Laila is in a Meeting: Design and Evaluation of a Contextual Auto-Response Messaging Agent

Pranut Jain
University of Pittsburgh
Pittsburgh, PA, USA
pranutjain@pitt.edu

Rosta Farzan
University of Pittsburgh
Pittsburgh, PA, USA
rfarzan@pitt.edu

Adam J. Lee
University of Pittsburgh
Pittsburgh, PA, USA
adamlee@cs.pitt.edu

ABSTRACT

The ease of smartphone communications has created an expectation of constant connectivity. While the adoption of virtual assistants has improved, their capabilities for handling proactive communication tasks remain underexplored. We present the design, implementation, and evaluation of a Contextual Auto-Response agent to communicate users' situational awareness. The agent creates auto-responses by modeling availability using smartphone sensors and sharing contextual information on behalf of the user. In a two-week study with 12 participants, we evaluated the perception of this agent and its impact on device usage behavior. Many participants found the agent useful for signaling unavailability, with some caveats. Participants also reported altering device and agent usage based on their understanding of its functions. Our findings indicate the importance of transparency in proactive agent designs and the need for personalization to enable an enhanced and cooperative human-agent interaction.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

user modeling, awareness, virtual assistant, privacy, context-sharing, AI explanations

ACM Reference Format:
Pranut Jain, Rosta Farzan, and Adam J. Lee. 2022. Laila is in a Meeting: Design and Evaluation of a Contextual Auto-Response Messaging Agent. In *Designing Interactive Systems Conference (DIS '22), June 13–17, 2022, Virtual Event, Australia.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3532106.3533493

1 INTRODUCTION

Laila is in a meeting, their phone buzzes, and a new notification of the fifth text message on the phone flashes. The meeting is important, it will be rude to check the phone, but Laila also worries about the friend sending the messages feeling ignored. They wish they could quickly send a message of "busy in a meeting". With the prominence of communication through text messaging, this experience is well-known to many of us. Attending to the constant flow of messages during our daily tasks causes disruption [56] and impacts task completion and task performance negatively [3, 15, 47]. On the other hand, ignoring the notification might also lead to an undesirable outcome. Past studies have shown that people start to speculate when they do not receive a response within their expected time [60]. These speculations are often negative [24] (e.g., being upset or angry and feeling ignored) and are often further fueled by inaccurate availability indicators of messaging applications [39, 50]. To rectify the negative social experience caused by this, recipients often have to apologize and explain delays in responding [22, 28, 54, 61].

Virtual assistants are growing in prominence in a variety of areas [55]. The ability and scope of these virtual agents have improved with language processing techniques to allow interaction through natural language in text (e.g., chatbots) or speech (e.g., voice assistants like Amazon Alexa and Apple Siri). There is ongoing research to improve their utility and add more capabilities to voice assistants, as evident from recent publications on developing frameworks for managing a user's calendar [13], supporting the elderly and disabled with daily tasks [32, 59] and even content acquisition [44]. These works promote the potential for virtual assistants to manage communication for their users. In particular, a proactive agent design can help with not only managing interruptions but also with improving situational awareness in communication [48]. The agent can be designed to reduce disruptions if it can work autonomously without user intervention. However, a key challenge to the design of such an agent is detecting and communicating contextualized unavailability while maintaining user expectations of privacy.

In this work, we build upon the prior literature to explore the design and implementation of a fully automated approach for generating and sending auto-responses as a means to improve situational and unavailability awareness. Our approach entails an evaluation of each new messaging session to predict the availability of the message recipient. If deemed unavailable, an auto-response is generated, which shares the predicted relevant recipient's context, using a user attentiveness model. With the implementation of the messaging agent, we aim to achieve three goals (1) reducing users' engagement with their devices when they are busy with other tasks; (2) improving situational awareness for users' social contacts; and (3) maintaining users' privacy through mutual awareness of the information shared through the agent. We present this agent's design, implementation, and evaluation concerning these goals from the perspective of its users.

In particular, we explore the following research questions:

DIS '22, June 13–17, 2022, Virtual Event, Australia

ACM ISBN 978-1-4503-9358-4/22/06…$15.00
https://doi.org/10.1145/3532106.3533493

- What are important considerations in designing an automated availability management agent to reduce device engagement while maintaining user privacy?
- What are users' perceptions and interpretations of the auto-responses generated from the information captured from sensor data?
- In what ways does the presence of an availability managing agent affect user behavior?

To answer these research questions, we developed and evaluated an availability management agent through an empirical two-week-long study with 12 participants who used our messaging agent on their smartphones for the study period. Our findings suggest that participants found the agent useful for communicating unavailability when they could not get to their phone. Participants also reported altering their behavior based on their understanding of the agent's design and function to appropriate it in their desired way. We also learned how inaccuracies in the agent's behavior lead to a sense of loss of interaction control. This occurred when the information shared to message senders by the agent was considered either irrelevant or inappropriate by the message recipients. This also resulted in an increased effort by message recipients to explain the agent's actions to their contacts.

Overall, our work contributes to the field of designing interactive systems by (1) presenting a novel design of a fully automated messaging agent that learns from users messaging behavior to identify and share relevant context related to their unavailability state (Section 3); (2) by describing ways in which this agent can be useful (Section 6.1) and what factors affect its utility for its users (Section 6.2). (3) by providing insights on how presenting mid-level sensed information (Section 3.2) rather than the inferred state can be perceived by users and under what circumstances such messages can be effective or misinterpreted (Section 6.3); and (4) by empirically evaluating the role of the agent in both positive and negative users' behavior changes (Section 6.4).

2 BACKGROUND AND RELATED WORK

Our research builds upon prior work in (1) Virtual assistants and technological behavior change; (2) Context-based designs and explanations; (3) Availability management in messaging. This section explores how they support our work and how our work extends the existing research. We start by discussing prior work on availability management in mobile messaging and the gaps in research that our work targets. Next, we discuss the utility of virtual assistants in reducing distractions and the importance of utilizing sensed contextual information in terms of helping with agent tasks and being utilized for explanations of agent actions.

2.1 Virtual assistants and technological behaviour change

Virtual assistant technologies are an active area of research, as evidenced by the research publications presenting designs for various general applications. In particular, virtual assistant technologies have been explored and utilized in tasks such as smart-home automation for controlling appliances with voice or text commands for people with special needs [43], academic guidance for students [40], navigation [46] and cooking assistance [45]. Most of these studies either present a novel design for a virtual assistant to accomplish the outlined task or augment/extend their capabilities (e.g., conflict resolution in case of multiple users [43]). More recent research has also looked at the use of virtual assistants for reducing distractions [21, 30]. Kimani et al. [30] designed and evaluated a virtual assistant, called *Amber*, as a conversational agent that allowed users to interact with the agent to schedule tasks and breaks. The agent also intervenes when it detects a user is spending more than a set time on social media. Similarly, persuasive systems [42] often utilize virtual assistants to provide an interface to their users to make them aware of their activities and allow them to reflect and foster change in behavior or attitudes [63]. These persuasive systems have been evaluated in health and physical activity [12] as well as for improving productivity through DPAs (Digital Productivity Systems) [63].

While researchers have evaluated virtual assistants for their role in behavior changes related to fields such as health and physical activity [12] and medicine adherence [4], we are mainly focusing on behavior changes related to the use of technology such as social media and any appropriation of technology to better suit individual needs [52]. The findings of Kimani et al. [30] reported that participants found agent suggestions around breaks and reflection useful and reported behavior changes in their routine with the use of the agent. Further, Grover et al. extended the work of Kimani et al. by introducing anthropomorphic features through a voice assistant and observed that this improved agent perception and its use for some participants [21]. Similarly, in persuasive agent designs, agent nudges have been observed to help with reducing time spent on social media [62]. In addition to self-behavior changes related to the use of technology, people have also been observed to appropriate technology to suit their needs better. For instance, in terms of communication, Retore et al.'s findings suggested that people tailor the way they use different controls on messaging applications (such as Slack and WhatsApp) depending on the *context-of-use* i.e., based on their situations and types of controls offered. These findings suggest that virtual assistants continue to show potential for improving the general well-being of their users. While virtual assistants can block notifications or silence smartphones to reduce disruptions, there is a stronger sense of obligation to respond to incoming messages. Even if ignored, the lack of awareness of the recipient state can negatively affect social relations and often requires effort to repair these social situations (e.g., by apologizing and explaining delays [28, 61]).

