Computers in Human Behavior Reports 14 (2024) 100411

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Proactive behavior in voice assistants: A systematic review and conceptual model

Caterina Bérubé [a,i,1], Marcia Nißen [b,i,1], Rasita Vinay [b,c,i], Alexa Geiger [b,i], Tobias Budig [b,i], Aashish Bhandari [d], Catherine Rachel Pe Benito [e], Nathan Ibarcena [e], Olivia Pistolese [e], Pan Li [e], Abdullah Bin Sawad [f], Elgar Fleisch [a,i], Christoph Stettler [g], Bronwyn Hemsley [e], Shlomo Berkovsky [h], Tobias Kowatsch [a,b,c,i], A. Baki Kocaballi [e,*]

[a] ETH Zurich, Switzerland
[b] University of St.Gallen, Switzerland
[c] University of Zurich, Switzerland
[d] Delhi Technological University, India
[e] University of Technology Sydney, Australia
[f] King Abdulaziz University, Saudi Arabia
[g] Bern University Hospital, Switzerland
[h] Macquarie University, Australia
[i] Centre for Digital Health Interventions, Department of Management, Technology, and Economics at ETH Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Human-agent interaction
Proactivity
Voice assistants
User experience
Systematic review
conversational agents
proactive behavior
voice-based agents

ABSTRACT

Voice assistants (VAs) are increasingly integrated into everyday activities and tasks, raising novel challenges for users and researchers. One emergent research direction concerns proactive VAs, who can initiate interaction without direct user input, offering unique benefits including efficiency and natural interaction. Yet, there is a lack of review studies synthesizing the current knowledge on how proactive behavior has been implemented in VAs and under what conditions proactivity has been found more or less suitable. To this end, we conducted a systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist. We searched for articles in the ACM Digital Library, IEEExplore, and PubMed, and included primary research studies reporting user evaluations of proactive VAs, resulting in 21 studies included for analysis. First, to characterize proactive behavior in VAs we developed a novel conceptual model encompassing context, initiation, and action components: Activity/status emerged as the primary contextual element, direct initiation was more common than indirect initiation, and suggestions were the primary action observed. Second, proactive behavior in VAs was predominantly explored in domestic and in-vehicle contexts, with only safety-critical and emergency situations demonstrating clear benefits for proactivity, compared to mixed findings for other scenarios. The paper concludes with a summary of the prevailing knowledge gaps and potential research avenues.

1. Introduction

Voice assistants (VAs) have become an integral part of people's everyday lives (Clark et al., 2019). Improvements in natural language processing, advances in the semantic web, and the development of powerful processors (McTear, Callejas, & Griol, 2016) together with affordable prices of smart speaker devices running such VAs are the key drivers behind this increasing adoption. VAs fall within the broader category of conversational agents (CAs), which encompasses chatbots, embodied CAs, and VAs. Although there is no consensus either in the

* Corresponding author. School of Computer Science, Faculty of Engineering and IT University of Technology, Sydney 15 Broadway, Ultimo, NSW, 2007, Australia.
E-mail addresses: berube.caterina@gmail.com (C. Bérubé), marcia.nissen@unisg.ch (M. Nißen), rasita.vinay@ibme.uzh.ch (R. Vinay), alexa.geiger@student.unisg.ch (A. Geiger), tobias.budig@unisg.ch (T. Budig), aashish.bhandari@student.rmit.edu.au (A. Bhandari), cpebenito88@gmail.com (C.R. Pe Benito), nathan.j.ibarcena@student.uts.edu.au (N. Ibarcena), olivia.pistolese@student.uts.edu.au (O. Pistolese), pan.r.li@student.uts.edu.au (P. Li), asawad@kau.edu.sa (A.B. Sawad), efleisch@ethz.ch (E. Fleisch), christoph.stettler@insel.ch (C. Stettler), bronwyn.hemsley@uts.edu.au (B. Hemsley), shlomo.berkovsky@mq.edu.au (S. Berkovsky), tobias.kowatsch@uzh.ch (T. Kowatsch), baki.kocaballi@uts.edu.au (A.B. Kocaballi).
[1] Caterina Bérubé and Marcia Nißen contributed equally to this work and share first authorship.

https://doi.org/10.1016/j.chbr.2024.100411
Received 15 October 2023; Received in revised form 19 March 2024; Accepted 19 March 2024
Available online 20 March 2024
2451-9588/

definition or the classification of CAs, chatbots are commonly associated with text-based conversational systems, and embodied CAs correspond to CAs either text- or voice-based with a virtual avatar displayed on a screen. VAs can be defined as CAs with speech as their primary modality.

A recent marked trend within the domain of CAs – and more specifically VAs – is the shift toward proactive behavior. *Proactive* behavior in CAs refers to situations in which an agent initiates a dialog or interaction without requiring direct input from the users (Reicherts, Zargham, Bonfert, Rogers, & Malaka, 2021; Shum, He, & Di Li, 2018). This contrasts with *reactive* behavior in which an agent acts only in response to a user input or prompt. While in reactive scenarios CAs wait for users to initiate the interaction; in proactive scenarios, CAs look for the opportune moments to start the interaction with their users. Although the reactive interaction model works well in the majority of situations and is the dominant model of interaction, proactive behavior has been attracting increasing interest in human-computer interaction (HCI) and artificial intelligence (AI) research. Furthermore, commercial VAs such as Google Assistant and Amazon Alexa are increasingly opening their products to proactivity, where users can set up routines according to time or even detect sounds (e.g., water flowing, coughs, and a crying baby, cf. Bizzaco, 2022). Proactive behavior has been discussed as useful for purposes such as supporting user engagement (Smith, Sumner, Hedge, & Powell, 2020, 2021; Liao, Davis, Geyer, Muller, & Shami, 2016; Shum, He, & Di, 2018), independence (Greuter, Balandin, & Watson, 2019, pp. 429–435), organization (Smith, Sumner, Hedge, & Powell, 2021, 2023), personal health (Jovanovic, Baez, & Casati, 2021), education (Dyke, Adamson, Howley, & Penstein Rosé, 2012).

There are three main developments behind the recent interest in supporting proactive behavior in CAs.

First, the advancements in generative AI improve natural language processing and speech recognition accuracy (Alharbi et al., 2021; Gozalo-Brizuela & Garrido-Merchan, 2023; Radford et al., 2023) removing some of the barriers to adoption and enabling the optimal use of CAs across settings and scenarios, including in noisy or multi-user environments.

Second, there have been improvements in activity detection methods via sensor fusion or image recognition (Dey, 2009; Nweke, Teh, Mujtaba, & Al-Garadi, 2019) – a key component in proactive systems which typically requires the detection of contextual and activity information to initiate interaction with users.

Third, there are limitations with the current reactive model of interaction with CAs: there can be situations, especially safety-critical or emergency situations, where waiting for users' input is not needed, or delayed responses would be detrimental to users. Beyond safety-critical or emergency situations, proactive VAs can be suitable in situations where using hands is simply not practical, for instance, while cooking (Hwang, Oza, Callison-Burch, & Head, 2023, pp. 2233–2248), when looking at a screen is not possible, for instance, while driving (Meck, Draxler, & Vogt, 2023; Pakdamanian et al., 2022; T. Wu, Martelaro, Stent, Ortiz, & Ju, 2021), when users are visually (Abdolrahmani, Howes Gupta, Vader, Kuber, & Branham, 2021; Metatla, Oldfield, Ahmed, Vafeas, & Miglani, 2019; Thoo, Jeanneret Medina, Froehlich, Ruffieux, & Lalanne, 2023), or cognitively or intellectually impaired (Balasuriya, Sitbon, Bayor, Hoogstrate, & Brereton, 2018, pp. 102–112; Masina et al., 2020).

One recurring concern, though, concerns the potential intrusiveness of proactive behavior by a VA. Given that the VA takes the lead in initiating interactions, there's a risk of users perceiving these actions as interruptions, especially since voice assistants are usually not embodied but usually deprived of other visual or non-verbal cues or any kind of tangible "presence" and therefore, for instance, not as socially present as social robots. If executed with poor timing, proactive initiatives could easily be perceived as inappropriate or misaligned which could then have the potential to degrade the user-VA rapport and erode trust in the system (Kraus, Wagner, & Minker, 2020). We therefore argue that proactive behavior in voice-only interactions requires new research approaches and conceptual models.

Yet, despite the promising application areas and use cases for proactive VAs, a conceptual model of proactive behavior in VAs is still missing, and empirical studies that investigate the effects of proactivity in VAs on user attitudes and behavior are scarce and lack standards. This work seeks to provide a comprehensive overview of the current state-of-the-art by answering the following three research questions.

RQ1 How is proactivity implemented in VAs?
RQ2 How are proactive behaviors in Virtual Assistants evaluated and what key outcomes have been identified?
RQ3 What are the prevailing knowledge gaps and potential research avenues?

To answer these research questions, we will summarize related work and conduct a systematic review of empirical studies that have investigated the effects of proactive behaviors in VAs on user attitudes and behaviors in the following sections, develop a conceptual model of proactive behavior in VAs and discuss what we know and what we do not know about the suitability of proactive behavior in VAs.

2. Related work

2.1. On the unique nature of voice assistants: advantages and challenges of proactive behavior

The unique nature of voice-only interaction gives VAs specific advantages and drawbacks: VAs prove invaluable in settings that necessitate hands-free and minimal-distraction environments (Monteiro, Goncalves, Coelho, Melo, & Bessa, 2021), from disease management (F. Wu et al., 2023), over driving (Meck et al., 2023; Pakdamanian et al., 2022; T. Wu et al., 2021) to culinary activities (Hwang et al., 2023, pp. 2233–2248). By eliminating the need to use hands or devote one's full visual attention to a screen, VAs provide a more accessible mode of interaction. Moreover, they increase accessibility by assisting users with motor disabilities (Masina et al., 2020), speech impairments (Duffy, Synnott, McNaney, Brito Zambrano, & Kernohan, 2021; Smith et al., 2021, 2023), visual impairments (Abdolrahmani et al., 2021, pp. 1–16; Boyle & O'Brolcháin, 2023; Metatla et al., 2019; Pradhan, Mehta, & Findlater, 2018; Thoo et al., 2023), autism (Allen, Shane, & Schlosser, 2018; Greuter et al., 2019, pp. 429–435; Pradhan et al., 2018; Yu et al., 2018), and potentially for people with low literacy or intellectual and cognitive disabilities (Balasuriya et al., 2018, pp. 102–112).

Although many interactions that can be performed through graphical user interfaces (GUIs) can also be performed via voice, VAs also come with their unique set of challenges (Clark et al., 2019): Speech is a temporal medium to transfer information, so it is hard to present multiple options and long responses through speech. It demands more time from the users and basic operations that can be done easily via GUIs such as undoing or browsing different options are harder to perform with VAs. Therefore, VAs need to be designed with some extra attention to the interface to be as accessible and inclusive as possible, without overwhelming the user's auditory memory or invading their privacy, while being able to understand them regardless of their language, accent, or speech abilities (Duffy et al., 2021; Smith et al., 2021).

Proactivity in interactive systems can be defined as "the ability to autonomously initiate anticipatory action based on reasoning, meant to impact people and/or their environments" (Grosinger, 2022, p. 2). In VAs, the initiation can be based on either contextual information or prior direct user input. While the former can use information such as time, location, activity recognition, and system usage (Meurisch et al., 2020), the latter typically involves users explicitly setting preferences or allowing for implicit and automatic data collection about their behavior (Kraus, Fischbach, Jansen, & Minker). This implicit sensing function is intended to remain latent and requires the VA to continually gather data on its surroundings, such as detecting the state of a smart environment (e.g., lights, home appliances) or recognizing the user's activity (e.g., voice, movement).

Proactive behavior is intended to support users in performing some tasks such as appointments or medication reminders. However, such behavior requires careful design so that the machine-initiated conversation is useful, appropriate, and non-invasive. Based on Eric Horvitz's principles of mixed-initiative user interfaces (Horvitz, 1999, pp. 159–166), Yorke-Smith, Saadati, Myers, and Morley (2012) set out nine design principles for proactive CAs: valuable for the user; pertinent to the situation; competent with respect to the system's abilities and knowledge; unobtrusive; transparent; controllable; deferent to the user; anticipatory about the current and future needs and opportunities; and safe. Eventually, the degree of proactivity should be tailored to the *specific context and use case* (Meurisch, Ionescu, Schmidt, & Mühlhäuser, 2017), ranging from reactive responses (awaiting user prompts) to fully autonomous actions (independent of user input).

2.2. *Towards a preliminary conceptual model of proactive behavior in voice assistants: context-awareness as a necessity for proactivity in voice assistants*

Contextual information plays a pivotal role for proactive VAs, allowing them to accurately identify the appropriate moments to initiate interaction without being intrusive. Generally, context-awareness refers to the capability of a technology to collect data about the environment that is relevant to the user-application interaction and use it to adapt that same interaction (Dey, Abowd, & Salber, 2001). Fig. 1 illustrates a preliminary conceptual model of context-aware proactive behavior in voice assistants inspired by Dey et al.'s (2021) "Context Toolkit". According to this preliminary model, contextual information can be related to the *status* or *activity, time, location,* or *identity*, and it can be used to *store* or *present information* to the user, deliver *recommendations*, and/or automatically *execute* a function (Dey et al., 2001). Moreover, according to Dey et al. (2001), context-aware applications can use a combination of *context* elements (identity, location, status or activity, time) to trigger one or more *actions* (store, present, recommend, execute), either through a *synchronous* or an *asynchronous* service. A synchronous service triggers an action directly without waiting for the user to respond, while an asynchronous service informs the application that an action is ready to be delivered once the user is ready to receive it (Dey et al., 2001). This preliminary model was later used as an initial coding scheme for the data extraction during the systematic literature review (cf. section 3.3 and section 4.2) to analyze proactive behavior in VAs.

