechanism Accountability (Bonferroni-adjusted).

Mechanisms Compared	Low-Complexity	High-Complexity
Command-Automation	*z = -4.10, p < 0.001**	*z = -3.25, p = .007**
Command-Recommendation	*z = -2.97, p = 0.018**	*z = -2.97, p = .018**
Command-Instruction	*z = -2.83, p = 0.02**	*z = -2.546, p = .065*
Automation-Recommendation	*z = -1.131, p = 1.0*	*z = -.283, p = 1.0*
Automation-Instruction	*z = -1.273, p = 1.0*	*z = -.707, p = 1.0*
Recommendation-Instruction	*z = -.141, p = 1.0*	*z = -.424, p = 1.0*

had had that experience and "mostly" or "completely" blamed their washing machine. This seems to dispel a notion that perceived accountability skews towards self in such situations.

Perceived Accountability Rankings: Figure 1 shows participants' rankings of mechanism accountability for the portrayed outcome.

Friedman tests on task complexity found automation level, varied through mechanism type statistically significant for accountability ranking. For low-complexity tasks, $\chi^2(3, N = 15) = 18.28, p < 0.001*$; for high-complexity, $\chi^2(3, N = 15) = 13.32, p = 0.004*$. Post-hoc analysis (Table 1) with Bonferroni correction of confidence interval found that for both high and low complexity tasks, *Command* had significantly lower accountability than *Full Automation* and *Recommendation*. For low-complexity tasks, the *Command* mechanism had significantly lower accountability than *Instruction*.

Control Capability scores: We analyzed participant scores for each mechanism in the Control Capability (CC) dimensions of *Transparency*, *Predictability* and *Reliability*. Differences between CC scores for the *Recommendation* and *Instruction* mechanisms (Figure 2) suggest that option delivery impacts experience of control over the interaction. For example, though *Recommendation* and *Instruction* offer similar choices to users, *Recommendation* was consistently seen as less transparent, predictable and accountable.

Figure 2 reports average ratings for CC dimensions by VUI mechanism; it shows a trend suggesting that increased automation is linked to reduced perceived transparency and predictability. We found statistical significance only for predictability for the high complexity task. However, in our post-hoc test with Bonferroni alpha adjustment, we were not able to find any statistical significance between specific mechanisms for predictability.

Accountability and user satisfaction for low-complexity task: The trend of the average satisfaction scores in Figure 3 suggests that participants preferred the *Automation* mechanism to those requiring more user involvement, for both low and high-complexity tasks. Participants also reported higher satisfaction with *Instruction* and *Command* for high task complexity. However, we did not find statistical significance for either tasks (High Complexity: F(2.075, 29.05), p=0.576; Low Complexity: F(1.885, 26.391), p=0.258).

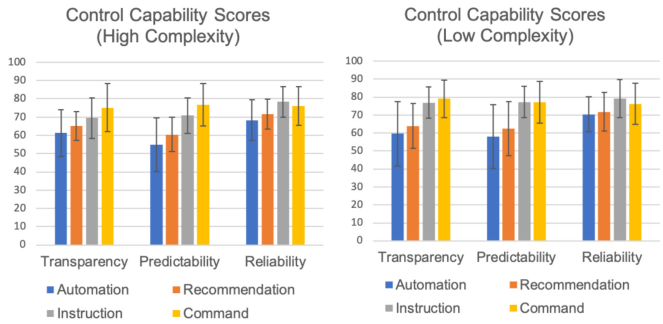

Figure 2: Participant scores (1-100) on control capabilities of VUI mechanisms (15 samples / bar). Error bars are standard error of the mean.

Table 2: ANOVA results for Control Capabilities and User Satisfaction

Ctrl Capability	Low Complexity	High Complexity
Transparency	*F(1.848, 25.877), p=1*	*F(1.887, 26.423), p=0.314*
Predictability	*F(1.514, 21.193), p=0.079*	*F(1.94, 27.164), p=0.028**
Reliability	*F(1.917, 26.834), p=0.305*	*F(2.044, 28.616), p=0.275*
Satisfaction	*F(1.885, 26.391), p=0.258*	*F(2.075,29.05), p=0.576*

Greenhouse-Geisser results are presented due to the violation of Sphericity. A post-hoc test with Bonferroni alpha adjustment was not significant for predictability.

6.2 Qualitative Results: Interviews

Eight questionnaire participants (4 female), selected through snowball sampling [66], were interviewed (Section 5.5). All were adults living with others who self-identified as primary or secondary users of home appliances, meaning they had purchasing power for home appliances in their households. No compensation was provided.