Our work augments this body of knowledge on virtual assistants by presenting and evaluating a novel design of an agent to manage user communication. By being cognizant of its user's state, the virtual assistant described in this work can act as a mediator between message senders and the owner of the assistant. Our work forms the first step in realizing the potential of virtual assistants to act more like real assistants, which often also handle communication for their users where they communicate unavailability and share some context related to it.

2.2 Context-based designs and explanations

Sharing context or an explanation surrounding a prediction can boost users' trust in the algorithmic predictions of systems [1, 58].

Persuasive systems such as the ones discussed in Section 2.1 share context about the users' activities to enable them to reflect and adjust their behavior. For instance, information shared such as cycling data [38], sleep data [35], and physical activity [25] can bring about behavior change if they are relatable (via correlations with users' understanding of their activities) and interpretable by their users, which in turn boosts users' trust in the system [38].

Research in recommender systems has also reported the importance of personalized explanations along with recommendations through the use of artificial intelligence and machine learning techniques [17, 41]. For instance, the work by Baltrunas et al. explored the design and evaluation of a context-based mobile app that recommended applications for users to install on their phone based on their contexts, such as currently installed applications, time of day, weather, and location [5]. They used this information to generate both recommendations and explanations. For instance, the explanation for the recommendation could include the relevance of time of day and location on the recommendation. While the general perception of these explanations was positive, their findings indicated that using a more personalized approach will result in more acceptable recommendations and more relevant explanations.

Our work explores an agent design with personalized explanations. At the same time, we extend this prior research by exploring how personalized and context-aware explanations provided by a system on behalf of a user to others are perceived and utilized by the user, as opposed to the explanations provided on behalf of the system to the direct user of the system. This adds another layer of complexity to the agent's perception and the information it shares, as now the agent represents its owner to their contacts and can potentially influence their social interactions.

2.3 Availability Management in Communication

Most mobile messaging applications, including SMS (when using Rich Communication Services or RCS), have availability indicators such as online/offline status, last-seen time, custom status, and read receipts. Although, how reliably they communicate unavailability is questionable since they are inaccurate as indicators [50] and multiple publications have also highlighted privacy concerns stemming from their use to infer availability [11, 24, 50]. To reduce distractions for message recipients, multiple previous studies on interruption management have investigated techniques such as deferring notifications [49, 66] or silencing a user's phone [53]. More recently, Apple announced the implementation of this in the form of Focus mode with the upcoming release of iOS 15[1]. This mode (and associated profiles) can either be turned on manually or set to turn on based on context (e.g., a fixed time and location). It can defer or silence notifications based on their type when turned on. The iMessage app would also be able to share if the focus mode is on with contacts. Although, it is currently limited to only Apple devices, and further investigation is required to understand its effectiveness in improving situational awareness.

However, just deferring notification or silencing the phone does not help with the situational awareness of the message recipient. Multiple recent works have looked at different ways of informing the availability of message recipients. Cho et al. investigated a manual status sharing approach for KakaoTalk[2] messaging application [10]. The proposed approach is proactive, i.e., the status is shared automatically with selected contacts on receiving a new message. While manual status sharing can provide more accurate and relevant status information, the reliance on users to set and update their status can lead to unreliable or outdated status information [6, 8]. Pielot et al. instead proposed a method to share the attentiveness level of message recipients [50]. They predict this level using machine learning with features representing information captured on a smartphone, such as ringer mode, screen status, or proximity. Their evaluation showed this automated approach to be accurate (71%) and perceived as less privacy-invasive in their user study. Wu et al. developed the IMStatus application to understand further the perception of different receptivity statuses [64]. Their application shared either computed attentiveness, responsiveness, or interruptibility status. One of these statuses was randomly chosen and was further shown in one of three different ways, i.e., textually, numerically, or graphically. Their findings suggested that participants in their study preferred an attentiveness or responsiveness status over interruptibility status, and in terms of presentation, they preferred textual descriptions over graphical or numerical. This design of a standalone application would require additional effort by message senders to check availability status. Further, a lack of mutual awareness between the two contact parties can have privacy implications since the recipient would not know who is checking their status. Finally, with both Pielot et al. and Wu et al., there is the issue of trust in the receptivity status since it would be not clear to message senders how this status was computed [1, 50].

Instead, our work proposes a fully automated design of an agent which can intercept communication from all major messaging applications on the user's phone and send auto-responses within the same communication thread. Further, the agent sends these auto-responses only when there is an attempt to start communication, and it predicts the recipient to be unavailable, limiting accessibility and enabling mutual awareness. Finally, auto-responses can also help improve situational awareness by including contextual information from a user's availability model, potentially supplementing trust in the unavailability status.

3 DESIGN OF AUTOMATED RESPONSE AGENT

The main design goals for the agent are to (1) reduce users' engagement with their devices when they are busy with other tasks; (2) improve situational awareness for users' social contacts; (3) maintain users' privacy through mutual awareness of user context. In this section, we present the design of the agent to achieve these goals.

3.1 Fully automated agent design

In order to reduce device engagement, the agent needs to act autonomously without requiring user intervention. An automated agent design would allow users to focus on their ongoing tasks, thereby reducing distractions. Furthermore, as discussed previously,

[1] https://www.apple.com/newsroom/2021/06/ios-15-brings-powerful-new-features-to-stay-connected-focus-explore-and-more/

[2] https://www.kakaocorp.com/page/service/service/KakaoTalk?lang=en

users are inconsistent in updating their status, so the agent should also ensure consistency in sharing users' status information and provide awareness to their social contacts. We designed a fully automated agent by modeling the users' messaging behavior and using this model to detect and share unavailability-related contexts.

3.1.1 Detecting and classifying messaging sessions. Similar to Avrahami et al. [2], to define a new messaging session, we used 5 minutes threshold since the last message from the same contact. This helped distinguish new messaging sessions from ongoing conversations and focus only on new sessions in modeling attentiveness rather than all incoming messages. In addition to tracking session initiation, the model also tracked when the user attended a message to generate class labels. We consider a session as *attended* if the user (1) *removes* the associated notification, (2) *opens* messaging application associated with the session, or (3) *accesses* the message on another device[3] (e.g., WhatsApp Web). For a session to be classified as "attended to", one of these events had to happen within *7.2 minutes* from the time the message was received. The choice of 7.2 minutes threshold comes from prior literature on attentiveness modeling, representing the average median attend time as the threshold for classifying attentiveness [26, 50].

3.1.2 Sensors and features used to define context. We used 58 features[4] to create the user model. We derived the feature set from phone sensor data as well as phone usage data, based on previous works on using smartphone data to create user models [26, 49, 50]. We logged two main types of information, (1) *time since an event* features - where events were cases such as change of screen's state (e.g., time since screen unlocked) or communication (e.g., time since phone was last answered); (2) *current status* features such as screen state (locked, unlocked, or covered), connectivity state (e.g., WiFi signal strength), or ringer mode (normal, silent, or vibrate). In addition to these, we also logged additional information such as (1) location, which the users semantically labeled as work, home, or other frequented locations; (2) level of background noise, using frequent processing of background sound through the phone mic [19]; and (3) Calendar information to represents events with which the users might be engaged [28, 31].