3. Methods

3.1. *Systematic literature review*

To collect a comprehensive and representative sample of studies investigating users' experiences with proactive behavior in VAs, we followed a structured systematic literature review protocol. A systematic literature review is an effective, robust, and rigorous method to answer our research questions, as it allows for a comprehensive and structured examination of existing knowledge. Drawing on the insights from relevant studies in the field and evaluating the current state-of-the-art in proactive behavior within VAs enables us to synthesize and analyze *how* proactivity is implemented in VAs (RQ1), to analyze *when* proactive behavior is beneficial or unsuitable (RQ2), and to address prevailing knowledge gaps and potential future research trajectories (RQ3). Overall, the systematic review protocol followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols 2015 (PRISMA) – a widely used checklist to facilitate the design of robust protocols for systematic reviews (Shamseer et al., 2015). The following sections explain the systematic review protocol in detail.

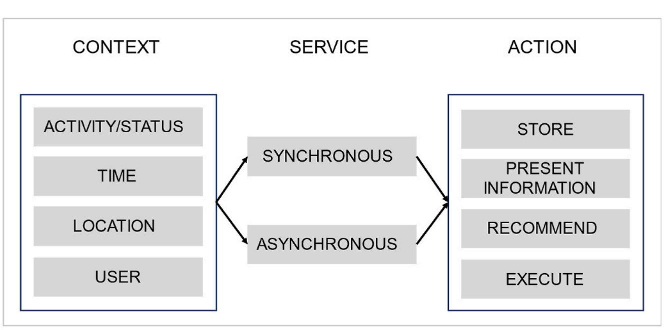

Fig. 1. Preliminary conceptual model for proactive behavior in VAs based on the conceptual model for context-aware computing by (Dey et al., 2001).

3.2. *Search strategy*

We employed search terms related to "technology" and "proactivity" (cf. Table 1). The terms were searched within titles and abstracts. The search was conducted in three electronic databases: ACM Digital Library, IEEE Xplore, and PubMed due to their recognized coverage of pertinent literature on conversational agents in computer science, human-computer interactions, and life science.

The first sampling round took place in May 2022 and included papers published until April 2022. The second sampling round took place in December 2023 as a "search update" and was limited to papers published between May 2022 and December 2023. In the second sampling round, we additionally used Google Scholar to conduct a Forward Reference Search (Vom Brocke et al., 2015) for all 12 papers that had been included in the first sampling round to identify further seminal articles that cited one of the classification works since they had been published. Please note that only during data extraction and analysis of the second sample, a thirteenth paper for the first sample was identified (Semmens, Martelaro, Kaveti, Stent, & Ju, 2019), which increased the final sample number of the first sample to n = 13.

3.3. *Selection*

3.3.1. *Criteria*

We included studies (1) that were primary research studies involving users experiencing or reflecting on a VA showing proactive behavior; (2) in which the agent was voice-based, that utilized speech input and

Table 1
Search terms.

Technology terms	Operator	Proactivity terms
"conversational agent" OR "conversational AI" OR "intelligent assistant" OR "relational agent" OR "intelligent agent" OR "virtual agent" OR "virtual assistant" OR "voice assistant" OR "voice agent" OR "speech-based" OR "voice-based" OR "voice-activated" OR "spoken-language" OR "text-to-speech" OR "chatbot" OR "chatterbot" OR "chatterbox" OR "socialbot" OR "digital assistant" OR "conversational UI" OR "conversational interface" OR "conversation system" OR "conversational system" OR "dialog system" OR "dialog system" OR "Siri" OR "Alexa" OR "Google Assistant" OR "Smart speaker" OR "Amazon Echo" OR "Apple Homepod"	AND	"proactive" OR "proactivity" OR "context-aware" OR "context-awareness" OR "context-sensitive" OR "driver-aware" OR "system-led" OR "mixed-initiative" OR "system-initiated"

output; (3) which reported qualitative and/or quantitative results on behavioral and/or attitudinal data from an interaction with a system, regardless of whether the system was a working prototype or based on the Wizard-of-Oz (WOZ) method (Dahlbäck, Jönsson, & Ahrenberg, 1993).

Papers were excluded if the study being reported (1) involved an interactive service via telephone (e.g., interactive voice response); (2) focused on testing machine learning algorithms only; or (3) examined a CA that was embodied (avatar or robot), not voice-based, or proactive.

We also excluded papers not written in English, workshop papers, dissertational theses, opinion papers or editorials, literature reviews, posters, and other presentations. In addition, we excluded papers not accessible to the authors after making all efforts to retrieve them.

3.3.2. *Processing*

During the first search round, all references were downloaded and inserted into a Google spreadsheet; and duplicates were removed. Two independent evaluators (i.e., CB and NI) conducted the screening for inclusion and exclusion criteria in two phases: first, papers were assessed based on their title and abstract, and, if included, they were assessed based on their full text. After both phases, we assessed inter-rater agreement by calculating Cohen's kappa (Cohen, 1960) and categorizing it as 0.00-0.20 (poor), 0.21-0.40 (fair), 0.41-0.60 (moderate), 0.61-0.80 (good), and 0.81-1.00 (very good) (Altman, 1990). The two evaluators consulted a third investigator (i.e., BK) in case of disagreements, which were then resolved by the majority.

The second search round followed the same procedure but all data was converted into an MS Excel spreadsheet. The initial title and abstract screening were conducted by MN, TB, and RV>, and the full-text screening by MN, AG, TB, and RV.

The results of the selection process are reported in a PRISMA figure (cf. Fig. 2, section 4.1) (Shamseer et al., 2015).

3.4. *Data extraction: Codebook and conceptual model development*

In the first sampling round, a total of eight researchers (i.e., CB, CPB, NI, OP, PL, AB, ABS, and BK) *first* extracted data from the eligible papers. Therefore, two researchers (i.e., CB and NI) reviewed all the entries for correctness and completeness. The following information was recorded in a Google spreadsheet: (1) study aims, (2) study methods, (3) VA behavior, (4) quantitative results, (5) qualitative results, and (6) key insights. Second, using the preliminary conceptual model for proactive behaviors in VAs introduced before (cf. section 2.2, Fig. 1), we classified *how* proactivity was implemented in the VA studies (cf. RQ1). This step required refining several elements of the preliminary conceptual model (Fig. 1) to the unique characteristics of VAs which was discussed by all authors. Namely, the "service" component was relabeled to "initiation", and its elements "synchronous" vs. "asynchronous" were rededicated to "direct" vs. "indirect". Further, five new elements within the "action" component were identified, namely "signal", "notification", "question", "suggestion", and "performance" instead of "store", "present information", "recommend", and "execute". The final Conceptual Model of Proactive Behavior for VAs and its components and elements are presented in detail in section 4.2, Fig. 3. In the second round, two researchers, MN and RV, followed the same procedure but converged all data (from the first and second sample) into an MS Excel spreadsheet. For each study, we captured how proactive behavior was represented by coding whether an element from the conceptual was represented or not (cf. Table A1 in the Appendix for the detailed analysis).

4. Results

4.1. *Screening results*

The first round encompassed a total of 478 unduplicated citations that were screened; 12 papers were eligible for inclusion and analysis. We included papers (yes/no) if they fulfilled all the inclusion criteria outlined in section 3.3.1 (i.e., (1) primary research study involving users' experiences with or reflections on (2) a proactive, voice-based conversational agent, and (3) quantitative or qualitative results were reported). If one of the criteria was not fulfilled, a paper was excluded. To ensure that the synthesis of the literature is conducted with a high

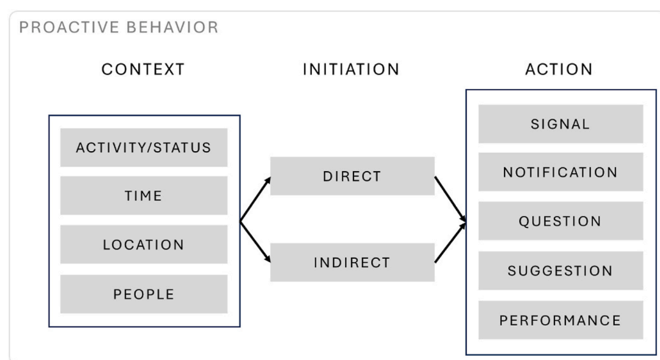

Fig. 3. The conceptual model of proactive behavior for VAs.

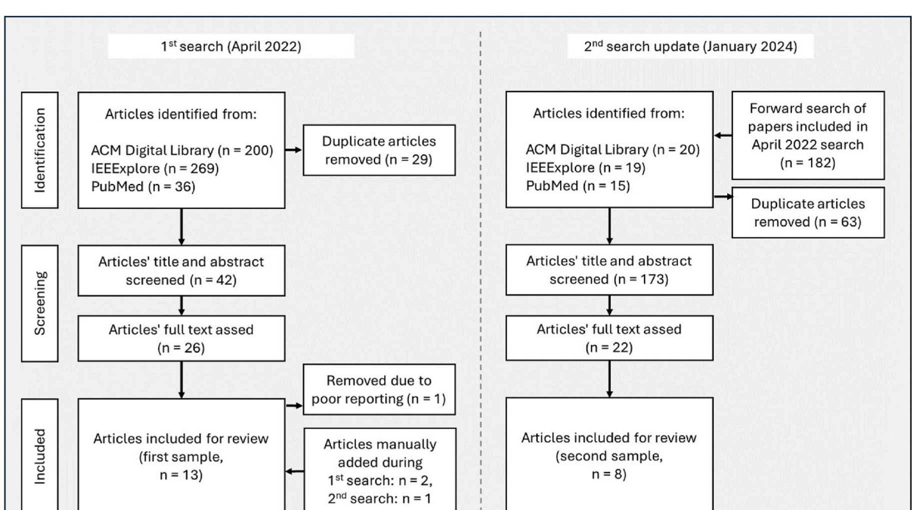

Fig. 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of included articles.

level of agreement among raters, we were eager to calculate Cohen's Kappa as an objective measure of reliability, consistency, and applicability of our inclusion and exclusion criteria, and thus, eventually, as an indicator of the robustness of our findings. Cohen's kappa for title and abstract screening initially amounted to $\kappa = 0.59$, indicating a moderate level of agreement. However, eventually, all disagreements were resolved by consensus before data extraction. One article (Schmidt, Bhandare, Prabhune, Minker, & Werner, 2020), was removed during the data extraction of included articles, due to insufficient reporting. Two papers published after the search date and that the authors knew about were manually added after the search process (Wei, Tag, Trippas, Dingler, & Kostakos, 2022; Zargham et al., 2022) – these two later also reappeared in the second search round and would have been included to the first sample latest by then. One paper was added manually to the first sample during the second search when comparing extracted data revealed a relevant thirteenth paper (Semmens et al., 2019), resulting in n = 13 studies eventually.

The second search resulted in 236 references, 54 from the search update and 182 from the forward search. 63 duplicates were removed, so 173 articles were screened for title and abstract, and 22 papers were screened based on the full-test. Cohen's kappa for title and abstract screening amounted to κ=0.91, indicating good levels of agreement. All disagreements were resolved by discussions each time before moving to the next step. Eventually, another 8 papers could be included for analysis from the second sample, leading to n = 21 studies in total. Fig. 2 summarizes the selection process results of both rounds in a PRISMA flow diagram.

4.2. A conceptual model for proactive behavior in VAs: how proactive behavior is implemented in VAs

To answer our first research questions (RQ1: How is proactivity implemented in VAs?) and to analyze proactive behavior in VAs in the included studies, a conceptual model was developed, that encapsulates three central components (i.e., context, initiation, action) each comprising multiple elements (cf. Fig. 3).

Context is multifaceted, spanning *activity/status, time, location,* and *people* dimensions (Dey et al., 2001). As emphasized in section 2.2, contextual information plays a pivotal role for proactive VAs, allowing them to accurately identify the appropriate moments to initiate interaction without being intrusive. *Activity/status* refers to the events that are occurring about an entity, such as talking, vocalizing, or heart rate for a person, a specific noise in a place, or an object being turned on or off. *Time* assigns chronological information to the context and is usually translated into a timestamp. *Location* encompasses both the position and relationships between entities. *People* identifies individuals present, including the main user. The context is monitored with rules, which, when satisfied, would then trigger an *initiation*.