In the following we organize our analysis of the interview transcripts as laid out in Section 4. Mechanism names here replace the numerical labels that participants used to refer to the videos.

When asked to revisit their ranking of accountability across the VUI mechanisms, the majority of participants identified *Full Automation* as the most accountable. However, this was not unanimous: P3 suggested all mechanisms were "completely accountable", and P8 found the *Recommendation* mechanism most accountable. These individual variations further justify the use of Control Capabilities to identify design factors that contribute to accountability.

Automation is seen as most accountable: *Automation* was deemed by the majority of participants (both primary and secondary) as the most accountable, because the machine gives minimal information and selects the washing cycle by itself. "*Not given a choice*" [P4] and "*machine do[es] whatever... it feels the best*" [P7] are reasons that participants ranked the automation mechanism as most accountable.

> *'I think in [Automation], the machine should take the most responsibility since it makes all the decisions. . . '[P2]*

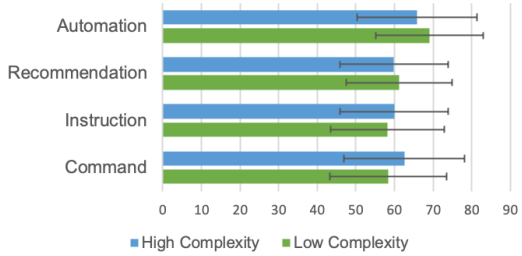

Figure 3: Average user satisfaction score (1-100) for low complexity and high complexity tasks, by VUI mechanism. Error bars: standard error of the mean.

Shared accountability: *Recommendation* and *Instruction* were viewed as generating a sense of shared accountability, by giving suggestions for participants to choose from. Participants are still accountable for the final decision: if "*something goes wrong, it should be your fault instead of the machine*" [P5]. However, "*maybe the user will blame the machine for giving [...] the wrong recommendation*" [P6] in the case of an error. Some suggested that since the system "*understands the situation [...] it's more accountable*" [P7].

Control Capability Dimensions:

A. Transparency – Participants generally agreed that all mechanisms were transparent enough for them to understand the interactions. As in Figure 3, most of participants rated and mentioned *Command* and *Instruction* as "more transparent" [P3, P4, P5, P7]. Some viewed *Command* as "*the most transparent*" [P4], since the user has complete control in this traditional method of doing laundry.

> *'...To me transparency relates to the extent to which they use or understand what's going on within the machine, so in video number one that I mentioned a machine is pretty much making the selections on behalf of the user [...] so the user doesn't really know what's going on. Whereas [Full Automation] for instructions, the machine just told me what the user is saying so the user one hundred percent still all the time what's going on. . . '[P7]*

Instruction was also viewed as transparent since it provides detailed information and participants understand the procedure.

> *'...the [instruction-based] with lots of details. Also, it's transparent. Although it's a bit annoying but it is transparent. . . '[P4]*

Some responses seem to indicate that the participants found the amount of information excessive. Additionally, some participants found *Automation* clearer since the interaction process was less complex and the "*machine takes care of everything*" [P7].

B. Predictability – Participants tended to consider *Instruction* as predictable since it is "*the most specific*" [P6] and it "*give[s] explanations*" [P1]. Participants claimed that they "*trust it most"* [P1] and described it as an expert guide:

> *'It's so smart, the machine acts as a teacher to teach you like, uh, what should you do? I don't have to worry about anything. He just tells you everything. . . ' [P5]*

The *Command* mechanism received high predictability ratings. Some participants suggested that the user in the video must have been familiar with the system already:

> '[Command based mechanism]... cause the user knows what he or she wants and maybe that's because he/she has already tried it before. Then for the [instruction based mechanism] because you have all the descriptions of the options the results would also be predictable' . [P7]

C. Reliability – Participants also tended to view *Instruction* as the most reliable for both simple and complicated tasks, since they gather the most information and it "*seems to know a lot*" [P1, P2].

> *'[...] if something goes wrong, it should be your fault instead of the machine. Because the machine let you know all the consequences before.'[P5]*

> *Interviewer: Why do you think that the instruction based mechanism is more reliable? . . . " "... I believe that the machine knows what it is doing because it has all the information about the laundry process that I don't"[P4]*

Satisfaction: Participants' satisfaction with the interaction depended on both the VUI mechanism and task complexity. The "*concise [..] and very quick"* [P3] *Automation* mechanism was the most satisfying for some, especially for routine tasks because "*you don't have to think about what you're doing*" [P6].