3.1.3 Modeling. We used personalized modeling of attentiveness [27, 50, 51] to predict when a user is not available to attend to their messages. Prior work has shown personalized models (1) more accurately predict users' attentiveness to their mobile messages [27]; (2) are more flexible in terms of the modeling process, optimization, and retraining of the model [26]; and (3) can better support users' privacy by enabling modifications to individual models based on comfort with specific information used to model behavior without having to retrain the general model [26]. We used a tree boosting algorithm called XGBoost [9] for creating these individual models. We used *binary logistic* as the objective method. We scaled the positive class weight to the ratio of the positive and negative class instances in our data to deal with potential class imbalance. The rest of the parameters were set to their default (*learning rate = 0.3*, *max depth = 6*, *minimum child weight = 1*) as they usually performed the best when testing on a dataset from another study [49]. Based on Jain et al. [27], we retrained these personal models once a day using cumulative data samples collected in the preceding days.

3.1.4 Detecting and sharing relevant context. For the design of the agent, rather than sharing a status type, we chose to use auto-responses within the same thread of conversation to inform unavailability. This allows the users to keep track of the agent's behavior within the specific context of a communication thread. A sample auto-response is shown in Figure 1. We distinguish each auto-response from regular messages through the use of text *"AUTO-RESPONSE:"* to signal to message senders that the message came from the agent. The base auto-response is a simple phrase *"[NAME] may not be available to respond"*. We further augment this base auto-response to include specific context, which may help explain the unavailability prediction to the auto-response recipient. The motivation for this design comes from how a human assistant may communicate unavailability, for instance, by including information such as "they are in a meeting" or "they have left the office". We illustrate this in Figure 1. The context shared in this case is the noisy background and the calendar information.

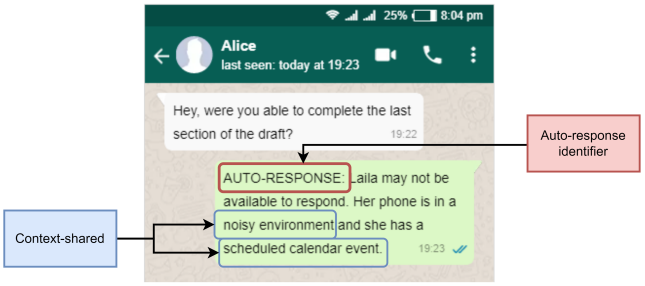

Figure 1: Sample Auto-response with two types of information being shared, device-state (noise level) and user-state (calendar event).

The next design consideration was identifying what information is relevant to share to form these augmented auto-responses. Since the availability models can achieve high levels of accuracy in predicting attentiveness [2, 26], understanding and interpreting the learned model can help identify relevant features that are associated with unavailability. We used the tree explainer component of SHAP [37] that utilizes Shapley values to produce local interpretations of each messaging session to identify these factors. Figure 2 visualizes an example of one such local interpretation for one of our participants. While these local interpretations may not link to causality, they still help identify patterns for each local prediction. We limited the number of features included in the auto-response to at most three to limit the amount of information that is shared and also reduce the cognitive load in understanding multiple items of information [58].

[3]While we tried to detect web-interfaces of messaging application (e.g., WhatsApp Web), due to the nature of notifications on Android, this detection was not always reliable. This led to some messages being falsely flagged as a new session when the participant used the web interface on another device. Two participants reported being affected by this. One participant reported being annoyed and described the agent as an "intruder in the conversation". The other participant reported the event as rare and were not significantly affected by it.

[4]List of all features used in modeling is available at https://docs.google.com/spreadsheets/d/1S59ZCWfAmVDA1Wuc0KXCZu4FMb4QVhvJUyqasAuwC4o/

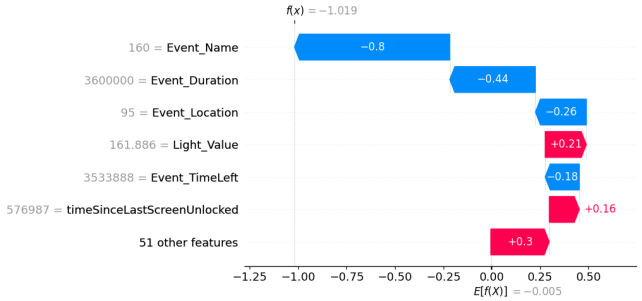

Figure 2: The figure shows a sample local interpretation for Participant P1 generated using SHAP waterfall visualization. Here, the y-axis represents features and their encoded values, while the x-axis bars represent the push of a specific feature towards a particular model output. The bars pointing towards the left or negative axis represent features pushing the model output towards unavailability. In contrast, the bars pointing to the right are pushing the model output towards available prediction. Based on this interpretation, *'Event_Name'*, which signifies a calendar event, has the most significant push towards the unavailability state. At the same time, the high *luminance* and short *time since the phone was last unlocked* are pushing the model output towards the available state.

3.2 Privacy Considerations

Ensuring users' privacy is one of the key aspects of our design decisions. By design, the agent sends auto-responses only for (1) new incoming session initiations; (2) when the model predicts unavailability; (3) contacts saved in the address book. This limits access to status information compared to typical online/offline and attentiveness [50, 64] indicators which constantly broadcast application usage status. Further, this enables mutual awareness and transparency since the message recipient is aware that information was shared with their contact through an auto-response in the same thread of conversation [10]. This approach provides high transparency between the social contacts of who has what contextual information about their availability instead of social contacts passively checking a user's status on the application.

As a design decision, we ensured that auto-response messages were neither too low level (e.g., detailed sensor data such as actual decibel noise levels or proximity readings), nor too high level (i.e., the agent is not making any inferences of the *actual* activity of the user). We call this mid-level sensor data. For instance, the agent might report that the user is in a 'dark environment' and 'silent environment' rather than inferring an associated state (e.g., sleeping). Additionally, we aggregated some low-level context values into bundles or categories. For instance, the agent shares the application category instead of sharing the last application used (e.g., productivity and communication). Similarly, instead of sharing precise location coordinates, the agent shares only labeled semantic locations which follow a circular radius along a point of reference (GPS coordinate) that the user is willing to share.

Further, the content of the message is not tracked or parsed by the agent. While the agent does use the contact name from the notification to identify new message sessions, it does not use this information for modeling availability. Finally, the identification of new sessions is local to the device, ensuring the privacy of both message content and contact information. While sensor data was sent to a remote server for modeling and prediction, as mobile devices become more capable of handling ML/AI tasks using neural co-processors, this processing can also be performed locally on the device, improving privacy even further.

4 IMPLEMENTATION OF AUTO-RESPONSE AGENT AND MESSAGES

We illustrate the design components of the application in Figure 3. The individual User, Interpretation, and Response Construction modules were deployed on a remote server, while sensor data collection and session identification were performed locally on the user's smartphone. The AWARE Framework was used as a library in our Android application [19] for sampling data from some sensors, while for the rest, we added manual listeners using Android's SensorManager class[5].

4.1 Supporting multiple applications

One of our objectives for implementing the agent was to support multiple messaging applications. Approximately 36% of smartphone users have multiple messaging applications on their phones (not including the preinstalled SMS application). People might use these applications either for different purposes (e.g., Slack for work-related discussions and WhatsApp for personal conversations [52]) or for interacting with different types of contacts [57]. Thus, the agent's utility might be limited if it cannot support multiple applications. We used Android's Notification Listener Service to intercept all notifications on the phone. We leveraged Android's Quick Reply feature, which allows users to send responses within the notification without launching the messaging application. Using either Notification Actions (introduced in Android API level 19) or Wearable Actions (on older <19 API level versions), we were able to use the Quick Reply feature to send auto-responses programmatically. This approach allowed our application to support all applications that supported this feature. Messaging and Social media applications which supported this feature at the time of the study included WhatsApp, Facebook Messenger (and Messenger Lite), Telegram, Signal, Instagram, Google Messages (and other SMS applications), and Slack.