Initiation was newly introduced and substituted Dey et al.'s (2001) "service" dimension, which we deemed not as applicable or relevant to characterize VAs. *Initiation* can be either *direct* or *indirect* (Yorke-Smith et al., 2012; Zargham et al., 2022). Direct initiation prompts action without user intervention such as a VA pushing an alert to pay attention during an autonomous ride (Pakdamanian et al., 2022) or initiating daily prompts in an experience sampling study (Ding et al., 2022, pp. 1–19). Direct initiation prompts are usually elicited based on predefined triggers, such as time-based reminders or alerts based on real-time information from environmental sensors. In contrast, indirect initiation always seeks user permission first through verbal (e.g., "I have something to tell you") or nonverbal cues (e.g., visual, auditory, or haptic signals).

Actions can manifest as *signals, notifications, questions, suggestions, or performances* (Kraus et al., 2020; Yorke-Smith et al., 2012). *Signals* mark events with basic cues like a chime. *Notifications* provide specific event details, like "Message from Ash. *Questions* involve the VA soliciting user input, often seen in ecological momentary assessments (Shiffman, Stone, & Hufford, 2008) and experience sampling methods (Cha et al., 2020; Conner, Tennen, Fleeson, & Barrett, 2009). *Suggestions* entail VA recommendations, for instance, "Please pay attention" (Pakdamanian et al., 2022), while *performance* denotes VA-initiated actions, like making a call or turning on lights (Mennicken et al., 2016; Völkel, Buschek, Eiband, Cowan, & Hussmann, 2021; Zargham et al., 2022). These actions can be used in tandem.

In the following sub-sections, we will answer our first research question by analyzing how proactivity was implemented in the VAs in all included studies with regards to this conceptual model, describing the domain, for which the respective VAs have been implemented, the motivation for their proactive behavior, context elements, initiation type, and actions.

To visualize the frequency of each context element, initiation type, or action across all studies, we also present a network diagram in Fig. 4. This diagram shows the contextual elements on the left and the action types on the right. The thickness of the connecting edges reflects the number of times a contextual element or an action type was used in direct or indirect initiations. The ovals in the center show the type of initiation and the number of times the selected studies included an initiation that was direct and/or indirect. A detailed breakdown of the implemented proactive behavior components and elements per study can be found in Table A1 in the Appendix.

4.2.1. Domains and motivation for proactive behavior

Table 2 yields an overview of the domains and motivations for proactivity across all studies.

The majority of the studies of the first sample (7 out of 13 papers) investigated proactivity in the domestic contexts (Wei, Dingler, & Kostakos, 2021a, 2021b; Cha et al., 2020; Mennicken et al., 2016; Völkel et al., 2021; Wei et al., 2022; Zargham et al., 2022), while three cases assessed the in-vehicle environment (Schmidt & Braunger, 2018; Semmens et al., 2019; T. Wu et al., 2021). The remainder unique cases investigated proactivity in social (Jarusriboonchai, Olsson, & Väänänen-Vainio-Mattila, 2014, pp. 98–106), entertainment (Szpektor et al., 2020), and financial (Guo, Guo, Yang, Wu, & Sun, 2021, pp. 1–11) domains.

Of the second sample (n = 8) two studies investigated proactive behavior in the domestic domain (Dubiel, Bongard-Blanchy, Leiva, & Sergeeva, 2023, pp. 1–6; Zargham, Reicherts, Avanesi, Rogers, & Malaka, 2023). Four research papers focused on in-vehicle application areas (Mathis, Werner, & Schmidt, 2023; Meck, 2023; Meck et al., 2023; Pakdamanian et al., 2022) while one paper each represented the entertainment (Marques, Abreu, & Santos, 2023) and the healthcare

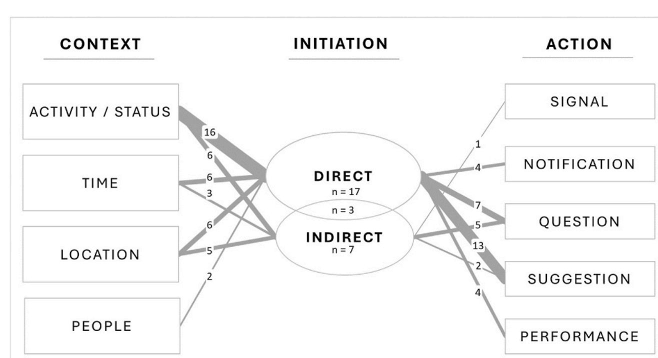

Fig. 4. A network diagram of context, initiation, and action across all studies (n = 21)
Notes. n refers to the unique number of studies implementing Direct or Indirect initiation. The numbers on the connecting edges show the number of times a contextual element or an action type was used. One study can use multiple contextual elements or action types. Therefore, the total numbers can be greater than the number of studies.

Table 2
Domains and motivation for proactivity.

	Total *(n = 21)*	First sample *(n = 13)*		Second sample *(n = 8)*	
	Count	Count	*References*	Count	*References*
Domain/Application Area					
Domestic	9	7	*(Cha et al., 2020; Mennicken et al., 2016; Völkel et al., 2021;* Wei et al., 2021a, 2021b, 2022; *Zargham et al., 2022)*	2	*(Dubiel et al., 2023,* pp. 1–6; *Zargham et al., 2023)*
In-vehicle	7	3	*(Schmidt & Braunger, 2018; Semmens et al., 2019; T. Wu et al., 2021)*	4	*(Mathis et al., 2023; Meck, 2023; Meck et al., 2023; Pakdamanian et al., 2022)*
Social	1	1	*Jarusriboonchai et al. (2014)*	-	-
Entertainment	2	1	*Szpektor et al. (2020)*	1	*Marques et al. (2023)*
Financial	1	1	*Guo et al. (2021)*	-	-
Healthcare	1	-	-	1	*Ding et al. (2022)*
Motivation for proactivity					
Monitoring and personalization	**11**	8	*(Cha et al., 2020; Mennicken et al., 2016; Schmidt & Braunger, 2018; Völkel et al., 2021;* Wei et al., 2021a, 2021b, 2022; *Zargham et al., 2022)*	3	*(Dubiel et al., 2023,* pp. 1–6; *Marques et al., 2023; Zargham et al., 2023)*
Social facilitation and interaction	**2**	1	*Jarusriboonchai et al. (2014)*	1	*Ding et al. (2022)*
Information and support	**5**	3	*(Semmens et al., 2019; Szpektor et al., 2020; T. Wu et al., 2021)*	2	*(Mathis et al., 2023; Meck et al., 2023)*
Safety and alert	**3**	1	*Guo et al. (2021)*	2	*(Meck, 2023; Pakdamanian et al., 2022)*

(Ding et al., 2022, pp. 1–19) domain. Social and financial domains were not represented in the second sample.

The combined total sample (n = 21) therefore included ten studies on the domestic domain, six on an in-vehicle environment, two on an entertainment area, as well as one paper each on a social, financial, or healthcare domain.

The motivation for proactive behavior in a VA in the included studies was not always clearly stated but similarly distributed across both samples. We identified the main motivations: monitoring and personalization, social facilitation and interaction, information and support, and safety and alert. Some cited its instrumental benefits for users, such as interfacing with a smart home (Mennicken et al., 2016), alerting of potential online-payment fraud (Guo et al., 2021, pp. 1–11), informing and soliciting (Semmens et al., 2019; T. Wu et al., 2021) or supporting the user while driving (Mathis et al., 2023; Meck, 2023; Meck et al., 2023; Pakdamanian et al., 2022; Schmidt & Braunger, 2018), fostering social interactions (Jarusriboonchai et al., 2014, pp. 98–106), enhancing topical discussions (Szpektor et al., 2020), or engaging older adults in a cognitive assessment (Ding et al., 2022, pp. 1–19). Other studies posited broader benefits, like enhancing monitoring (Wei et al., 2021a, 2021b, 2022) or aiding daily home tasks (Cha et al., 2020; Ding et al., 2022, pp. 1–19; Dubiel et al., 2023; Marques et al., 2023, 2023, pp. 1–6, 2023; Völkel et al., 2021; Zargham et al., 2022, 2023, 2023, 2023).

4.2.2. *Context*

Activity/status emerged as the dominant element, featured in 12 out of 13 studies in the first sample (Cha et al., 2020; Guo et al., 2021, pp. 1–11; Mennicken et al., 2016; Schmidt & Braunger, 2018; Semmens et al., 2019; Szpektor et al., 2020; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021; Zargham et al., 2022) and in 8 out of 8 studies in the second sample (Ding et al., 2022, pp. 1–19; Dubiel et al., 2023, pp. 1–6; Marques et al., 2023; Mathis et al., 2023; Meck, 2023; Meck et al., 2023; Pakdamanian et al., 2022; Zargham et al., 2023).

In the first sample, location (Jarusriboonchai et al., 2014, pp. 98–106; Mennicken et al., 2016; Schmidt & Braunger, 2018; Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021) and time (Cha et al., 2020; Mennicken et al., 2016; Schmidt & Braunger, 2018; Wei et al., 2021a, 2021b, 2022) followed, appearing in seven and six studies, respectively. In the second sample, though, location and time only appeared in one study each (location: Meck et al., 2023; time: Mathis et al., 2023).

Several studies combined activity/status with time and/or location (Cha et al., 2020; Mennicken et al., 2016; Schmidt & Braunger, 2018; Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021). In total, only two incorporated the “people" context: one in the first (Zargham et al., 2022), and one in the second sample (Marques et al., 2023).

Notably, of the studies presenting proactive behaviors in differing contexts (scenarios) to users (cf. section 4.3.1), in the first sample, Mennicken and colleagues (Mennicken et al., 2016), trialed six domestic scenarios in different rooms of a smart home, spanning morning to evening routines (cf. p. 124 for the complete list of scenarios). Schmidt and Braunger (Schmidt & Braunger, 2018) described to participants three driving scenarios to rate in terms of satisfaction: looking for a parking lot without success and the VA suggesting one; stepping out of the car in the morning at work and the VA reminding of a meeting; driving to the same location for the third time in a week and the VA suggesting storing the address (cf. p. 79 (Schmidt & Braunger, 2018)). Völkel and colleagues (Völkel et al., 2021) prompted participants to design the ideal VA for nine distinct daily scenarios where the user faces a problem. Finally, Zargham et al. (Zargham et al., 2022) had participants rank nine storyboards for usefulness, appropriateness, and intrusiveness. These ranged from a VA interrupting a historical discussion among friends to offer accurate information, to a VA detecting a fire, alerting the fire brigade, and notifying the household (cf. p. 4 for the complete list). In the second sample, two studies presented scenarios in online studies (Marques et al., 2023; Zargham et al., 2023): Marques et al. (2023) let users assess the relevance of proactive behavior in a set of Television usage scenarios (cf. p. 315 for an overview); Zargham et al. (2023) let participants assess their preference for humorous vs. non-humorous proactive behavior in seven different scenarios (cf. pp 298-299 for an overview). In contrast, Mathis et al. (2023) assessed users' attitudes and behaviors towards proactive behavior by an in-vehicle VA in various driving situations (cf. p. 419 for an overview).

4.2.3. *Initiation*

In both samples, direct initiation was predominant, appearing in 17 studies, 9 in the first sample (Guo et al., 2021, pp. 1–11; Mennicken et al., 2016; Schmidt & Braunger, 2018; Szpektor et al., 2020; Völkel et al., 2021; Wei et al., 2021a, 2021b, 2022; Zargham et al., 2022) and 8 in the second (Ding et al., 2022, pp. 1–19; Dubiel et al., 2023, pp. 1–6; Marques et al., 2023; Mathis et al., 2023; Meck, 2023; Meck et al., 2023; Pakdamanian et al., 2022; Zargham et al., 2023). However, in the first sample, indirect initiation was also common, featuring in seven studies (Cha et al., 2020; Jarusriboonchai et al., 2014, pp. 98–106; Semmens et al., 2019; Wei et al., 2021a; Wei et al., 2022; T. Wu et al., 2021). In the second sample, no study investigated indirect initiation only. Notably, two studies of the first sample (Wei et al., 2021a; Wei et al., 2022) and one study of the second sample (Dubiel et al., 2023, pp. 1–6) explored both forms of initiation by directly comparing them. In Wei et al. (2021b), Wei et al. (2021b voice application on a Google smart speaker posed questions about availability, mood, and current activities. The voice application was either starting the questions directly, after an earcon, or by first asking “Hey, are you available?” and asking the questions only after the participant would say “yes.” Dubiel et al. (2023, pp. 1–6) compared unsolicited (direct) vs. solicited (indirect) feedback by a VA on users' food choices in a food ordering scenario against a baseline with no feedback at all.

4.2.4. Action

In both samples, suggestions emerged as the primary action in six studies of the first sample (Cha et al., 2020; Guo et al., 2021, pp. 1–11; Mennicken et al., 2016; Schmidt & Braunger, 2018; Völkel et al., 2021; Zargham et al., 2022) and in all studies of the second sample. *Questions* follow closely in seven studies in the first sample (Jarusriboonchai et al., 2014, pp. 98–106; Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021) and four studies in the second sample (Ding et al., 2022, pp. 1–19; Mathis et al., 2023; Meck et al., 2023; Zargham et al., 2023). *Notifications* appeared in four studies in the first sample (Guo et al., 2021, pp. 1–11; Mennicken et al., 2016; Schmidt & Braunger, 2018; Zargham et al., 2022), but in no study of the second sample. *Performances* were presented in three studies of the first sample (Mennicken et al., 2016; Völkel et al., 2021; Zargham et al., 2022) and in one study of the second sample (Mathis et al., 2023). *Signals* were only present in the first sample (Jarusriboonchai et al., 2014, pp. 98–106).