For complex, critical or high-stakes washing tasks such as "really expensive clothes" [P6], the "*detailed instructions*" [P3] of instruction-based mechanisms were considered more satisfying. Participants appreciated additional information when the perceived cost of error was high (*e.g.*, damaging expensive clothes).

> *'... it depends on what you're trying to wash like if I'm gonna wash really expensive clothes that I cannot mess up. Uh, then I would have done the third one because whatever I don't know what to do, it tells me exactly what's the stuff to take separate the clothes and all that stuff.'[P6]*

7 Discussion

This study demonstrates a difference in accountability between our designed mechanisms. The results of both qualitative and quantitative analysis supports our first hypothesis of a positive relationship between system automation and accountability. This relationship is well represented by P2's interview response that *"...the machine should take the most responsibility since it makes all the decisions."*. A similar trend has been reported by Sheridan *et al.*: *"... individuals using the [automated] system may feel that the machine is in complete control, disclaiming personal accountability for any error or performance degradation*" [57].

Though our results showed that *Automation* takes the highest accountability, two participants provided insights on other potential design factors that affect accountability. P5 argued that *Instruction* should be accountable when the system fails because users "*wasted"* their time listening to its verbose instructions. P6 indicated the importance of claims about the system: *"Actually that really depends on what the machine says it can do, you know if it says like 'I'm gonna be able to distinguish colour clothes from regular clothes and I'm not gonna mess up.' and it messes up then it's the machine's fault."* Setting realistic expectations about a system's abilities may help manage its accountability.

Our results echoed Suchman's prediction that automation can lead to shared accountability between humans and machines [64], and further that level of automation impact perceived accountability.

Users did not experience the washing machines firsthand, a limitation imposed by the state of technology. However, participants empathized with the common experience of doing laundry sufficiently to report a projected level of satisfaction with the interaction. They showed no difficulty in bridging the gap between experienced and imagined scenarios, making comments such as *"there is common ground between me and the machine"* [P3].

Our results suggest that task complexity does influence user satisfaction. Our qualitative analysis made it clear that for the high complexity laundry task, participants were more willing to ask for guidance and more likely to include the VUI agent in the decision-making process compared to the low-complexity task. However, they might prefer *Command* or *Automation* once they became comfortable with the system. Multiple users expressed the desire to transition to command-based systems once they had learned about the washing machine's hidden functionality through the instruction-based or recommendation-based mechanisms. This key finding motivates VUIs during naive access, which could be invaluable for secondary users of home appliances.

The qualitative results also support our quantitative analysis outcomes. Some participants stated that only experienced users who found the washing machine predictable would use the command-based mechanism. This may have contributed to an inflated predictability rating for the command-based mechanism. Though re-

sults suggest *Automation* is perceived as the most accountable, and *Command* the least, it is difficult to make a conclusive judgment on shared accountability for *Recommendation* and *Instruction*. Each of these mechanisms was scored differently in one Control Capability; however, the difference was not significant.

7.1 Implications for design

We encapsulate these findings in VUI design recommendations:

1. Accountability-aware design must consider context, specifically *task complexity*, *type of user(s)* and *ambiguity* of the interactions.
2. Automation has opposing effects on *accountability* and *user satisfaction*. A highly automated system may be satisfying to use, but in case of failure, users are more likely to find it blameworthy. Designers should consider this duality in the unique context of their product and its anticipated use.
3. User perceptions of *Reliability*, *Transparency* and *Predictability* depend both on available choices, and how those choices are presented, particularly for high complexity tasks. Designers should consider providing justification for system-presented choices, especially for high-stakes tasks.
4. Results suggest that detailed instruction- and recommendation-based mechanisms improve learnability, but could eventually be too repetitive. Designers should consider transitional mechanisms, in which system operation gradually provides less explanation and becomes more automated.

8 Conclusions and Future Work

We investigated the concept of accountability in home appliance VUIs. We examined automation level as a parameter that could impact accountability delegation, by designing and studying four mechanisms which varied automation and user involvement in decision-making, in simple and complex tasks.

Our primary study sought to characterize differences between these mechanisms. We found our use of video prototypes a successful basis for initial discussions on design concepts, providing non-obvious insights.

Qualitative and quantitative results support our first hypothesis of a positive relationship between automation and accountability, which held whether for both high and low complexity tasks.