4.2 Generating auto-response messages

Using the top features returned by the Interpretation module, we generate multiple auto-response patterns in the Response Construction module. These are patterns since we generate the actual response on the user's phone to include their name and gender preferences for an auto-response. We did not store any individual information on the application's server-side. The response types, along with their description and an example, are listed in Table 1.

[5]https://developer.android.com/guide/topics/sensors/sensors_overview

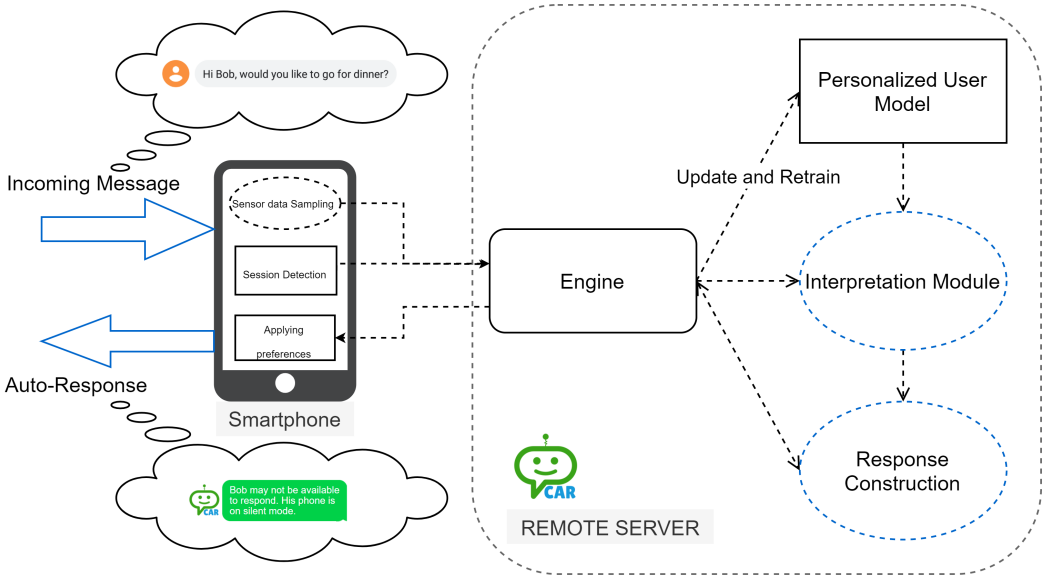

Figure 3: Agent System Design

For each messaging session where the agent predicts the user as *'unavailable'*, it generates multiple auto-response types as listed in Table 1. The single feature auto-responses further includes the top feature, second-best feature, and the third-best feature auto-responses to make a total of seven auto-response types. We construct the *weighted top-features ensemble* auto-responses by using a simple heuristic approach as follows: From the list of features and weights returned by the Interpretation module, the Response construction module picks those features that pushed the model output towards an unavailability state. If the normalized sum of weights for the top 2 features makes up 80% of the overall weight (for the unavailability prediction), then those two features are used in the auto-response; otherwise, the top 3 features are used. Suppose some constituents of an auto-response type are missing (e.g., no device-state features returned by the interpretation module), then the agent skips that auto-response type. After generating all auto-response types, the agent randomly picks one of them to send to the message sender.

The Response construction module also included a rule-base. This rule-base defined rules for phrasing different feature-value pairs and combining multiple phrases to form coherent auto-responses. Additionally, the rule-base also defined a hierarchy of features based on the type of information at different levels, such as high level (e.g., user-state or device-state) or low level (e.g., connectivity or location). This prevented the creation of multi-feature auto-responses, which included highly correlated features such as those shown in Figure 2. For example, only one of *event_name* or *event_location* features will be picked for an auto-response since they both represent a calendar event. Although, we would still consider them independent if they had a unique feature encoding defined, i.e., if *event_name* feature has a different auto-response phrase than *event_location* feature.

4.3 Pilot run

The research team carried out a month-long pilot run to (1) determine what controls to add for primary agent function; (2) fine-tune phrasing of auto-responses, especially multi-feature ones; and (3) detect and iron out any bugs in the application. The controls added based on the result of the pilot run included the ability to add a *delay* before sending an auto-response (default: 1 minute). The purpose of the delay value was to give the user a chance to respond when they were available contrary to the model prediction. This value was customizable and could range from 0 (instant auto-response) to 7 minutes (the threshold used for modeling). Another observation from the pilot run was the frequency of auto-responses for some contacts. These contacts with whom a user engages in conversation multiple times a day may get multiple auto-responses throughout the day. Multiple auto-responses in a short period can lead to annoyance and raise privacy concerns due to oversharing contextual data. Instead of limiting the number of auto-responses for a specific contact, we added a contact interval setting to the application, preventing another auto-response from being sent to the same contact for the set amount of time (default: 2 hours).

Regarding phrasing of auto-responses, we observed that some features, when chained together, resulted in redundant information in an auto-response. For example, if the top 2 features are the "high number of unattended notifications" and the "long time since screen unlocked", the resultant multi-feature auto-response would be *"Laila may not be available to respond. She has* ***not been checking her notifications*** *and has* ***not been using her phone for a while****"*. In this example, these statements sound redundant when taken together as they directly relate to one another. To prevent this redundancy, we assigned a category to each auto-response feature based on the type of information it represented, i.e., whether it was related to a *device-state* information (charge level, ringer mode, etc.)

Table 1: Auto-response types generated by the agent

Sub-type	Description	Example
No-context/Simple	This response type did not share any additional context as part of the auto-response.	Laila may not be available to respond at this time.
Single-feature	These response types use a single feature value from the Interpretation module.	Laila may not be available to respond. Her phone is covered (in a bag or pocket).
User & device ensemble	User-state information represents information about the activities or tasks of the user and their environment. In contrast, device-state information relegates information about the device (e.g., screen-state, ringer mode) [28].	Laila may not be available to respond. Her phone is covered (in a bag or pocket) and she has a scheduled calendar event.
Weighted top-features ensemble	This response type combined responses from multiple top features returned by the Interpretation module to form a single cumulative auto-response.	Laila may not be available to respond. She has not been using her phone for a while and has a scheduled calendar event and her phone is currently locked.
Clustered ensemble	The third type of multi-feature auto-response included features from at least two of three dimensions of locality, time, and task information, comprising of top features as returned from the Interpretation module.	Laila may not be available to respond. She is at work and has a scheduled calendar meeting and has not unlocked her phone for a while.

or *user-state* information (current location, calendar information, etc.). We then augmented the agent rule-base to prevent a multi-feature auto-response from including two features from the same category. The agent then picked the features with the higher weight from the Interpretation module to be included in the auto-response.

5 USER STUDY

We evaluated our approach to modeling and implementing the auto-response agent in a two-week user study with 12 participants. We recruited our participants through advertisements on the university news web page, flyers around campus, and social media listings. They were briefed remotely about the description and requirements of the study. In the 15-minutes session, we also described our Android application and answered any participants' questions. Following this, we sent the participant a link to install the application and a web-based guide describing the functions and controls of the application. Week one of participant recruitment was dedicated to data collection to build an initial model. The agent generated and sent auto-responses during week two using the participant's personalized attentiveness model trained using week 1 data. The application also sent daily questionnaires during week two. Participants were paid 30 USD for participating and completing the study. It is worth noting that the study took place between September to December 2020, when most organizations and universities were operating remotely due to the COVID-19 pandemic. These circumstances may have impacted our study results, as we will discuss later in this paper.

Ethical Considerations. The Institutional review board (IRB) of the university approved the study. During the briefing, we disclosed all the data collected by the application and the permission the application needs to be able to function to the participants. This information was also made available through the study web-page[6] sent out in an email after the briefing. As mentioned in section 3.2, the agent did not send any text message or contact information to the remote server.