Notably, while in the first sample, *questions* were often the sole action (in five studies) (Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021), in the second sample they were paired with suggestions and in half of all studies (Ding et al., 2022, pp. 1–19; Mathis et al., 2023; Meck et al., 2023; Zargham et al., 2023). Furthermore, in the first sample, *notifications* and *suggestions* frequently paired together in five studies (Guo et al., 2021, pp. 1–11; Mennicken et al., 2016; Schmidt & Braunger, 2018; Wei et al., 2022; Zargham et al., 2022) – a pattern not repeated in the second sample. *Performances*, besides being present in a few studies (n = 3), were always considered together with *notification, question,* and/or *suggestion* in both samples alike (Guo et al., 2021, pp. 1–11; Mathis et al., 2023; Mennicken et al., 2016; Zargham et al., 2022).

4.3. Analyzing study approaches to and user evaluations of proactive behavior in VAs

To answer our second research question (RQ2: “How are proactive behaviors in Virtual Assistants evaluated, and what key outcomes have been identified?“), we examined study designs, methods, measures, and outcomes used by the studies to assess the effects of proactivity on VA users. A comprehensive summary of all studies can be found in Table A2 in the Appendix.

4.3.1. Study designs

Comparing the study designs of the reviewed studies (cf. Table 3), they can be classified into three main types with regards to what they assessed: either user perceptions of a) proactive behavior (without comparisons), b) proactive vs. reactive behavior, or c) differently designed proactive behaviors against each other. User experience was either assessed 1) within one scenario or 2) across multiple scenarios/ situations.

Among the studies simply assessing proactive behavior (without comparison), in the first sample, one study measured user perceptions of proactive VAs within a domestic scenario (Wei et al., 2022) and one in a social scenario (Jarusriboonchai et al., 2014, pp. 98–106); four studies compared its suitability across multiple domestic scenarios (Cha et al., 2020; Wei et al., 2021b; Zargham et al., 2022) or across various situations during a drive (Semmens et al., 2019; T. Wu et al., 2021). In the second sample, there was only one study comparing proactive behavior across various situations during a drive (Mathis et al., 2023).

Among the studies comparing user perceptions of proactive vs. reactive behavior, in the first sample, two studies did so within an entertainment (Szpektor et al., 2020) or financial setting (Guo et al., 2021, pp. 1–11), and three studies compared user perceptions further across various domestic (Mennicken et al., 2016; Völkel et al., 2021) or in-vehicle scenarios (Schmidt & Braunger, 2018). In the second sample, only two studies compared proactive vs. reactive behaviors, one within a healthcare scenario (Ding et al., 2022, pp. 1–19), and one across multiple scenarios in an entertainment setting (Marques et al., 2023).

Eventually, differently designed proactive behaviors where only studied by one paper of the first sample within a domestic scenario (Wei et al., 2021a), Wei et al., 2021but by four studies of the second sample, thereof one within an e-commerce setting (Dubiel et al., 2023, pp. 1–6), two within a driving situation (Meck et al., 2023; Pakdamanian et al., 2022) and one comparing differently designed proactive behaviors (humorless vs. humorous) across multiple domestic scenarios (Zargham

Table 3
Study design comparison.

Study focus: Assessing user perceptions of …	Total (*n = 21)*	First sample *(n = 13)*		Second sample *(n = 8)*	
	Count	Count	*References*	Count	*References*
… proactive behavior	**8**	**7**		**1**	
Within one scenario	2	2	*(Jarusriboonchai et al., 2014*, pp. 98–106; *Wei et al., 2022)*	-	-
Across multiple scenarios	6	5	*(Cha et al., 2020; Semmens et al., 2019; Wei et al., 2021b; T. Wu et al., 2021; Zargham et al., 2022)*	1	*Mathis et al. (2023)*
… proactive vs. reactive behavior	**7**	**5**		**2**	
Within one scenario	3	2	*(Guo et al., 2021*, pp. 1–11; *Szpektor et al., 2020)*	1	*Ding et al. (2022)*
Across multiple scenarios	4	3	*(Mennicken et al., 2016; Schmidt & Braunger, 2018; Völkel et al., 2021)*	1	*Marques et al. (2023)*
… differently designed proactive behavior	**6**	**1**		**5**	
Within one scenario	4	1	*Wei et al. (2021a)*	3	*(Dubiel et al., 2023*, pp. 1–6; *Meck et al., 2023; Pakdamanian et al., 2022)*
Across multiple scenarios	2	-	-	2	*(Meck, 2023; Zargham et al., 2023)*

et al., 2023) and one across different driving-related applications (Meck, 2023).

4.3.2. Methods

In the first sample, six articles were based on field experiments (Cha et al., 2020; Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021), none in the second sample. Two articles in the first sample conducted a lab experiment (Jarusriboonchai et al., 2014, pp. 98–106; Mennicken et al., 2016), while four ran an online study (Guo et al., 2021, pp. 1–11; Schmidt & Braunger, 2018; Völkel et al., 2021; Zargham et al., 2022). In the second sample, five articles comprised lab experiments (Ding et al., 2022, pp. 1–19; Dubiel et al., 2023, pp. 1–6; Mathis et al., 2023; Meck et al., 2023; Pakdamanian et al., 2022), two articles online studies (Meck, 2023; Zargham et al., 2023) and one study was conducted in a focus group setting (Marques et al., 2023). In the first sample, one article included both a lab and an online evaluation (Szpektor et al., 2020).

Overall, we observe that in the first sample, all the field experiments involved an indirect initiation of proactive behavior (Cha et al., 2020; Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021), two of which included both direct and indirect initiation (Wei et al., 2021a; Wei et al., 2022), while all online studies involved a direct initiation (Guo et al., 2021, pp. 1–11; Schmidt & Braunger, 2018; Völkel et al., 2021; Zargham et al., 2022). In the second sample, there was no field study, but one lab study that compared indirect (here called: "solicited") to direct (here called: "unsolicited") initiation (Dubiel et al., 2023, pp. 1–6), all other studies investigated direct initiation.

Regarding the implementation methods for testing proactive VAs, in the first sample, one study used a Wizard of Oz (WOZ) approach (Jarusriboonchai et al., 2014, pp. 98–106), three implemented working prototypes (Cha et al., 2020; Guo et al., 2021, pp. 1–11; Szpektor et al., 2020), and six used mock-ups (Mennicken et al., 2016; Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021). In the second sample, three studies used a Wizard-of-Oz approach (Dubiel et al., 2023, pp. 1–6; Mathis et al., 2023; Meck, 2023), two of them in vehicle mock-ups (Mathis et al., 2023; Meck, 2023). One study simulated the interaction with a working VA prototype in a vehicle mock-up (Pakdamanian et al., 2022). One study employed a working prototype of a VA for cognitive assessments (Ding et al., 2022, pp. 1–19).

Further worth mentioning, three studies of the first sample presented scenarios exclusively (Schmidt & Braunger, 2018; Völkel et al., 2021; Zargham et al., 2022), which depicted imaginary situations, with one study having participants enact scenarios using a mock-up (Mennicken et al., 2016). Of the second sample, two studies only presented scenarios depicting imaginary situations in online studies (Marques et al., 2023; Zargham et al., 2023), one Wizard-of-Oz study with a vehicle mock-up tested proactive behavior in various driving situations (e.g., speed changes, heavy traffic, lane change, etc.; cf. Mathis et al., 2023, p. 4, for an overview).

Therefore, overall, the majority of all studies (first sample: n = 10 of 13, second sample: n = 5 of 8) provided participants with some experiential context, whether through WOZ, mock-ups, or prototypes (Cha et al., 2020; Ding et al., 2022, pp. 1–19; Dubiel et al., 2023, pp. 1–6; Guo et al., 2021, pp. 1–11; Jarusriboonchai et al., 2014, pp. 98–106; Meck et al., 2023; Meck, 2023; Mennicken et al., 2016; Pakdamanian et al., 2022; Semmens et al., 2019; Szpektor et al., 2020; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021). In general, studies offering concrete experiences presented a specific proactive VA for evaluating participants' behaviors and responses, whereas scenario-based setups took a more exploratory approach, gauging participants' hypothetical attitudes.

4.3.3. Measures

To gauge the impact of proactive behavior in VAs, we analyzed outcomes related to its influence on user behavior and attitudes, including aspects like preference, satisfaction, and trust. Overall, there's a notable lack of standardization in evaluation methods.

In the first sample, seven studies focused solely on behavior (Cha et al., 2020; Jarusriboonchai et al., 2014, pp. 98–106; Semmens et al., 2019; Völkel et al., 2021; Wei et al., 2021b; Wei et al., 2022; T. Wu et al., 2021), three only on attitude (Mennicken et al., 2016; Schmidt & Braunger, 2018; Zargham et al., 2022), and three on both (Guo et al., 2021, pp. 1–11; Szpektor et al., 2020; Wei et al., 2021a). In the second sample, only one study focused solely on behavior (Dubiel et al., 2023, pp. 1–6), five studies solely on attitude (Ding et al., 2022, pp. 1–19; Marques et al., 2023; Mathis et al., 2023; Meck, 2023; Meck et al., 2023), and two on both (Pakdamanian et al., 2022; Zargham et al., 2023).

In the first sample, among studies assessing attitudes (Guo et al., 2021, pp. 1–11; Mennicken et al., 2016; Schmidt & Braunger, 2018; Szpektor et al., 2020; Wei et al., 2021a) half used Likert scales (n = 3) (Guo et al., 2021, pp. 1–11; Schmidt & Braunger, 2018; Wei et al., 2021a), two relied on unspecified ratings (Mennicken et al., 2016; Szpektor et al., 2020) and one employed ranking (Zargham et al., 2022). In the second sample, among studies assessing attitudes, seven studies employed Likert scales (Ding et al., 2022, pp. 1–19; Dubiel et al., 2023, pp. 1–6; Mathis et al., 2023; Meck, 2023; Meck et al., 2023; Pakdamanian et al., 2022; Zargham et al., 2023), one employed interviews (Marques et al., 2023).

For behavioral investigations, methods in the first sample varied between behavior annotation (Cha et al., 2020; Jarusriboonchai et al., 2014, pp. 98–106; Semmens et al., 2019; Wei et al., 2021a; T. Wu et al., 2021) and interaction logs (Guo et al., 2021, pp. 1–11; Szpektor et al., 2020; Völkel et al., 2021; Wei et al., 2021a, 2021b, 2022). In the second sample, only interaction logs were present (Marques et al., 2023; Pakdamanian et al., 2022).

4.3.4. Outcomes

To increase comparability of the findings and to answer under what conditions proactive behavior in VAs is beneficial or unsuitable (RQ2), we will report outcomes grouped by the underlying study design as derived in section 4.3.1, this is with regard to what they assessed, either user perceptions of a) proactive behavior (without comparisons), b) proactive vs. reactive behavior, or c) differently designed proactive behaviors against each other. We will further report the results by whether user experiences were assessed 1) within a single scenario or 2) across multiple scenarios.

4.3.4.1. User perceptions of proactive behavior. Among the studies simply assessing user reactions to proactive behavior within a set scenario, Jarusiboonchai et al. (2014) discerned that when participants were guided by a proactive VA to communicate with another user in a social setting, they exhibited a tendency to interact more extensively with the VA rather than with the human interlocutor. Wei et al. (2022) found that users adopted four distinct strategies to prevent or resolve errors in their interactions with proactive VAs in an experience sampling scenario: (i) raising voice and approaching, (ii) repeating, (iii) phonetic and lexical changes, and (iv) getting help from others.

When presented with multiple hypothetical domestic scenarios for proactive behavior, participants in a study by Zargham et al. (2022) ranked emergencies as the most valuable, fitting, and least intrusive; proactive VA interventions during a social game as the least beneficial and apt, and a VA stepping in to mediate a disagreement as the most intrusive (Zargham et al., 2022). Another study by et Wei et al. (2021b) comparing proactive behavior across various domestic scenarios found that user interruptibility was heightened during activities tied to entertainment and work/study, particularly when ambient light and noise levels were subdued (Wei et al., 2021b). Contrasting this evidence, Cha et al. (2020) elucidate that across dormitory settings, students engaged in chores, relaxation periods, digital entertainment (e.g., internet or smartphone usage, watching videos), and during transitions between activities, appeared to be more amenable to interruptions than when engaged in work or study.