Concerning our second hypothesis (that system automation increases user satisfaction), the quantitative result (N=15) was not statistically significant, but trended towards users preferring the most automated system. Interviews consistently supported H2 in that increased user involvement reduced satisfaction. This creates a dilemma for designers of automated systems, who must minimize users' cognitive load without saddling the system with complete accountability for errors.

We found participants more receptive to instructions and recommendations when they were concerned about the outcome of a process, or when they first used a system. We recommend that VUI designers implement guided interfaces as well as command-based ones. This gives users the freedom to transition from guided use to command-based use without leaving the system (and its designers) accountable for mistakes.

Future Work: From this foundation, we recommend next steps.

Sample Size – Our study size was appropriate for this early stage of investigation, revealing clear trends supporting the possibility of directing perceptions of accountability in users to support greater investment (more realistic study approaches) in this idea. However, increasing the size and diversity of even this exploratory approach might provide higher power of statistical tests and more significant quantitative insights.

Mechanism design – We examined four distinct mechanisms in isolation. As suggested in Section 7.1, we propose a mechanism that adjusts its automation as the user becomes more familiar with the device. The benefits of such a mechanism would need to be confirmed in a longitudinal study.

Metrics – User satisfaction is a volatile metric. In this study it is especially so because participants did not interact with a physical prototype. To minimize this limitation on user empathy, we assessed common moderate failure outcomes instead of complete failures (*i.e.*, a stained rather than a destroyed shirt). We will have more realistic results when users can reflect on their satisfaction level by observing the laundry process outcome on their own clothes.

Real results – A functional VUI system and, separately, a machine that truly enacts its instructions would advance the reality of the participants' experience and make their responses more reliable. A real system would succeed more often than fail, as opposed to our scenarios which aimed to make use of a short study session. When studied longitudinally within real homes in actual use, we can follow the development of trust and familiarity over time.

References

[1] 4 key measurements for customer satisfaction, Jan 2017.

[2] Demographics of mobile device ownership and adoption in the united states, Jun 2019.

[3] "accountability". In *Merriam-Webster.com*. 2020. Date Accessed: May 25, 2020.

[4] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3174156

[5] Amazon. Design process: The process of thinking through the design of a voice experience, 2020.

[6] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, and et al. Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300233

[7] T. Ammari, J. Kaye, J. Y. Tsai, and F. Bentley. Music, search, and iot: How people (really) use voice assistants. *ACM Trans. Comput.-Hum. Interact.*, 26(3), Apr. 2019. doi: 10.1145/3311956

[8] V. Bellotti and K. Edwards. Intelligibility and accountability: Human considerations in context-aware systems. *Hum.-Comput. Interact.*, 16(2):193–212, Dec. 2001. doi: 10.1207/S15327051HCI16234_05

[9] D. Bohn. Amazon says 100 million alexa devices have been sold - what's next?, Jan 2019.

[10] D. Boos, H. Guenter, G. Grote, and K. Kinder. Controllable accountabilities: the internet of things and its challenges for organisations. *Behaviour & Information Technology*, 32(5):449–467, 2013. doi: 10.1080/0144929X.2012.674157

[11] M. Braun, A. Mainz, R. Chadowitz, B. Pfleging, and F. Alt. 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*, CHI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300270

[12] G. Button and W. Sharrock. The organizational accountability of technological work. *Social Studies of Science*, 28(1):73–102, 1998. doi: 10.1177/030631298028001003

[13] C. Castelfranchi and R. Falcone. Towards a theory of delegation for agent-based systems. *Robotics and Autonomous Systems*, 24(3):141 – 157, 1998. Multi-Agent Rationality. doi: 10.1016/S0921-8890(98)00028-1

[14] C. M. Cheung and M. K. Lee. Understanding the sustainability of a virtual community: model development and empirical test. *Journal of Information Science*, 35(3):279–298, 2009.

[15] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, and et al. What

makes a good conversation? challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300705

[16] V. Clarke and V. Braun. Thematic analysis. *The Journal of Positive Psychology*, 12(3):297–298, 2017. doi: 10.1080/17439760.2016.1262613

[17] M. H. Cohen, M. H. Cohen, J. P. Giangola, and J. Balogh. *Voice user interface design*. Addison-Wesley Professional, 2004.

[18] M. L. Cummings. Automation and accountability in decision support system interface design. 2006.