[6]https://people.cs.pitt.edu/~pranut/messaging_study/index.html

5.1 Application interface

On the first launch of the application, participants had to enter details such as name, gender, and age. Following this, the application presented the consent form describing the purpose of the study. The main screen had buttons to start and stop the background services and an options pane to customize aspects of the application. Upon hitting the 'start' button for the first time, the application prompted the participants to label some locations of interest. They were informed that these location labels would be used for prediction and could also be shared in auto-responses.

The server kept track of when the application was started and stopped and alerted the participant if the application was stopped or crashed for more than 6 hours. Upon stopping the application, all data collection was ceased, and the agent stopped sending auto-responses. At the end of the day, around 9:00 PM, the application generated a notification asking the participant to complete a daily questionnaire asking for their feedback on the use of the agent. This was also available within the application in case the participant accidentally dismissed the notification or wanted to take the questionnaire earlier in the day. Participants could also take the questionnaire multiple times a day, and only unevaluated auto-responses were shown to them. All participants used the option to start the questionnaire from the application, sometimes even multiple times a day.

After two weeks, participants were sent an end-of-study survey within the application, which consisted of general questions related to the overall perception of auto-responses. The survey and the daily questionnaire responses guided the semi-structured end-of-study interview, which lasted about 45 minutes on average.

5.2 Participants

We reached saturation in terms of new high-level findings after around 12 interviews, and at that point, we stopped recruiting. In total, we recruited 14 participants. One participant could not run the application on their phone and had to withdraw after one week. After the briefing, another participant withdrew from the study due to being uncomfortable with sending auto-responses to their contacts. We discarded any collected data for these two participants and only presented the analysis results of the remaining 12 participants. In terms of the demographics of our participants, six were in the age group 18-24, three in the group 24-34, two in the group 35-44, and one in the group 55-64—seven of our participants identified as female, and five identified as male.

5.3 Analysis

The primary researcher remotely conducted the interviews with all participants, which were audio/video recorded. The recorded interviews were transcribed with the built-in transcription of the recording software and further fixed by the primary researcher. We performed inductive thematic analysis on interview transcripts [7] and used Nvivo software for creating and categorizing codes. The primary researcher developed the initial set of codes from half (six) of the interview transcripts, which were then improved upon and categorized into themes and sub-themes during multiple discussion sessions among the research team. Another researcher not part of this project's research team coded one of the interview transcripts. We achieved a Kappa value of 0.813 after performing a reliability analysis. Given the high level of agreement, the primary researcher coded the remaining six interview transcripts.

There were 105 initial codes such as "customizing: contact-blocking", "perception of noise value", and "usefulness for family". Upon iterating and refining these nodes, some nodes were split and recategorized. For instance, we split the "perception of noise value" code into two parts: "perceived utility of noise value" and "interpretation of noise value" categories. From these final set of nodes, we identified 16 first-level categories such as "interpretation", "customization", and "agent accuracy". Through rounds of discussion between the research team, we identified four major themes: "varying preferences related to agent function", "effect of misclassifications", "understanding of the agent and appropriation", and "utility of auto-response information type".

6 RESULTS

During the two-week agent deployment, 310 auto-responses were sent ($\mu = 25.333, \sigma = 16.036$) with a minimum of 6 for P2 and a maximum of 61 for P10. The most common auto-response type was the phone-usage ("*Laila has not been using her phone for a while*"), which was sent 86 times (27.74% of all auto-responses). We expected this since phone usage was in the top features for multiple personalized models similar to prior works [26, 27]. The overall accuracy was 70%, the false-positive rate was 0.21, and the false-negative rate was 0.55. We discuss these metrics in more detail in Section 7.1.1.

Next, we discuss the major themes emerging from our interview data. The overall response from our participants about the agent and auto-responses was positive, with participants noting less obligation and pressure to respond and to explain their delayed responses. While half of our participants reported less engagement with their phones, which was the agent's goal, it was highly context-dependent. Factors such as the message's urgency, strength and nature of social relationships, and the format and content of auto-response messages all played a role in defining how beneficial the agent was for the users. The type of information that the agent shared, in particular, was an important consideration, as our participants noted that its misinterpretations could be consequential. Additionally, there were indications of behavior change related to device and agent usage arising from the understanding of the agent function and effort to fix mistakes made by the agent. We expand on each of these in this section.

6.1 An auto-response agent can be a useful tool to communicate unavailability

Multiple participants reported various perceived benefits of using the agent and auto-responses, as we detail below.

6.1.1 Agent reduced pressure and obligation to respond. Overall, participants found the agent useful in reducing their attention to their phones. We observed an average of *5 minutes* increase in time to attend to new incoming messages among our participants; i.e., in the first week of study, when there were no auto-responses, they took an average of *18 minutes* to attend to their messages. In contrast, in the second week, when the agent started sending auto-responses, they took an average of *23 minutes*. P4 mentioned that they felt less pressure to check their phone and focused better on their tasks. *P4: "It really put less pressure on me to have to check my phone and my messages all the time to just make sure that people knew I was okay and receiving them, it kind of took that off my plate, and I could be more focused on what I was doing in the moment and then at night or later in the day kind of check back to see if messages required more meaningful response from myself, but oftentimes I could just leave it at that, i.e., auto-response. So I really enjoyed it"*. Indeed, our data confirmed that P4 took significantly more time to attend to incoming messages with the presence of the auto-response, from 8.5 minutes in the first week (no auto-responses) to 19 minutes in the second week (with auto-responses). Although, it could be related to their unique situation described in the next section. Nevertheless, they ascribed lower engagement to the use of the agent.

6.1.2 Agent can help stay focused on important tasks. Beyond the general utility of the agent, we learned that there could be certain situations where auto-responses were particularly useful for some participants. For example, four of our participants felt that auto-responses would be helpful while driving, *P7: "A big bad habit I have is that when I'm at stoplights, I'll check my phone. With the auto-responses, I did not do that"*. Similarly, P4 mentioned being on a trip when auto-responses started (during the second week), *P4: "I was in a unique position because when it started up the auto-responses, I was on a long 10-hour road trip. So it's really helpful not to kind of have to worry about responding to people knowing that the app would respond for me, and it did"*.

Another participant (P11) brought up the usefulness of auto-responses while studying, as messages can be distracting during that time, *P11: "They were especially useful when studying because I*

like to put my phone away and Yeah, I guess, like the biggest drive to pick up that phone is to make sure no one has contacted me". Another unique situation brought up by P11 was when they were at a doctor's appointment, *P11: "One time it was useful was when I was in a doctor's appointment. First thing in the morning on my birthday, and there were a bunch of people texting me because it was my birthday. And I was like, well, I'm at the doctor for an hour".*

6.1.3 Agent reduced the need to explain unavailability. Six participants indicated that they had to provide fewer explanations when the agent sent an auto-response. P2 attributed this to an accurate representation of their unavailability state in auto-responses, *P2: "Oh, because it was already laid out for me as to what was happening and why I wasn't available during that time".* Similarly, P13 described how they felt that agent explanations were sufficient to justify delays, *P13: "Before I used that application and I was away from my phone. I would always get from the other party like where are you, what are you doing, how come you didn't message me back. And then I would have to sit there and just, you know, lengthily explain what I was doing. That's why I felt those auto-responses were helpful".*

6.2 An auto-response agent is more useful in some situations

We identified multiple factors influencing how participants felt about the agent and auto-responses in our analysis.

6.2.1 Urgent vs Non-urgent messages. Our participants reported variations in the usefulness of auto-responses based on the urgency of the messages. Out of the six participants who brought up urgent/time-sensitive incoming messages, three felt auto-responses were not helpful for urgent messages, while the other three felt they were. The participants who preferred auto-responses in urgent situations gave reasons such as stronger emotions linked to urgent messages and making the sender aware so that they can reach out to someone else, *P9: "When it's someone texting when it's urgent or important, then I'd really want them to get an auto-response, just so they know what's going on. I think that would be really useful because if they know that you're not available or something, then they could reach out to someone else".* While participants who preferred not sending auto-responses in urgent situations felt that they needed to handle those situations themselves, *P4: "usually those (urgent) messages in the nature of my work on campus are more pointed towards me and are more time-sensitive. I guess that's the only reason".* This finding falls in line with previous work that surveyed people on their perception of sharing contextual information, confirming that urgency matters and has varying implications on agent usefulness for different people [28].