Two studies assessing user perceptions of proactive behavior across various driving situations, by Semmens et al. (2019) found that users were more prone to be interrupted, a) when driving on the correct driving route (81.3% YES) compared to off-road route (66.9% YES), b) when the steering wheel is straight, c) when the car is moving at a constant speed, d) when the car is completely stopped, e) when the car was accelerating (vs. braking), f) when the participant was releasing the brake versus applying the brake. T. Wu et al. (2021) re-analyzed the dataset by Semmens et al. (2019), by letting annotators assess the safety of drivers' likeliness to decline interactions with an in-vehicle proactive VA, and ascertained that drivers frequently declined interactions with in-vehicle proactive VAs, even in traffic scenarios deemed sufficiently safe for such engagements. Another study by Mathis et al. (2023) found that during an automated drive most driving situations and the performance of various non-driving related activities did not impact users' perceived opportuneness of a proactive interaction. In contrast to the results reported by Zargham et al. (2022), only an extreme traffic situation with an approaching emergency vehicle was considered inopportune in the study by Mathis et al. (Mathis et al., 2023). However, it should be noted, that in Zargham et al. (2022), proactive behaviors were based on imaginary proactive behaviors related to the emergency, whereas Mathis et al. (2023) studied context-unaware, unrelated proactive behaviors. Based on a qualitative analysis of interview data, the authors (Mathis et al., 2023) further contest that travel time and the current state of the user need to be considered for the selection of an opportune moment.

4.3.4.2. User perceptions of proactive vs. reactive behaviors. Comparing user perceptions of proactive vs. reactive VAs, two studies of the first sample demonstrated a more favorable disposition towards the proactive VA in an entertainment context with the VA engaging in a conversation about the NBA basketball league (Szpektor et al., 2020) and in a financial context, where the proactive version alerted customers when a suspected online-payment fraud was detected (Guo et al., 2021, pp. 1–11). Similarly, a study among older adults by Ding et al. (2022, pp. 1–19) found more favorable assessments of proactive (compared to reactive) behaviors in a VA-based screening and diagnosis tool for neurocognitive disorders (e.g., dementia).

In one of the studies where proactive and reactive behaviors were compared against each other across various domestic scenarios, users were prompted to conceptualize interactions with an ideal VA across nine distinct scenarios (Völkel et al., 2021): Surprisingly, across all scenarios, only a mere 8.1% envisioned interactions with a proactive VA, most users (91.9%) designed dialogues that were initiated by the user. Notable exceptions were scenarios where the VA detected that the user was listening to loud music even though their neighbors were sensitive to noise (the volume scenario), that the lights were still on, while the user tended to fall asleep while reading (the "IoT" scenario) and that the alert was not yet set, despite an important meeting early next morning" (the "alert" scenario). In these three cases, the VA was designed to initiate the dialogues by 27.3% of all users in the "volume" scenario, 12.7% in the alert scenario respectively, and 10.7% in the "IoT" scenario. Similarly, a study by Mennicken et al. (2016), contrasting a highly proactive VA against a less proactive counterpart, found a marked preference among users for the latter across various domestic scenarios. This inclination was primarily attributed to concerns over privacy. The overtly proactive VA was often perceived as "patronizing", "crossing social boundaries", and as a "stranger living in their home". The only study from the second sample investigating user perceptions of reactive vs. proactive behavior across various domestic scenarios by Marques et al. (2023) found mixed results, where proactive behaviors were preferred in three out of six scenarios (i.e., continue watching a missed content, continue watching an interrupted content, or start watching a related content) and reactive behaviors in the other three (i.e., continue watching a YouTube content, recommendation of content for co-viewing, recommendation of a TV maintenance service). In contrast to users' reservations towards proactive VAs in the above domestic examples, a study by Schmidt et al. (2018) overall observed a preference for proactive behaviors across three different in-vehicle scenarios (i.e., searching for a parking lot, reminding of an appointment, saving a frequently searched address as a new entry in the address book).

4.3.4.3. User perceptions of differently designed proactive behaviors. The only study from the first sample investigating differently designed proactive behaviors within a single (domestic) scenario by Wei et al. (2021a) consists of a comparative study of initiation styles in user interactions (Wei et al., 2021a). Their findings revealed that non-verbal cues, specifically earcons, more frequently elicited behavioral responses than their verbal counterpart, exemplified by the prompt, "Hey, are you available?". However, it is noteworthy that the same study (Wei et al., 2021a) highlighted a user preference for the verbal cue. In the second sample, one study in an e-commerce setting by Dubiel et al. (2023, pp. 1–6) discovered that unsolicited (i.e., direct initiation) feedback by a VA on users' food choices in a food ordering scenario was perceived as more appropriate than solicited (i.e., indirect initiation) feedback. Two studies compared differently designed proactive behaviors within a driving-related scenario. Here, Meck et al. (2023) found differences in users perceptions of a proactive VA's intelligence and positiveness based on the design of linguistic markers in the proactive prompts (e.g., sentence length, position of sub-clauses, form of address, politeness, etc.) in a lab study with a Wizard-of-Oz based-VA implemented in a vehicle mockup. Pakdamian et al. (2022) compared perceptions of speech-based always warnings by a proactive VA (i.e., context-independent) against context-dependent choices of different advisory warning modalities such as text, vibrotactile, speech, visual for different non-driving related task types (i.e., gaming, talking, reading) in an automated driving situation. They found that a context-dependent choice of modality had "statistically significant effects on safer takeover behavior, improved driver situational awareness, less attention demand, and more positive user feedback, compared with uniformly distributed speech-based warnings across all [non-driving related tasks]" (p. 75).

Two studies from the second sample further assessed differently designed proactive behaviors across different scenarios: Zargham et al.'s (2023) study demonstrated that the impact of humor on the desirability of a proactive statement is contingent on participants' perceptions of voice assistants in general, their subjective judgment of the humor, and rather independent of the scenario. A study of linguistic markers in proactive prompts by a VA by Meck (2023) with in-vehicle use cases discovered that users' preferences for the position of sub-clauses differed significantly between comfort-oriented (i.e., relaxing mode) and functional domain (i.e., information on the remaining fuel range) prompts.

5. Discussion

5.1. Theoretical contribution

This work makes some substantial theoretical contributions to the field of proactive behavior in VAs.

Firstly, the newly developed conceptual model allows for a comprehensive characterization of proactive behavior in VAs, providing scholars with a new lens for studying proactive interactions with VAs, and offering a framework for further investigation into this emerging area of research. Our conceptual model encompasses three key components: context, initiation, and action. 'Context' pertains to information related to activity, time, location, and individuals involved. 'Initiation' denotes how a VA begins a conversation. It can either be direct, initiating a conversation without the user's explicit consent, or indirect, requiring user permission beforehand. 'Action' captures the range of tasks a VA might undertake proactively, with our model delineating five distinct action types: signal, notification, question, suggestion, and performance. We particularly contend that our model lays a robust groundwork for a systematic exploration of the interplay between various contextual factors, initiation styles, and action types in shaping the proactive behavior of VAs, thus allowing us to answer our RQ1 ("How is proactivity implemented in VAs?"):

Among the studies reviewed, 'Activity' emerged as the most prevalent contextual factor, followed by 'time' and 'location'. The presence of omther individuals was only considered in one study (Zargham et al., 2022). Direct initiation saw more frequent use than its indirect counterpart. While direct initiation is often apt for emergencies, bypassing the need for user consent, indirect initiation tends to be more appropriate in various other scenarios. Over time, a proactive VA could harness historical data from indirect initiations to discern suitable moments for direct engagement. As for action types, there was an even distribution across categories, apart from 'signal', which was adopted just once (Jarusriboonchai et al., 2014, pp. 98–106). Signals, characterized by their simplistic, minimalist output, are most effective when users can easily discern their meaning. In the absence of context or clear cues, signals might engender confusion. Thus, adequate context is pivotal. For instance, a singular beep could function in two different ways in our model: as an indirect initiation prompt or a direct standalone signal. To ensure users decode a signal correctly, prior contextual information is essential. One study (Szpektor et al., 2020) highlighted a distinctive form of proactivity. Rather than initiating a conversation, the VA proactively augmented a user-initiated conversation with pertinent, yet unsolicited, information. While our model predominantly addresses the inception of a conversation and does not specifically explore the dynamics of guiding an ongoing conversation, this behavior can arguably align with a 'direct notification' within our framework. Here, the proactive VA delivers information that directly resonates with the topic at hand. Such an approach can be viewed as a nuanced, in-conversation form of proactivity, particularly relevant to advanced dialog management.

Secondly, the systematic review of user studies investigating the effects of proactive behaviors in VAs on user attitudes and behaviors enhances our understanding of users' current mental models regarding proactive VAs. However, based on the available evidence, our second research question (RQ2: "How are proactive behaviors in Virtual Assistants evaluated, and what key outcomes have been identified?") can only be answered partially: Prior work suggests that the degree of proactivity should be tailored to the *specific context and use case* (Meurisch et al., 2017), ranging from reactive responses (awaiting user prompts) to fully autonomous actions (independent of user input). To this end, our review partially confirms that contextual factors overall seem to sway the perceptions of appropriateness, invasiveness, and usefulness associated with VA's proactive behaviors. Especially situations pertaining to critical safety and emergencies emerged as the most appropriate, least intrusive, and most valuable in user evaluations if the proactive behavior was related to the emergency (Guo et al., 2021, pp. 1–11; Zargham et al., 2022) and least suitable if the proactive prompt was unrelated to the emergency-situation (Mathis et al., 2023). Other than that, the heterogeneity of the sample hardly allows us to derive generalizable answers regarding beneficial or unsuitable conditions.

Examining users' mental models of proactive VAs, findings overall indicate an absence of proactive behavior in their perception and conceptualization of VAs (Völkel et al., 2021). When users were introduced to hypothetical scenarios with proactive VAs, there was a discernible openness to their use (Schmidt & Braunger, 2018). However, this acceptance was accompanied by concerns, particularly around privacy—such as the potential for interruptions during personal interactions at home (Zargham et al., 2022). These apprehensions were mirrored in a laboratory mock-up study (Mennicken et al., 2016) and an in-vehicle field study (Semmens et al., 2019; T. Wu et al., 2021), which found that users could become overwhelmed by the VA's proactive tendencies. From a social context standpoint, proactive VAs do not seem to be considered active participants in human-to-human dialogues. Instead, they are viewed predominantly as tools for individual tasks. For example, a study (Jarusriboonchai et al., 2014, pp. 98–106) noted that while users would engage with the proactive VA when prompted, they often did not incorporate it into everyday conversations. Furthermore, many users were resistant to the idea of a proactive VA intervening in active disputes to offer potential solutions (Zargham et al., 2022). Given these nuances, it's imperative to finely tune the level of intrusiveness in proactive VAs and ensure they adapt to individual user preferences, such as intervening appropriately within conversations (Strauss & Minker, 2010). As proactivity remains a relatively novel feature with an inherent intrusive quality, users might benefit from a familiarization phase. This notion is further articulated by Mennicken et al. (2016) as follows:

> *"[…] we could think about the […] agent as a flatmate who moves in and is instantly given access to all the functions in the home, the data we own, and that passively tracks our conversations all day long. For an actual flatmate whom we do not know well yet, we would give the trust relationship time to develop before opening up and giving access to all of this. Designing for a phase with limited proactive behavior and a dialogue to "get to know each other" could be beneficial for inhabitants to feel more comfortable about the autonomy and proactivity that [the] agent might exhibit later on."* (Mennicken et al., 2016, p. 129)

Therefore, after an onboarding phase in which a proactive VA and a user get to know each other, it may be a useful strategy to design the proactive behavior in a dynamic way where the proactive VA learns and adapts to the users' preferences and changing needs over time. Studies (Mennicken et al., 2016; Zargham et al., 2022) propose that to optimize the user-friendliness of human-VA interactions, a gradual introduction to proactive behavior is essential. Over time, the VA should adapt its behavior to align with user preferences. A compelling illustration of this concept is a short film by Superflux and Mozilla, where a voice assistant continuously gathers user data and tailors its behavior to complement their daily routines.[2]

5.2. Design implications

The present research offers actionable guidelines and a salient framework that can guide practitioners from the first day in designing, developing, and implementing proactive behavior in VAs.

First, we offer an explicit representation of proactive behavior as a determining factor for a VA's design. Having a common understanding and definition of proactive behavior in VAs and being aware of the distinct nature of VAs and the particularities of proactive behavior in VAs might help to lay a solid foundation for streamlining the design process of proactive VAs, which in turn reduces designers' efforts, cost, and time to develop and implement new VA-based services and prevent communication problems within companies between product managers and developers.

Second, while the design components detailed in the conceptual model are descriptive, the conceptual model gives designers the flexibility to add and combine these components to prototype and tailor the VA development quickly to any desired target group or use case while taking into account boundary conditions and restrictions (e.g., available budget or development expertise).

5.3. Limitations and future research

Despite its contributions, this study has several limitations that

[2] Our Friends Electric | A Short Film by Superflux and Mozilla (2017): https://www.youtube.com/watch?v=PsjunTAH-2A.

warrant consideration and that give rise to many open research directions (RDs). This last section, thus, eventually also allows us to answer our last research question, "What are the prevailing knowledge gaps and potential research avenues? (RQ3).

First, the conceptual model itself presents certain limitations. Its categorization of initiation and action may not fully capture the entire range of behaviors in voice assistants, and its approach to context, though grounded in context-aware computing (cf. section 2.2) and covering the major types of contextual information, might overlook some less common context elements, such as environmental temperature, utilized by proactive voice assistants in diverse settings. Future research could explore additional contextual dimensions beyond those covered in the current model to enhance the model's applicability across diverse domains, settings, and use cases (RD1). The key will be to consider the relevant contextual elements needed to initiate and perform proactive actions effectively in voice assistants.