[19] A. Datta, S. Sen, and Y. Zick. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *2016 IEEE Symposium on Security and Privacy (SP)*, pp. 598–617, May 2016. doi: 10.1109/SP.2016.42

[20] P. Dourish. *Accounting for System Behavior: Representation, Reflection, and Resourceful Action*, p. 145–170. MIT Press, Cambridge, MA, USA, 1997.

[21] P. Dourish. *Where the action is: the foundations of embodied interaction*. MIT press, 2004.

[22] C. Dwork and D. K. Mulligan. It's not privacy, and it's not fair. *Stan. L. Rev. Online*, 66:35, 2013.

[23] M. Giuliani, M. E. Foster, A. Isard, C. Matheson, J. Oberlander, and A. Knoll. Situated reference in a hybrid human-robot interaction system. In *Proceedings of the 6th International Natural Language Generation Conference (INLG 2010)*, 2010.

[24] B. Goodman and S. Flaxman. European union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, 38(3):50–57, Oct. 2017. doi: 10.1609/aimag.v38i3.2741

[25] Google. Conversation design guideline, 2020.

[26] R. A. Harris. *Voice interaction design: crafting the new conversational speech systems*. Elsevier, 2004.

[27] S. Hergeth, L. Lorenz, R. Vilimek, and J. F. Krems. Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human Factors*, 58(3):509–519, 2016. PMID: 26843570. doi: 10.1177/0018720815625744

[28] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, p. 241–250. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/358916.358995

[29] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, p. 159–166. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/302979.303030

[30] K. Höök. Steps to take before intelligent user interfaces become real. *Interacting with Computers*, 12(4):409 – 426, 2000. doi: 10.1016/S0953-5438(99)00006-5

[31] R. Kang, A. Guo, G. Laput, Y. Li, and X. hen. Minuet: Multimodal interaction with an internet of things. In *Symposium on Spatial User Interaction*, SUI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3357251.3357581

[32] B. Knijnenburg, P. Bahirat, Y. He, M. Willemsen, Q. Sun, and A. Kobsa. Iuiot: Intelligent user interfaces for iot. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, IUI '19, p. 139–140. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3308557.3313121

[33] J. Kowalski, A. Jaskulska, K. Skorupska, K. Abramczuk, C. Biele, W. Kopeundefined, and K. Marasek. Older adults and voice interaction: A pilot study with google home. 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, 2019. doi: 10.1145/3290607.3312973

[34] J. A. Kroll, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu. Accountable algorithms. *U. Pa. L. Rev.*, 165:633, 2016.

[35] T. Kulesza, M. Burnett, W.-K. Wong, and S. Stumpf. Principles of explanatory debugging to personalize interactive machine learning. In O. Brdiczka and P. Chau, eds., *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pp. 126–137. ACM, New York, USA, January 2015. doi: 10.1145/2678025.2701399

[36] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, and M. Bailey. Emerging threats in internet of things voice services. *IEEE Security Privacy*, 17(4):18–24, July 2019. doi: 10.1109/MSEC.2019.2910013

[37] J. La Delfa, M. A. Baytaş, H. Ngari, R. Patibanda, R. Khot, and F. Mueller. Drone chi: Somaesthetic human-drone interaction. 04 2020. doi: 10.1145/3313831.3376786

[38] L. Leahu, M. Cohn, and W. March. How categories come to matter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 3331–3334. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2466455

[39] J. R. Lewis. Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *Int. J. Hum.-Comput. Interact.*, 7(1):57–78, Jan. 1995. doi: 10.1080/10447319509526110

[40] B. Y. Lim and A. K. Dey. Assessing demand for intelligibility in context-aware applications. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, UbiComp '09, p. 195–204. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1620545.1620576

[41] B. Lubars and C. Tan. Ask not what ai can do, but what ai should do: Towards a framework of task delegability. In *Advances in Neural Information Processing Systems*, pp. 57–67, 2019.

[42] T. Malche and P. Maheshwary. Internet of things (iot) for building smart home system. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 65–70, Feb 2017. doi: 10.1109/I-SMAC.2017.8058258

[43] Y. Mehdi and Y. Mehdi. Windows 10: Powering the world with 1 billion monthly active devices, Mar 2020.