6.2.2 Agent's personality and its content representation. While most participants felt that the tone and framing of the auto-responses were fine, i.e., not too formal or casual, there was a mixed response as to whether they would like auto-responses to sound like them or take on an independent agent personality. Four of our participants felt that auto-responses sounding like them would improve their acceptance for their contacts, *P10: "The person who gets those auto-responses will believe that these responses are from me".* Further, seven of our participants also wanted to customize some aspect of the auto-responses by adding a personal touch, *P1: "Personalization messages are really big for me. I really like value using my own voice. And so I would definitely want to see that".* Other participants preferred auto-responses not to sound like they would, to be distinctive from their own responses, and not confuse their contacts. P8 elaborated on this *P8: "I had a friend who used voicemail with, "Hello, are you there?". It sounded like she was actually picking up, and that always drove me nuts because I would try and actually talk to her. I feel like if it (the agent) sounded more like me, it might get more responses unnecessarily".*

6.2.3 Usefulness for different contact groups. Another avenue of varied response was the utility of auto-responses for different contact groups. The qualitative design of the study allowed us to inquire about the perception of the agent for more distinctive contact groups, unlike previous survey-based studies, which were limited to two or three coarse groups [28, 31]. In addition, to close vs. distant groups, our participants noted the relevance of more fine groups such as higher-authority figures (e.g., boss and advisor), family, friends, coworkers, and even personalized contact types (e.g., their doctors' offices, special-needs contact) as we discuss below.

Four participants mentioned agent usefulness towards an interesting contact group: a higher-authority group such as a boss, advisor, or professor. P9 emphasized the usefulness of auto-responses for their boss, *P9: "More important people like, say like a boss or someone that you always want to be more responsive to, you know, or keep them more in the loop".*

In terms of close vs. distant contacts, our participants again had mixed perceptions of auto-response utility for these contact types. Some participants did not feel comfortable sending auto-responses to infrequent contacts, *P1: "Basically there are two kinds of people who contact me: people whom I think of as close friends and people who are acquaintances, or maybe who I don't know at all. And so for people whom I don't know at all or not very well or like acquaintances, I definitely don't want auto-responses to go to them because I don't feel the need to tell them anything about me until I've decided whether or not I want to engage".* In contrast, some participants specifically found auto-responses useful for contacts they did not engage with frequently, such as distant family, *P2: "I have a cousin that's in [redacted] right now. It would have been really useful for her because there were times where I can't always get to her, and I hear sometimes I'm just entirely too busy to respond to her".*

Similarly, some participants felt that close contacts already knew about their availability and schedules, making the auto-responses less helpful, *P8: "I think people whom I text very frequently, it was less useful. Like if people are already fairly aware of my schedule and (they) can kind of anticipate. It's not necessarily providing any new information".* Two participants mentioned that while auto-responses were less useful for frequent contacts in general, they were helpful for their families, *P4: "All my family really liked it. I'd say my parents probably benefited the most from it while I was away on vacation. They enjoyed being able to "keep up with me" but know that I was safe and would respond at a later point. And then when we were driving it auto responded to my cousin whose house we were staying at, and she found that helpful as well".* On the other hand, personal situations also made auto-responses to close family members such as parents not useful for some participants, *P2: "There might be people who just don't want the auto-responses*

to go to like my mom because she might actually need something at that point in time. She's more of a person that I need to get to right away because of health issues".

There were also instances brought up by participants discussing contact types that are more specific to them. For instance, P4 mentioned how auto-responses could be confusing to a special-needs person they interact with through messaging, *P4: "One of the individuals has special needs so, with her, I have to be very direct and blunt with the messages. So I just didn't want to confuse her".* Similarly, P9 mentioned wanting auto-responses to their doctor's office even if they are not on their contacts list to inform them of their unavailability.

6.3 Perception and interpretation of information shared by the agent

Our participants evaluated 263 auto-responses in the daily questionnaire. In terms of mean ratings for different categories listed in Table 1, we did not observe any significant difference concerning the usefulness and comfort of these auto-response types. Although, there were implications related to the content of the auto-responses, as we detail below.

6.3.1 Is the reason convincing? Our participants discussed multiple factors as to what constituted a good auto-response. One of them is that the reason shared has to be convincing, *P10: "It's about what the information is, what the reason is, it could be very long, but [if] there is no specific reason, or there is no convincing reason, then I don't think the other person would be very friendly to you".* P8 had a similar opinion and elaborated using an example auto-response that the agent sent to their contact, *P8: "The ambient noise one, I'm like, just playing music in my own house. I don't think [it] makes me less likely to respond".* Similarly, P9 felt that silent environment (noise value) auto-response may not be indicative of unavailability in most cases, *P9: "I feel like there's a lot of cases where you're in a silent environment, but you're still available to respond. You're just like, say, in your room just like reading a book or whatever, like you're not necessarily you know focused on something very important or like, if you're in the library studying, Well, I guess, in that case, then it [would] be different but yeah I think there's just too many cases with that when that wouldn't be a good response".*

Most participants liked the auto-response sharing phone usage. P1 and P9 also noted the reduced privacy risk from sharing phone usage information compared to other user-state information such as location, calendar, and app usage. *P1: "Because it doesn't really tell you what I'm doing, it tells you what I'm not doing, and since what I'm not doing is relevant to their needs, then it makes a little more sense in terms of alignment to me".* P9 expressed a similar sentiment, *P9: "I think that might be one of the best ones just because like you know it's like general, It doesn't give too much information, but it gives enough to infer to the other person that he is not using his phone so he's probably just not available".* Similarly, ringer mode had a positive reception, *P9: "I thought that was really useful because I feel like when my phone is on silent mode, I probably won't want to respond, so I think that's always a good time to send an auto-response".*

6.3.2 Privacy implications of sharing app usage information. There was an overwhelmingly negative response to sharing app usage information in an auto-response even though the agent was sharing the category of application (e.g., productivity, communication, and entertainment app) rather than the exact name of the application last used. P1 and P12 mentioned that they were not comfortable sharing app usage due to the potential of sharing highly personal usage information. *P1: "I basically almost never want them to know which apps I'm using on my phone because if I want to look at [inappropriate content]. That's my own thing, not just good, but yeah, I definitely don't love that."* Sharing app usage was not always perceived negatively. P10 pointed out a stark variation in their perception of sharing app usage based on the type of app category shared in the auto-response. One of those auto-responses shared that they were last using an educational app, whereas the other said they were last using an entertainment app, *P10 (for education app): "I think this reason tells them that he's working on some project or something, educational and should not be disturbed."* Whereas, when the agent shares 'entertainment app', *P10 (entertainment app): "They might think like he's ignoring me but he's also using an entertainment app."*

6.3.3 Speculative and misinterpreted context. P7 felt that in addition to being convincing, the auto-responses should also not leave room for speculation, *P7: "I like the ones that are just a little bit shorter and clearish. I don't want [the sender] reading too much into it".* As noted in Section 6.2.3, there was also some variation in preferences related to auto-response information concerning different contact groups. In general, while sharing that the user has a calendar event had a positive reception from most participants, P11 felt that sharing that they have a calendar event may lead to speculations and more questions, *P11: "I thought the calendar one was kind of unnecessary. It just kind of makes it begs like oh, what is the event or like begs more questions than a simple like not able to respond".* P13 mentioned a similar issue with sharing 'not at usual location'. The agent picks this auto-response when the user is in an unlabeled location that affects their availability, *P13: "They want to know what's going on and where am I, that's what they'll be thinking".*