Second, the limited number of studies and their heterogeneity across various factors, including application areas, study designs, reported outcomes, etc., currently make it implausible to draw generalizable conclusions. Only a greater base of primary research studies will allow future researchers to assess statistical effects or effect sizes in meta-analysis (RD2). While we are convinced that the focused scope of our systematic review provides future researchers with a realistic portrayal of the current landscape of user studies on proactive behavior in VAs, it simultaneously emphasizes the pressing need for further research in this area. We purposefully applied focused inclusion criteria to account for the unique nature of VAs and the likely different perception of proactive behavior in VAs compared to other types of CAs. Future research from various disciplines should and could combine their knowledge in comprehensive conceptual frameworks and design guidelines, that allow to compare and synthesize perceptions of proactive behavior in all types of CAs, from text-based chatbots to physically embodied social robots (RD3). For instance, early Media Equation experiments studied user perceptions of "dominant" vs. "submissive" behaviors of computers, which could be re-evaluated under the realm of the reactivity-proactivity spectrum. Another approach might be to consider other domains of proactive behaviors studied, for instance, in Marketing, Human-Computer-Interaction, and Information Systems Research, where one stream of research investigates the timing, intrusiveness, perception, and reception of push notifications delivered to users' smartphones and tablets (RD4) (Mehrotra, Musolesi, Hendley, & Pejovic, 2015; Mehrotra, Pejovic, Vermeulen, Hendley, & Musolesi, 2016). Considering, that many studies in our review were conducted with in-vehicle VAs, investigating the perfect reactivity-proactivity ratio across the different levels of autonomy in automatic driving might also be fruitful (RD5). This seems to be further highlighted by the slight shift towards studying proactive behavior in in-vehicle VAs (4 out of 8 studies of the second sample, vs. 3 out of 13 in the first sample, cf. Table 2). Eventually, the still limited number of studies stands in stark contrast to the increasing interest in VAs by various industries. If the limited number and heterogeneity of studies in our review are a reflection of the current state of research in the domain of proactive VA, this would be alarming, indicating a widening gap between research and practice. To capture industry experts' experience with proactive VAs, applying a different methodological approach to our research questions encompassing interview studies, focus groups, workshops, or Delphi studies could help to bridge this gap (RD6).

Third, of all the studies reviewed, only five employed operational or working prototypes of proactive VAs (Cha et al., 2020; Ding et al., 2022, pp. 1–19; Guo et al., 2021, pp. 1–11; Pakdamanian et al., 2022; Szpektor et al., 2020). The rest either investigated the suitability of proactive VAs through imagined experiences (Völkel et al., 2021; Zargham et al., 2022) or used simulated setups and systems such as WOZ (Jarusriboonchai et al., 2014, pp. 98–106) or mock-ups (Mennicken et al., 2016; Semmens et al., 2019; Wei et al., 2021a, Wei et al., 2022Wei et al., 2022; T. Wu et al., 2021). Although imaginary and simulated experiences yield valuable insights, the evidence on proactive VAs that are based on actual authentic user experience involving functional prototypes deployed in actual user settings remains scarce. While this lack of ecological implementation may be due to the technological limitations of current smart speakers, which do not allow for the desired level of proactivity implementation (except for reminders and time-based routines), future technologies will allow us to overcome this obstacle. Overall, we thus proclaim, that user studies investigating the effects of proactive behavior in VAs on user attitudes and behavior should be based on authentic user experiences with operational prototypes (RD6). Similarly, we recognize a slight shift in study design focus from studying user experiences of "proactive vs. reactive" behaviors towards studying differently designed proactive behaviors within one and across multiple scenarios (cf. Table 3) indicating another avenue for future research contributions (RD7).

We further argue that users' individual privacy concerns and awareness associated with monitoring contextual information and potentially high degrees of intrusiveness should always be taken into account as potential cofounding variables (RD8), instead of explicitly requesting participants "to temporarily set aside privacy and data protection considerations during the survey" (Zargham et al., 2023, p. 302). Interestingly, none of the studies of the second sample purposefully assessed or even comprehensively discussed users' privacy concerns, whereas four studies of the first sample did so (Cha et al., 2020; Guo et al., 2021, pp. 1–11; Mennicken et al., 2016; Zargham et al., 2022).

Eventually, to draw direct comparisons across studies or arrive at generalized conclusions regarding the appropriateness of proactive behavior in VAs, standardized evaluation measures are needed. While the general user experience with CAs can be gauged using validated questionnaires (Kocaballi, Laranjo, & Coiera, 2019), ensuring comparability, a standardized measure for user experience with proactive VAs remained absent in the reviewed studies. A promising research direction therefore involves creating an evaluation methodology for key aspects like appropriateness, usefulness, and invasiveness that span the three proactivity dimensions: context, initiation, and action (RD9). Furthermore, this evaluation framework should allow researchers to assess a proactive VA's adaptability to individual users.

6. Conclusion

This work offers a novel conceptual model for proactive VA behavior that serves both as an analytical framework for comprehending existing proactive VAs and as a design tool for configuring proactive behavior in new systems. A review of user experiences demonstrated that proactive VAs can offer distinct advantages, notably in safety-critical and emergency contexts, but revealed that future research should develop standardized evaluation methods for proactive VAs and gather data from prototypes in real-world settings across diverse user groups.

Funding statement

This work was funded in part by the Swiss National Science Foundation (SNSF), Sinergia Project "HEADWIND" (CRSII5 183569), and by the School of Medicine, University of St.Gallen, Switzerland.

CRediT authorship contribution statement

Caterina Bérubé: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Marcia Nißen:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Rasita Vinay:** Writing – review & editing, Formal analysis. **Alexa Geiger:** Writing – review & editing, Formal analysis. **Tobias Budig:** Writing – review & editing. **Aashish Bhandari:** Writing – review & editing, Formal analysis. **Catherine Rachel Pe Benito:** Writing – review & editing, Formal analysis. **Nathan Ibarcena:** Writing – review & editing, Formal analysis. **Olivia Pistolese:** Writing – review & editing, Formal analysis. **Pan Li:** Writing – review & editing, Formal analysis. **Abdullah Bin Sawad:** Writing – review & editing, Formal analysis. **Elgar Fleisch:** Writing – review & editing. **Christoph Stettler:** Writing – review & editing. **Bronwyn Hemsley:** Writing – review & editing. **Shlomo Berkovsky:** Writing – review & editing. **Tobias Kowatsch:** Writing – review & editing, Funding acquisition. **A. Baki Kocaballi:** Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used GPT-4 in order to improve readability and language of the final draft. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: TK, MN, AG, TB, RV, EF are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich, the Department of Management, Technology, and Economics at ETH Zurich, and the Institute of Technology Management and School of Medicine at the University of St. Gallen. The Centre for Digital Health Interventions is funded in part by CSS, a Swiss health insurer, Mavie Next, an Austrian health insurer, and MTIP, a Swiss digital health investor. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS, Mavie Next, MTIP nor Pathmate Technologies were involved in this research.

Data availability

All data utilized in this study are presented within the text or tables of the paper. The authors are willing to provide additional specific data upon request.

APPENDIX

Table A1
Methods, measures relative to proactivity, and most relevant outcomes.

	Context				Initiation		Action				
Sample/Study *First author and year [#]*	**Activity/ status**	**Time**	**Location**	**People**	**Direct**	**Indirect**	**Signal**	**Notification**	**Question**	**Suggestion**	**Performance**
Total sample (n = 21)	**20**	**7**	**9**	**2**	**17**	**7**	**1**	**4**	**10**	**14**	**4**
First sample (n = 13)	**12**	**6**	**8**	**1**	**9**	**6**	**1**	**4**	**6**	**6**	**3**
(Cha et al., 2020)	x	x				x				x	
(Guo et al., 2021, pp. 1–11)	x				x			x		x	
(Jarusriboonchai et al., 2014, pp. 98–106)			x			x	x		x		
(Mennicken et al., 2016)	x	x	x[1]		x			x		x	x[1]
(Semmens et al., 2019)	x		x			x			x		
(Schmidt & Braunger, 2018)	x	x	x[1]		x			x		x[1]	
(Szpektor et al., 2020)	x				x						
(Völkel et al., 2021)	x[1,3]				x					x	x[1,3]
(Wei et al., 2021a)[4]	x	x	x		x	x[1]			x		
(Wei et al., 2021b)[4]	x	x	x		x	x[1]			x		
(Wei et al., 2022)	x	x	x		x[2]				x		
(T. Wu et al., 2021)	x		x			x			x		
(Zargham et al., 2022)	x			x[1]	x			x		x	x[1]
Second sample (n = 8)	**8**	**1**	**1**	**1**	**8**	**1**	**-**	**-**	**4**	**8**	**1**
(Ding et al., 2022, pp. 1–19)	x				x				x	x	
(Dubiel et al., 2023, pp. 1–6)	x				x	x				x	
(Marques et al., 2023)	x			x	x					x	
(Mathis et al., 2023)	x	x			x				x	x	x
(Meck, 2023)	x				x					x	
(Meck et al., 2023)	x		x		x				x	x	
(Pakdamanian et al., 2022)	x				x					x	
(Zargham et al., 2023)	x				x				x	x	

Notes. [1]the elements refer to multiple scenarios/conditions of proactive behavior; [2]deducted from authors through indirect information; [3]based on results, instead of methods; [4]note that these rows contain two articles stemming from the same experiment.

Table A2
Methods, measures relative to proactivity, and most relevant outcomes

Study First Sample	Methods	Measures	Outcomes
Cha et al. (2020)	- **Field experiment with working prototype:** Experience Sampling Method in shared dormitory rooms ("Is now a good time to talk?") - N = 40 (20 rooms/pairs of roommates)	- Factors of interruptibility: day; time; activity - Interview to assess the reason for interruptibility	- Influence of activity demand/urgency, psychological state, and movements: - Most uninterruptible: preparing to sleep (89%); working/studying (82%); having face-to-face interactions (74%); playing video games (66%); departures (65%) - Most interruptible: entrance (96%); using the Internet/smartphone (92%); doing chores (85%); resting (83%); activity transition (82%); watching videos (72%). - Small influence on interruptibility of presence of both roommates: Alone (55.35%) vs. Together (50.38%)
Guo et al. (2021)	- **Online study with working prototype (experiment + survey):** rule-based VA vs proactive VA (Shing) for risky payment detection alert support on a digital finance platform in China - N = 162	- Aborted transactions rate - Self-reports: trustworthiness; privacy concerns; - 7-point Likert scale (1 = strongly disagree; 7 = strongly agree)	- Aborted transaction rate: proactive VA (29.6%) vs rule-based VA (27.3%) - Pearson Chi-squared sign (p=.001) - Willingness to converse (provide transaction details): proactive VA (55.8%) vs rule-based VA (52.9%), Pearson Chi-squared sign (p=.001) - Trustworthiness: proactive VA (M = 5.76, SD = 1.55) vs rule-based VA (M = 5.56, SD = 1.75) - T-test ns. - Privacy concern: proactive VA (M = 4.18, SD = 2.19) vs rule-based VS (M = 4.45, SD = 1.88) - T-test ns
Jarusriboonchai et al. (2014)	- **Lab experiment with WOZ method:** two stranger PP were given a smartphone and brought to a semi-public coffee room; the VAs of both smartphones started by giving the same signal to inform they were connected and would then start asking questions to both PP - N = 36 (18 sessions)	- Thematic analysis of pairs interacting with VAs and with each other - *(video and audio recording)*	- VAs took a dominant role: PP mostly reacted to VA's proactive behavior (15/18) and let the VA moderate the conversation (12/18). - Only in a minority of cases conversations were developed beyond the answers given to the VAs (7/18).
Mennicken et al. (2016)	- **Lab experiment with scenarios and mock-up:** Extroverted and Cheerful (EC) vs. Conscientious, Kind, and Calm (CKC) smart home, communicating with the user through a VA and simulating scenarios of daily routine tasks - N = 41	- Questionnaire on "what they liked and disliked in the experience" - Rating of "several statements about trust privacy, personal preferences, and general experiences"	- Preference for CKC (61%) vs EC (39%) - Privacy concern: EC perceived as "patronizing", "crossing social boundaries", and as a "stranger in their home"
Semmens et al. (2019)	- **Field experiment with VA prototype in a real car:** Experience Sampling Method during a standardized road drive ("Is now a good time?") - N = 62	- Multi-channel: video of the cabin and road, audio, vehicle telemetry, position, inertial forces, and driver physiological data - Location and CAN data - Survey: Is now a good time (YES/NO)	- Probability of responding YES is higher when: o on correct driving route (81.3% YES) compared to off-road route (66.9% YES) othe steering wheel is straight othe car is moving at a constant speed othe car is completely stopped othe car is accelerating (vs. braking) othe participant is releasing the brake versus applying the brake
Schmidt and Braunger (2018)	- **Online study with survey on experience and with scenarios:** personalized proactive VAs with US and DE users - N = 1′550 (1′051 US)	- Satisfaction with current adaptive VAs - Attitude towards data-driven personalization and adaptive suggestions - Preference of in-vehicle proactive VA scenarios - All 5-points Likert scale rating	- Positive influence of in-vehicle VA usage frequency on user satisfaction - In general, compared to DE users, US users tend to have a more positive attitude towards data-driven personalization and adaptive suggestions - General preference for proactive VA giving notifying and suggesting, compared to proactive VA just notifying or saying nothing (non-proactive VA) - Slightly higher preference in US PP for proactive VA
Szpektor et al. (2020)	- **Lab experiment:** human-rated evaluation with working prototype - **Online experiment:** live experiment with working prototype - Comparison proactive VA vs Google Assistant in conducting dialogues on NBA - N==NR	- Lab study: Rate overall conversation experience for 200 dialogues with at least 3 turns; - scale from 1 to 5 - Online study: N of NBA queries, follow-up rate; NBA follow-up rate; other-sports follow-up rate; explicit positive feedback; explicit negative feedback	- **Lab study:** proactive VA 4.45/5 vs. Google Assistant 3.89/5 - **Online study:** results in change from Google Assistant to proactive VA oN of NBA queries +3.9; oFollow-up rate +2.9; oNBA follow-up rate +4.2; oOther-sports follow-up rate +3.1; oExplicit positive feedback +15.6; oExplicit negative feedback +23.5
Völkel et al. (2021)	- **Online experiment with scenarios:** Users create conversation with the "perfect" VA in multiple scenarios	- Conversation initiation rate (proactive VA vs user-initiated)	- 8.1% of cases had a proactive VA - Scenario Volume had the highest proactive VA rate (27.3% of dialogues) - "In summary, most people envisioned dialogues with a perfect voice assistant that were highly