[44] D. R. Millen. Rapid ethnography: Time deepening strategies for hci field research. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '00, p. 280–286. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/347642.347763

[45] Y. Mittal, P. Toshniwal, S. Sharma, D. Singhal, R. Gupta, and V. K. Mittal. A voice-controlled multi-functional smart home automation system. In *2015 Annual IEEE India Conference (INDICON)*, pp. 1–6, Dec 2015. doi: 10.1109/INDICON.2015.7443538

[46] K. L. Mosier, L. J. Skitka, M. D. Burdick, and S. T. Heers. Automation bias, accountability, and verification behaviors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 40(4):204–208, 1996. doi: 10.1177/154193129604000413

[47] S. Nobarany, L. Oram, V. K. Rajendran, C.-H. Chen, J. McGrenere, and T. Munzner. The design space of opinion measurement interfaces: exploring recall support for rating and ranking. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2035–2044, 2012.

[48] D. A. Norman. How might people interact with agents. *Commun. ACM*, 37(7):68–71, July 1994. doi: 10.1145/176789.176796

[49] X. Page, P. Bahirat, M. Safi, B. Knijnenburg, and P. Wisniewski. The internet of what?: Understanding differences in perceptions and adoption for the internet of things. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2:1–22, 12 2018. doi: 10.1145/3287061

[50] R. Parasuraman and V. Riley. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2):230, 06 1997. Copyright - Copyright Human Factors and Ergonomics Society Jun 1997; Last updated - 2017-11-09; CODEN - HUFAA6.

[51] R. Parasuraman, T. B. Sheridan, and C. D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3):286–297, May 2000. doi: 10.1109/3468.844354

[52] M. Porcheron, J. E. Fischer, S. Reeves, and S. Sharples. 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, 2018. doi: 10.1145/3173574.3174214

[53] E. Rader, K. Cotter, and J. Cho. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI*

Conference on Human Factors in Computing Systems, CHI '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173677

[54] R. Riffenburgh. *Statistics in Medicine (3rd ed.)*. Academic Press, 2012.

[55] I. Robertson and P. Kortum. An investigation of different methodologies for rating product satisfaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):1259–1263, 2019. doi: 10.1177/1071181319631071

[56] J. A. Rode and E. S. Poole. Putting the gender back in digital housekeeping. In *Proceedings of the 4th Conference on Gender & IT*, pp. 79–90, 2018.

[57] T. Sheridan, T. VÃ¡mos, and S. Aida. Adapting automation to man, culture and society. *Automatica*, 19(6):605 – 612, 1983. doi: 10.1016/0005-1098(83)90024-9

[58] B. Shneiderman. Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences*, 113(48):13538–13540, 2016. doi: 10.1073/pnas.1618211113

[59] J. SILVA. Increasing perceived agency in human-ai interactions: Learnings from piloting a voice user interface with drivers on uber. *Ethnographic Praxis in Industry Conference Proceedings*, 2019(1):441–456, 2019. doi: 10.1111/1559-8918.2019.01299

[60] L. Skitka, K. Mosier, and M. BURDICK. Accountability and automation bias. *International Journal of Human-Computer Studies*, 52:701–717, 04 2000. doi: 10.1006/ijhc.1999.0349

[61] B. Stigall, J. Waycott, S. Baker, and K. Caine. Older adults' perception and use of voice user interfaces: A preliminary review of the computing literature. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, OZCHI'19, p. 423–427. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3369457.3369506

[62] N. Stout, A. Dennis, and T. Wells. The buck stops there: The impact of perceived accountability and control on the intention to delegate to software agents. *AIS Transactions on Human-Computer Interaction*, 6, 03 2014. doi: 10.17705/1thci.00058

[63] L. Suchman. *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press, 2007.

[64] L. Suchman and J. Weber. Human-machine autonomies. *Autonomous weapons systems: law, ethics, policy. Cambridge University Press, Cambridge*, pp. 75–102, 2016.

[65] P. E. Tetlock and R. Boettger. Accountability: A social magnifier of the dilution effect. *Journal of personality and social psychology*, 57(3):388, 1989.

[66] S. J. Tracy. Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10):837–851, 2010. doi: 10.1177/1077800410383121

[67] D. Wang, D. Lo, J. Bhimani, and K. Sugiura. Anycontrol – iot based home appliances monitoring and controlling. In *Proceedings of the 2015 IEEE 39th Annual Computer Software and Applications Conference - Volume 03*, COMPSAC '15, p. 487–492. IEEE Computer Society, USA, 2015. doi: 10.1109/COMPSAC.2015.259

[68] J. Zimmerman. Video sketches: Exploring pervasive computing interaction designs. *IEEE pervasive computing*, 4(4):91–94, 2005.