P5 pointed out the ambiguity of sharing light value, *P5: "Oh, the low light one is kind of not useful. For me nor for them just because it could apply that my phone was just facing down".* P9 recalled that their contacts found the dark and silent environment auto-responses 'creepy' and raised concern for them, *P9: "A couple of people thought that some of the responses were overly specific or like, you know, kind of creepy. I think they had mentioned the light level one and the silent environment one".* Similarly, for most participants, noise value auto-responses raised concerns about being misinterpreted due to their potential locality inference. P2 pointed out an example of this. They had an auto-response sent saying that they were in a 'noisy environment' whereas they were in bed, sleeping. While discussing this auto-response, they recalled that it was probably due to their room's loud air conditioning, which their phone's mic might have picked up. So even though the information in the auto-response was technically correct, the auto-response itself was misconstrued by their contact, *"P2: Didn't think it was appropriate since it sounds like I was at a party and I wasn't, and that one was to my dad. So he's probably like, where is she?".* P7 and P11 raised similar concerns regarding noise value: *P7: "When I think of a noisy environment. I think it's like crowds, and if it's going to coworkers and my parents, that's not really the image I want to put forth". P11:*

"I feel like it gives the illusion that I'm in like it begs like where are they that's noisy". Prior survey-based studies which evaluated the comfort of sharing noise data did not report on the potential locality inference arising from sharing noise value, making this finding interesting [28, 31].

Another observation that P9 noted was related to potential long-term effects or assumptions based on the information shared in the auto-response. For instance, the participant mentioned that an auto-response sharing 'not responsive at this time of the day' might prevent contacts from initiating the conversation at that time in the future, *P9: "They just assume like yeah this, he doesn't want to be disturbed this time of day and I'll just hold off for later"*. Although, P10 felt that this auto-response was particularly useful for them since there were times in the week when they did not respond to messages, *P10: "I think this will clear up the fact that this is not a good time to text because anyway he won't text you back"*.

6.4 Behavior change related to the agent and device usage

We were interested in identifying how the agent as a whole and auto-responses influence a change in how our participants were using their devices. Our findings reveal both positive and negative aspects of using the agent to handle communication.

6.4.1 Reduction in device engagement when the agent works as expected. As described in Section 3, the main goal in the design of the agent was to reduce device engagement by enabling the agent to handle incoming communication. Thus, understanding the effect of the agent and auto-responses on device engagement was one of our focuses for the evaluation. Overall, half of our participants reported *reduced* device engagement with the use of the agent, while the other half reported an *increase*. Most participants initially reported *increased* engagement with their device due to feelings of curiosity regarding the tool's novel features. However, perceptions of engagement *decreased* in the latter part of the study, as indicated by the following quotes. *P7: "At first, whenever it first started sending the auto-responses, I checked like "oh did it send an auto-response cool!". After that initial checking of messages, I stopped checking them as much because I felt like it could explain if I was available or not available"*. P2 and P11 also felt that auto-responses would help them take a break from their device, *P11: "At times I thought it was actually helpful to not feel the need to be connected to my phone because of that (auto) response. So I thought that was good"*.

6.4.2 Mistakes of the agent can increase users' effort and decrease their sense of control. Mistakes by the agent, such as sending an auto-response when it was not needed, resulted in an increased effort by participants to provide explanations to repair a social situation, *P11: "It would send a response, and then two seconds later, I would see it and have to explain that. That (it) was just a false alarm"*.

Reasons for misclassifications. We computed the overall false-positive rate (FPR) and false-negative rate (FNR) based on our logged data of (1) when a user received a message, (2) whether they attended to it within the expected response threshold (7.2 minutes), and (3) whether the agent sent an auto-response. The computed FPR of 0.03 was quite low due to multiple factors such as sending auto-responses only for known contacts, auto-response delay setting, contact interval setting, and contact blocking. Without these filters, the FPR would have been 0.21, which is still not very high and is comparable to the results of prior studies [26, 50]. The false-negative rate was 0.55, which was much higher than FPR. However, even though the FPR was lower than FNR, our participants' perceptions of these misclassifications differed. All of our participants reported experiencing false positives, whereas only four mentioned experiencing false negatives, with only P2 and P5 reporting a high frequency of missed opportunities to send auto-responses. This indicates that most participants were more sensitive to the agent responding when not needed than not responding when it should.

What can be contributing to participants' perception of false positives incidences is how unavailability is defined, i.e., how long of a delay in responding is acceptable to send an auto-response? As discussed in Section 3.1.1, we used a threshold of 7.2 minutes for labeling attentiveness based on prior works, which used the average median time in their respective datasets for evaluation [26, 50]. Multiple participants felt that not attending to a message within 7 minutes does not warrant an auto-response, *P8: "I'd say probably somewhere between 20 and 30 (minutes) is fast enough to not warrant an auto-response"*. Another reason could be the particular circumstances of our study, which took place during the work-from-home and stay-at-home period due to the COVID-19 pandemic. Multiple participants reported unusually greater attention to their devices due to classes and work taking place remotely from home, making it harder for the agent to detect instances of unavailability (resulting in greater FNR as well), *P8: "I think just because of the way my work in school is, I'm online most of the time, or, you know, I'm within ready access of my phone most of the time when I'm awake. And if I'm not, it's like I'm on a certain kind of call or running or driving. I don't think there were a lot of opportunities for it to send one where it didn't"*.

Sharing irrelevant or unwanted context also resulted in participants putting in the effort to explain that context while also feeling more obligated to respond earlier than they would have. P13 described a situation where the agent sent an auto-response that they were listening to music which caused them to respond immediately, explaining themselves, *P13: "I was using an app, and I was playing music, and I saw an auto-response went out. I immediately got off the app and went into messenger. And I told my mom. I'm like, Hey, I'm available to talk to you. I'm just, you know, listening to music"*. Similarly, P14 recalled when the agent sent an auto-response saying they were last using a communications app, making them respond quicker than they would have since the auto-response indicated they had been messaging recently. As mentioned in Section 3.1.4, our approach utilizes correlation with the availability state rather than causation. This can lead to sharing irrelevant context that the user or their contacts may not link to unavailability and may lead to increased effort and loss of control over the interaction.

6.4.3 Uncertainty and lack of understanding of agent function negatively affects its usage. As mentioned earlier, about half the participants reported increased device engagement due to the use of the agent. Unfamiliarity with how the agent functioned was a significant reason for this behavior change. Some participants reported checking their phones more often to prevent an auto-response from

going out. For instance, P11 suspected that not using their phone for a certain amount of time was a trigger for an auto-response, *P11: "Sometimes I would check it even more frequently because I didn't want that auto-response to go through"*. Similarly, P8 described checking their phone more often in anticipation of an important message that they did not want the agent to respond to. This was another example of increased use due to the belief that not using the phone will trigger an auto-response, *P8: "I was checking more constantly because I was worried that it would send him (landlord) something, and I'd have to explain it. We don't talk a lot. So I think it would be kind of weird"*.

On the other hand, P2's experience with the agent auto-responses was quite the opposite. P2 reported issues where they expected the agent to auto-respond, but it did not. They explained that they would often go into messaging app to check if the agent sent an auto-response upon getting a message. If not, they would respond themselves, reducing the agent's utility and increasing their device engagement, *P2: "My engagement would have probably went down. I don't want to engage with my phone as much. I was trying to practice that a little bit in terms by leaving my phone away from me for a bit, but then I will pick it back up If it was like five minutes and I didn't see anything (auto-response)"*. This behavior projects the gap between understanding of the agent function and expectations. Since the agent learns from messaging behavior of its users, opening a message within 5 minutes of arriving, P2 was inadvertently *attending* to it. This would cause the agent to prevent sending an auto-response (if the delay setting is greater than 5 minutes) while also learning that the user is available in that context. Some participants also tried to align the agent's behavior based on their understanding of the agent, e.g., by turning off the app, *P8: "I knew, I was going somewhere (and) the algorithm would notice that you know, doing something different, (or) at a different location and I didn't want it to notice that. I didn't want it to send auto-responses (at that location)"*.