(continued on next page)

Table A2 (*continued*)

Study First Sample	Methods	Measures	Outcomes
	- Scenarios: Search, Music, Internet, Volume, Weather, Joke, Conversational, Alarm, Open scenario - N = 205		interactive and not purely functional; it is smart, proactive, and has personalized knowledge about the user. On the other hand, peoples' attitude towards the assistant's role and it expressing humor and opinions diverged."
Wei et al. (2021a)	- **Field experiment with mock-up** in private homes: Experience Sampling Method with different initiations: proactive VA started talking directly, after an earcon, or verbally solicits the user and waits for confirmation before talking ("Hey, are you available?") - N = 7	- Average activation rate (AR, i.e., times the voice application was successfully invoked) - Response rate (RR, i.e., times the user responded)	- Direct: oAR: 0.89 oRR: 0.38 - Earcon: oAR: 0.88 oRR: 0.43 - Verbal solicitation: oAR: 0.91 oRR: 0.36
Wei et al. (2021b)	- **Field experiment with mock-up** in private homes with recording-based voice application invocation: Experience Sampling Method with different initiations: proactive VA started talking directly, after an earcon, or verbally solicits the user and waits for confirmation before talking ("Hey, are you available?") - N = 13	- Factors of interruptibility: oself-rated availability; oself-rated boredom; oself-rated mood; oreported activity; oambient light level; onoise level; ouser proximity; - Questionnaire: appropriateness; - 5-point Likert scale	- PP are most interruptible during entertainment and study/work type of activities, are in proximity, to the proactive VA ($r = 0.128$, $p<.001$), when the ambient light is lower (only valid if speaker is in the living room; $r = -0.134$, $p<.001$), and the noise level is lower ($r = -0.083$, $p<.005$ - Preference initiations: verbal solicitation > earcon > direct
Wei et al. (2022)	- **Field experiment with mock-up** in private homes with recording-based voice application invocation: Experience Sampling Method - N = 13	- Interaction and data entry errors during EMS (see Wei et al., 2021 a; b), and user strategies	- System error frequency: 62.8% without errors; 23.6% one error; 13.6% two or more errors - Interaction termination: 14.4% caused by accumulation of consecutive system errors, 85.6% are caused by sudden termination - Erroneous numerical answers: availability (0.2%); boredom (0.7%); mood (0.3%) - Erroneous open-ended answers: 24.7%, in particular: partially missing (2.4%); partially incorrect (12.0%); totally incorrect - (9.3%); extra information (1.0%) - User strategies: raising voice and approaching (n = 6/13); repeating (n = 8/13); phonetic and lexical changes (n = 13/13); help from others (n = 4/13)
(T. Wu et al., 2021)	- **Field experiment with mock-up dataset:** comparison between a previous interruptibility dataset and annotations from third evaluators as safe/not-safe moment to talk to the driver - N = 46	- Rating safe/not-safe moment for the proactive VA to talk to the driver (using video snippets 6 s before 9 s after query); - 5-point Likert scale (definitely no, maybe no, unsure, maybe yes, definitely yes)	- 92% of "yes" responses were annotated as "safe" - 82% of "no" responses were annotated as "safe"
Zargham et al. (2022)	- **Online experiment with scenarios:** presentation of scenarios with proactive VA - Scenarios: Meeting Reminder, Health Risk warning, Recipe suggestions, Fact Checking, Disagreement Clarification, Nudging, Technical Support, Fact Spoiler, Emergency - N = 15	- Rank scenarios of proactive behavior in terms of: Usefulness; appropriateness; invasiveness	- Emergency: most useful, most appropriate, least invasive - Fact spoiler: least useful and least appropriate - Disagreement clarification: most invasive
Second Sample			
Ding et al. (2022)	- **Lab Study:** Comparison of reactive (RBC) vs. proactive backchanneling (PBC) cues in a voice-based screening and diagnosis tool for neurocognitive disorders (e.g., dementia) - N = 36	- **By expert assessors:** appropriateness of backchanneling behavior (y/n) - **By users:** Conversational satisfaction (7-point Likert scale); audio-recorded qualitative feedback ("think aloud")	- **Appropriateness:** RBC: 89.8%, PBC: 85.5% - **Conversational satisfaction:** not statistically significant differences between conditions - **Qualitative feedback:** RBC not as good as humans, PBC appreciated by older adults
Dubiel et al. (2023)	- **Lab (Wizard-of-Oz) study:** comparison of no (baseline) vs. solicited (indirect) vs. unsolicited (direct) feedback on food choice - N = 30	- **Appropriateness:** 11-point Likert-scale	- **Appropriateness:** baseline vs. solicited, *t*-test ns; baseline vs. unsolicited, *t*-test p = 0.015.
Marques et al. (2023)	- **Focus group studies,** scenario/vignette-based: - N = 5 users, 5 experts	- **Communication dimensions:** VAs voice: human-like vs. robotic VAs speech: cordial vs. rude - **Intrusion dimension:** appropriateness, access to user-related information - Relevance and usefulness of proactive behavior	- **Communication dimensions:** VAs voice: rather perceived as robotic VAs speech: cordial > rude - **Intrusion dimension: appropriateness:** >50% users considered it appropriate; experts invasive because VA need contextual information from user - **Relevance and usefulness** of proactive behavior: depends
Mathis et al. (2023)	- **Lab study (vehicle mock-up, Wizard-of-Oz VA):** evaluation of opportune and non-opportune moments (based on driving situation and driver activity) for proactive voice interaction	- **After prompt:** Usefulness, Opportuneness; Single-item 7-Likert scales	- **Most opportune moment:** standstill; NDRA2 > constant > emergency - No significant differences in usefulness between situations

(*continued on next page*)

Table A2 (*continued*)

Study First Sample	Methods	Measures	Outcomes
	- N = 32	- **After the ride:** qualitative feedback based on video-supported retrospective think-aloud study	- Qualitative feedbacks: timings often too late (20%)
Meck (2023)	- **Lab study (vehicle mock-up, Wizard-of-Oz VA):** Comparison of linguistically differently designed prompts (in terms of complexity, form of address, suggestive language) - N = 48	- Approval rate - 7-point Likert scales: positivity, intelligence, comprehensibility, naturalness	- Sentence structure: Parataxes - Sentence length: Short; Position of sub-clauses: prepositive Form of address: "I": positive | "you": intelligent politeness: no politeness: natural | politeness: intelligent voice: no significant preferences
Meck et al. *(2023)*	- **Scenario-based online studies:** Comparison of linguistically differently designed prompts (in terms of complexity, form of address, suggestive language) in proactive and not-proactive (informational, functional) VAs - N = 200 (x3)	- Preference (multiple A/B tests between two prompts)	- No differences in formulation preferences were found for functional and informational prompts. Within proactive prompts, preferences for the position of sub-clauses differed significantly between domains: Chi-squared = 6.4, $df = 2$, $p = .04$. Subsequent post-hoc Dunn Bonferroni tests showed a significant difference between the comfort-oriented and the functional domain ($p = .03$), although the effect size was found to be small, Cohen's $r = 0.007$.
Pakdamanian et al. (2022)	- **Lab Study:** Comparison of baseline (always speech-based warnings) vs. context-dependent modality (i.e., "context-aware advisory warnings" (CAWA): different advisory warning modalities such as text, vibrotactile, speech, visual for different NDRT types (gaming, talking, reading) - N = 20; - 456 TORs (19 Participants, 2 Trails, 12 true TORs per trail) - Left one participant out due to misleading biometric data	- By drivers: (1) takeover quality: reaction time; lateral vehicle control; (2) situational awareness (i.e., gaze behavior: percentage of driver looking at the road; fixation duration of when a driver's eyes are on/off the road); (3) stress, cognitive workload: heart rate variability, Driving Activity Load Index (DALI); (4) driver perception: safety, disruptiveness, urgency (5-point Likert); (5) qualitative/rankings: preference and usefulness	(1) reaction time significant faster in CAWA than in baseline based on ANOVA analysis; lateral vehicle control highest in "visual warning" modality (2) participants remain more vigilant in Baseline (3) no effect on stress; cognitive workload higher in Baseline (4) not statistically significant: but CAWA was rater safer with higher urgency, but more disruptive than Baseline
Zargham et al. *(2023)*	- **Scenario-based online experiment:** comparison of humorous (intervention) vs. non-humorous (baseline) version of VA scenario (based on cartoons) - N = 50	- Desirability: usefulness, appropriateness, invasiveness, consideration - 7-point Likert scales humor	- Results: (Dimension, Median, Inter Quartile Ranking) oInterest (5; 4) oEnjoyment (5; 3) oPerceived usefulness (4; 2.25) o- > general balanced but high IQR - > diverse viewpoints - Wilcoxon Signed-Rank test showed that scenarios without humor are ranked higher than those with humor - Only if VA is on similar social level, then user open for humor by VA

Notes. PP = participants; NR = not reported; sign = significant; ns = not significant.

References

Abdolrahmani, A., Howes Gupta, M., Vader, M.-L., Kuber, R., & Branham, S. (2021). Towards more transactional voice assistants: Investigating the potential for a multimodal voice-activated indoor navigation assistant for blind and sighted travelers. *Proceedings of the 2021 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3411764.3445638

Alharbi, S. S., Alrazgan, M., Alrashed, A., Alnomasi, T., Almojel, R., Alharbi, R., et al. (2021). Automatic speech recognition: Systematic literature review. *IEEE Access, 9*, 131858–131876. https://doi.org/10.1109/ACCESS.2021.3112535

Allen, A. A., Shane, H. C., & Schlosser, R. W. (2018). The Echo™ as a speaker-independent speech recognition device to support children with autism: An exploratory study. *Advances in Neurodevelopmental Disorders, 2*(1), 69–74. https://doi.org/10.1007/s41252-017-0041-5

Altman, D. G. (1990). *Practical statistics for medical research*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429258589

Balasuriya, S. S., Sitbon, L., Bayor, A. A., Hoogstrate, M., & Brereton, M. (2018). Use of voice activated interfaces by people with intellectual disability. *Proceedings of the 30th Australian conference on computer-human interaction*. https://doi.org/10.1145/3292147.3292161

Bizzaco, M. (2022). 5 unique and creative uses for Alexa's Sound Detection feature. *Digital Trends*. on 12.02.2024 from *www.digitaltrends.com/home/five-best-uses-for-alexa-sound-detection/*.

Boyle, B., & O'Brolcháin, F. (2023). Understanding the assistive potential of consumer technologies: A case example of smartphones, smart speakers, and internet of things technologies. In A. Rafael Garcia-Ramirez (Ed.), *Trends in assistive technologies* (Vol. 19). IntechOpen. https://doi.org/10.5772/intechopen.110452.

Cha, N., Kim, A., Park, C. Y., Kang, S., Park, M., Lee, J.-G., et al. (2020). Hello there! Is now a good time to talk? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(3), 1–28. https://doi.org/10.1145/3411810

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., et al. (2019). The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers, 31* (4), 349–371. https://doi.org/10.1093/iwc/iwz016

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass, 3*(3), 292–313. https://doi.org/10.1111/j.1751-9004.2009.00170.x

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies — why and how. *Knowledge-Based Systems, 6*(4), 258–266. https://doi.org/10.1016/0950-7051(93)90017-N

Dey, A. K. (2009). Context-aware computing. In J. Krumm (Ed.), *A Chapman & Hall book. Ubiquitous computing fundamentals* (pp. 321–352). Boca Raton, Fla., London, New York: CRC Press/Taylor & Francis. https://doi.org/10.1201/9781420093612-12.

Dey, A. K., Abowd, G. D., & Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction, 16*(2–4), 97–166. https://doi.org/10.1207/S15327051HCI16234_02

Ding, Z., Kang, J., HO, T. O. T., Wong, K. H., Fung, H. H., Meng, H., et al. (2022). TalkTive: A conversational agent using backchannels to engage older adults in neurocognitive disorders screening. *Proceedings of the 2022 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3491102.3502005

Dubiel, M., Bongard-Blanchy, K., Leiva, L. A., & Sergeeva, A. (2023). Are you sure you want to order that?: On appropriateness of voice-only proactive feedback strategies.

Proceedings of the 5th international conference on conversational user interfaces. https://doi.org/10.1145/3571884.3604312
Duffy, O., Synnott, J., McNaney, R., Brito Zambrano, P., & Kernohan, W. G. (2021). Attitudes toward the use of voice-assisted technologies among people with Parkinson disease: Findings from a web-based survey. *JMIR Rehabilitation and Assistive Technologies, 8*(1), Article e23006. https://doi.org/10.2196/23006
Dyke, G., Adamson, D., Howley, I., & Penstein Rosé, C. (2012). Towards academically productive talk supported by conversational agents. In S. A. Cerri (Ed.), *Proceedings of the 11th international conference on intelligent tutoring systems* (pp. 531–540). https://doi.org/10.1007/978-3-642-30950-2_69
Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A State of the Art Review of large Generative AI models. *ArXiv. Advance online publication*. https://doi.org/10.48550/arXiv.2301.04655
Greuter, S., Balandin, S., & Watson, J. (2019). Social games are fun: Exploring social interactions on smart speaker platforms for people with disabilities. *Extended abstracts of the annual symposium on computer-human interaction in play companion*. Extended Abstracts. https://doi.org/10.1145/3341215.3356308
Grosinger, J. (2022). On proactive human–AI systems. *Proceedings of the 8th international workshop on artificial intelligence and cognition. AIC'22)*.
Guo, J. J., Guo, J. J., Yang, C., Wu, Y., & Sun, L. (2021). Shing: A conversational agent to alert customers of suspected online-payment fraud with empathetical communication skills. *Proceedings of the 2021 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3411764.3445129
Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the 1999 SIGCHI conference on human factors in computing systems*. https://doi.org/10.1145/302979.303030
Hwang, A., Oza, N., Callison-Burch, C., & Head, A. (2023). Rewriting the script: Adapting text instructions for voice interaction. *Proceedings of the 2023 ACM designing interactive systems conference*. https://doi.org/10.1145/3563657.3596059
Jarusriboonchai, P., Olsson, T., & Väänänen-Vainio-Mattila, K. (2014). User experience of proactive audio-based social devices. *Proceedings of the 13th international conference on mobile and ubiquitous multimedia*. https://doi.org/10.1145/2677972.2677995
Jovanovic, M., Baez, M., & Casati, F. (2021). Chatbots as conversational healthcare services. *IEEE Internet Computing, 25*(3), 44–51. https://doi.org/10.1109/MIC.2020.3037151
Kocaballi, A. B., Laranjo, L., & Coiera, E. (2019). Understanding and measuring user experience in conversational interfaces. *Interacting with Computers, 31*(2), 192–207. https://doi.org/10.1093/iwc/iwz015
Kraus, M., Fischbach, F., Jansen, P., & Minker, W. A comparison of explicit and implicit proactive dialogue strategies for conversational recommendation. Proceedings of the 12th conference on language resources and evaluation, 429–435.
Kraus, M., Wagner, N., & Minker, W. (2020). Effects of proactive dialogue strategies on human-computer trust. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization* (pp. 107–116). https://doi.org/10.1145/3340631.3394840
Liao, Q. V., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What can you do?: Studying social-agent orientation and agent proactive interactions with an agent for employees. In *Proceedings of the 2016 ACM conference on designing interactive systems* (pp. 264–275). https://doi.org/10.1145/2901790.2901842
Marques, T., Abreu, J., & Santos, R. (2023). Proactivity in the TV context: Understanding the relevance and characteristics of proactive behaviours in voice assistants. In *Proceedings of the 2023 ACM international conference on interactive Media experiences* (pp. 314–319). https://doi.org/10.1145/3573381.3597219
Masina, F., Orso, V., Pluchino, P., Dainese, G., Volpato, S., Nelini, C., et al. (2020). Investigating the accessibility of voice assistants with impaired users: Mixed methods study. *Journal of Medical Internet Research, 22*(9), Article e18431. https://doi.org/10.2196/18431
Mathis, L.-A., Werner, K., & Schmidt, G. J. (2023). Proactive voice assistance in automated vehicles: Understanding when to engage the user. *Stuttgarter Symposium für Produktentwicklung*. https://doi.org/10.24406/PUBLICA-1499
McTear, M., Callejas, Z., & Griol, D. (2016). *The conversational interface*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3
Meck, A.-M. (2023). Secure, comfortable or functional: Exploring domain-sensitive prompt design for in-car voice assistants. In *Proceedings of the 5th international conference on conversational user interfaces* (pp. 1–5). https://doi.org/10.1145/3571884.3604314
Meck, A.-M., Draxler, C., & Vogt, T. (2023). How may I interrupt? Linguistic-driven design guidelines for proactive in-car voice assistants. *International Journal of Human-Computer Interaction*, 1–15. https://doi.org/10.1080/10447318.2023.2266251
Mehrotra, A., Musolesi, M., Hendley, R., & Pejovic, V. (2015). Designing content-driven intelligent notification mechanisms for mobile applications. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 813–824). https://doi.org/10.1145/2750858.2807544
Mehrotra, A., Pejovic, V., Vermeulen, J., Hendley, R., & Musolesi, M. (2016). My phone and me: Understanding people's receptivity to mobile notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 1021–1032). https://doi.org/10.1145/2858036.2858566
Mennicken, S., Zihler, O., Juldaschewa, F., Molnar, V., Aggeler, D., & Huang, E. M. (2016). "It's like living with a friendly stranger": Perceptions of personality traits in a smart home. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 120–131). https://doi.org/10.1145/2971648.2971757
Metatla, O., Oldfield, A., Ahmed, T., Vafeas, A., & Miglani, S. (2019). Voice user interfaces in schools: Co-Designing for inclusion with visually-impaired and sighted pupils. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–15). https://doi.org/10.1145/3290605.3300608
Meurisch, C., Ionescu, M.-D., Schmidt, B., & Mühlhäuser, M. (2017). Reference model of next-generation digital personal assistant. In *Proceedings of the 2017 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2017 ACM international symposium on wearable computers* (pp. 149–152). https://doi.org/10.1145/3123024.3123145
Meurisch, C., Mihale-Wilson, C. A., Hawlitschek, A., Giger, F., Müller, F., Hinz, O., et al. (2020). Exploring user expectations of proactive AI systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(4), 1–22. https://doi.org/10.1145/3432193
Monteiro, P., Goncalves, G., Coelho, H., Melo, M., & Bessa, M. (2021). Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics, 27*(5), 2702–2713. https://doi.org/10.1109/TVCG.2021.3067687
Nweke, H. F., Teh, Y. W., Mujtaba, G., & Al-Garadi, M. A. (2019). Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion, 46*, 147–170. https://doi.org/10.1016/j.inffus.2018.06.002
Pakdamanian, E., Hu, E., Sheng, S., Kraus, S., Heo, S., & Feng, L. (2022). Enjoy the ride consciously with CAWA: Context-aware advisory warnings for automated driving. In *Proceedings of the 14th international conference on automotive user interfaces and interactive vehicular applications* (pp. 75–85). https://doi.org/10.1145/3543174.3546835
Pradhan, A., Mehta, K., & Findlater, L. (2018). "Accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3173574.3174033
Radford, A., Kim, J. W., Xu, T., Brockman, G., Mcleavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th international conference on machine learning* (pp. 28492–28518). https://doi.org/10.5555/3618408.3619590
Reicherts, L., Zargham, N., Bonfert, M., Rogers, Y., & Malaka, R. (2021). May I interrupt? Diverging opinions on proactive smart speakers. In *Proceedings of the 3rd conference on conversational user interfaces* (pp. 1–10). https://doi.org/10.1145/3469595.3469629
Schmidt, M., Bhandare, O., Prabhune, A., Minker, W., & Werner, S. (2020). Classifying cognitive load for a proactive in-car voice assistant. In *2020 IEEE sixth international conference on big data computing service and applications (BigDataService)* (pp. 9–16). https://doi.org/10.1109/BigDataService49289.2020.00010
Schmidt, M., & Braunger, P. (2018). A survey on different means of personalized dialog output for an adaptive personal assistant. In *Adjunct publication of the 26th conference on user modeling, adaptation and personalization* (pp. 75–81). https://doi.org/10.1145/3213586.3226198
Semmens, R., Martelaro, N., Kaveti, P., Stent, S., & Ju, W. (2019). Is now A good time? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300867
Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ, 350*, Article g7647. https://doi.org/10.1136/bmj.g7647
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415
Shum, H. Y., He, X., & Di, L. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering, 19*, 10–26. https://doi.org/10.1631/FITEE.1700826
Smith, E., Sumner, P., Hedge, C., & Powell, G. (2020). Smart-speaker technology and intellectual disabilities: Agency and wellbeing. *Disability and Rehabilitation: Assistive Technology*, 1–11.
Smith, E., Sumner, P., Hedge, C., & Powell, G. (2021). Smart speaker devices can improve speech intelligibility in adults with intellectual disability. *International Journal of Language & Communication Disorders, 56*(3), 583–593. https://doi.org/10.1111/1460-6984.12615
Smith, E., Sumner, P., Hedge, C., & Powell, G. (2023). Smart-speaker technology and intellectual disabilities: Agency and wellbeing. *Disability and Rehabilitation: Assistive Technology, 18*(4), 432–442. https://doi.org/10.1080/17483107.2020.1864670
Strauss, P.-M., & Minker, W. (2010). *Proactive spoken dialogue interaction in multi-party environments* (1st ed.). New York: Springer. https://doi.org/10.1007/978-1-4419-5992-8
Szpektor, I., Cohen, D., Elidan, G., Fink, M., Hassidim, A., Keller, O., et al. (2020). Dynamic composition for conversational domain exploration. In *Proceedings of the web conference 2020* (pp. 872–883). https://doi.org/10.1145/3366423.3380167
Thoo, Y.-J., Jeanneret Medina, M., Froehlich, J. E., Ruffieux, N., & Lalanne, D. (2023). A large-scale mixed-methods analysis of blind and low-vision research in ACM and IEEE. In *Proceedings of the 25th international ACM SIGACCESS conference on computers and accessibility* (pp. 1–20). https://doi.org/10.1145/3597638.3608412
Völkel, S. T., Buschek, D., Eiband, M., Cowan, B. R., & Hussmann, H. (2021). Eliciting and analysing users' envisioned dialogues with perfect voice assistants. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–15). https://doi.org/10.1145/3411764.3445536
Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems, 37*(9), 205–224. https://doi.org/10.17705/1CAIS.03709

Wei, J., Dingler, T., & Kostakos, V. (2021a). Developing the proactive speaker prototype based on Google home. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3411763.3451642

Wei, J., Dingler, T., & Kostakos, V. (2021b). Understanding user perceptions of proactive smart speakers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5*(4), 1–28. https://doi.org/10.1145/3494965

Wei, J., Tag, B., Trippas, J. R., Dingler, T., & Kostakos, V. (2022). What could possibly go wrong when interacting with proactive smart speakers? A case study using an esm application. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–15). https://doi.org/10.1145/3491102.3517432

Wu, T., Martelaro, N., Stent, S., Ortiz, J., & Ju, W. (2021). Learning when agents can talk to drivers using the INAGT dataset and multisensor fusion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5*(3), 1–28. https://doi.org/10.1145/3478125

Wu, F., Nägele, M., Cleres, D., Haider, T., Fleisch, E., Ruschitzka, F., et al. (2023). Trends in voice characteristics in patients with heart failure (VENTURE) in Switzerland: Protocol for a longitudinal observational pilot study. *PLoS One, 18*(4), Article e0283052. https://doi.org/10.1371/journal.pone.0283052

Yorke-Smith, N., Saadati, S., Myers, K. L., & Morley, D. N. (2012). The design of A proactive personal agent for task management. *The International Journal on Artificial Intelligence Tools, 21*(1), Article 1250004. https://doi.org/10.1142/S0218213012500042

Yu, C., Shane, H., Schlosser, R. W., O'Brien, A., Allen, A., Abramson, J., et al. (2018). An exploratory study of speech-language pathologists using the echo Show™ to deliver visual supports. *Advances in Neurodevelopmental Disorders, 2*(3), 286–292. https://doi.org/10.1007/s41252-018-0075-3

Zargham, N., Reicherts, L., Avanesi, V., Rogers, Y., & Malaka, R. (2023). Tickling proactivity: Exploring the use of humor in proactive voice assistants. In *Proceedings of the 22nd international conference on mobile and ubiquitous multimedia* (pp. 294–320). https://doi.org/10.1145/3626705.3627777

Zargham, N., Reicherts, L., Bonfert, M., Voelkel, S. T., Schoening, J., Malaka, R., et al. (2022). Understanding circumstances for desirable proactive behaviour of voice assistants: The proactivity dilemma. In *Proceedings of the 4th conference on conversational user interfaces* (pp. 1–14). https://doi.org/10.1145/3543829.3543834