The presence of the messaging agent also affected some participants' contacts. For instance, P12 reported that their contact sent multiple messages upon getting an auto-response, *P12: "A lot of times they said that when they messaged me like they weren't sure if I got it or not. They messaged me almost three times the same message. I don't think they were 100% if [I] got the message or if it went through. I think they felt like sometimes it was blocking them or something"*. P5 had to stop their app because some of their contacts were trying to trigger agent response out of curiosity, *P5: "I kind of had to stop it (app). Just because I know some people were starting to mess with the app and, like you know, purposely responding to stuff just to see what would happen. And like I think it can get a bit too abusive with it"*.

7 DISCUSSION

Intelligent Personal Assistants or IPAs are designed to assist users in their tasks by utilizing contextual information available to them through sensors [16, 36]. We are starting to see IPAs take on more proactive tasks without requiring initiation by their users [65]. Our work on the availability management agent advances our understanding of facilitating awareness in mobile messaging through a virtual assistant. We present design implications from our findings, followed by the limitations of this study and the direction of future work.

7.1 Design Implications

7.1.1 Need for more cooperative human-agent interaction. As discussed in Section 6.4.2, mistakes made by the agent decreased users' sense of control and increased the required effort to explain agent actions to their social contacts. These mistakes or misclassifications, as reported by the participants, took three forms: (1) false positives - situations where the user was available to respond, but an auto-response was still sent; (2) false negatives - situations where an auto-response was expected to be sent but was not; and (3) auto-responses shared irrelevant information to explain users' unavailability. While the model's intelligence can continually be improved as more data becomes available, as we explain below, there are also cases specific to unforeseen circumstances such as the response from user's contacts. Here, we argue that intelligent agents need to be designed more as human partners, and their design should support feedback from their users.

Learning from the user. In addition to retraining the model daily, as mentioned in Section 4.3, to reduce potential false positives, we introduced a delay setting in the CAR application to allow users to set up a delay before the agent sends an auto-response; however, we restricted the maximum delay setting to be 7 minutes to conform to the threshold used for labeling (7.2 minutes). Nine participants adjusted this delay setting, with 6 participants changing the delay at least twice. The general reason given by the participants for increasing the delay was to give them a greater chance to respond if they were to become available. Feedback from contacts also affected how participants adjusted the delay setting. For instance, P7 reduced the delay as it caused their contact to misread the agent response as theirs, *P7: "[My boyfriend] was just like, you know, I really don't like it when there's such a delay between the auto-responses. It makes me expect that you actually responded to my message"*.

These interactions with the agent can provide helpful context about user preferences to the agent [18]. The agent can link each user interaction within its setting as a learning opportunity about the user. The agent uses past messaging behavior to create an availability model for their users. We learned that in some cases, users might be interested in pushing an auto-response even when the agent predicts them to be available correctly. Providing adequate controls to users in such cases while also allowing the agent to learn the user-specific context for future incidences can improve the human-agent interaction. This was also highly reflected in users' feedback about what context the agent should share. Most participants wanted to customize or add a personal touch to auto-responses. Allowing users to link or change auto-response presentations to specific contexts can help with improving the perception of the agent while at the same time reducing ambiguity associated with specific sensor values. For instance, a user can be allowed to change the term 'noisy' to another term more applicable to their context, such as 'busy', as demonstrated in the case of P8, who described wanting to change the noise value phrasing, *P8: "It wasn't really telling them anything helpful about where I might be, um, maybe if the person knew like noisy environment equals busy. Maybe if I were like a construction worker or something, but I'm definitely not. It was kind of unhelpful information in that context"*.

7.1.2 Intelligent Personal Assistants can teach their users about AI by being transparent. People typically appropriate technology to

suit their needs [52]. In this study, as discussed in Section 6.4.3, our participants tried to use their exposure to the agent to understand how the agent was functioning and, in some cases, reverse engineer the behavior of the agent by altering their own behavior (e.g., turning off the agent when moving to a new location) or trial and error to decode the agents' behavior some of which resulted in increased device engagement contrary to the purpose of the agent. This demonstrates a significant opportunity to design intelligent personal assistants as a medium to teach users about intelligent algorithms. Previous literature on agent design has emphasized the importance of making AI actions and machine learning predictions explainable as well as transparent to their users, which not only helps with improved system understanding [23] but can also help build trust in the system use [1]. Thus, for the design of the communication agent, it becomes essential to make the learned model open to the user and provide clarity towards agent actions and learning opportunities about the agent behavior. Users can interact with the agent to ask questions about the agents' behavior in different contexts. Improved agent understanding and the addition of proper controls such as modifying or removing any learned context (e.g., location) from the model can help users make more meaningful appropriations of the agent and gain a higher level of awareness about intelligent agents.

7.2 Limitations and Future work

Our study was affected by the COVID-19 pandemic. As discussed earlier, our participants reported that they had higher than usual access to their phones due to working from home during the pandemic. Increased access and quick attendance to notifications limited the opportunities by the agent to auto-respond and could have affected the results of our study. Further, the agent operated in conjunction with the regular availability indicators in messaging applications. We did not ask our participants to disable these indicators since we wanted to support multiple messaging applications for this study. It would have required effort on the part of the participants to find and disable these indicators, which might not even be possible for all available applications. This could have affected our results as well. However, we did not receive any feedback from our participants on how auto-responses worked in combination with these indicators, which might be helpful to explore in the future.

As discussed in Section 6.2.1 some participants preferred auto-responses in urgent situations while others preferred to handle urgent situations themselves. Further, there might be other situations where the agent's action could be undesired. For instance, P8 recalled a situation, *P8: "it's also sort of unpredictable, what kinds of responses warrant an auto-response versus not. But, um, I was asking someone for a letter of reference, who I primarily contact through text, and that person responded to me saying, I'm going to need a little extra time one of my parents died. And that's definitely the kind of message where I would want to respond personally and have some time to think about it. And so if the app is doing anything with message content, I would say like maybe scan for the message being kind of serious"*. While we developed a detection mechanism for the agent which prevented sampling of meta-messages such as reactions in Signal[7] app, we did not parse message content to detect emoticons or end-of-conversation behaviors [20] or messages such as *"goodbye", and "talk to you later"* [34]. This understanding of conversation will help prevent agent actions in these situations, potentially improving agent utility. Although parsing text messages can have privacy implications, and further research is needed to understand the balance between getting more context from conversations and user expectations of their privacy.

Finally, our participants used the agent for two weeks, within which the auto-responses were sent only for the second week. As noted in Section 6.4.1, users initially reported increased engagement due to curiosity about how the agent functioned. However, we might see more habituation and more considerable decreases in device utilization once users are comfortable with the agent. A more extended study can provide quantitative evidence regarding how beneficial the agent can be for its users. In addition to this, it would be interesting to see how users themselves and their contacts start sense-making of the information shared by the agent in the long term as these might further raise privacy concerns [33]. A long-term study will also help us understand whether over-trust and over-reliance could be a potential issue for this agent type [14, 29]. Through personalization, as the agent gets better at its task, users may start to rely on it even more, potentially impacting how they utilize the agent and engage with their contacts.

8 CONCLUSION

In this work, we reported on the design, implementation, and evaluation of a messaging agent called CAR that uses context captured from its user's smartphone to detect and communicate their unavailability to message senders. This work contributes to the field of designing interactive systems by describing the design of a proactive agent and a unique approach to sharing mid-level sensor-based context through auto-responses. The results of the agent's evaluation through a two-week user study provided important insights into the value of such an agent to improve users' experience in message-based communication and how personal and situational factors play a crucial role in the design of such agents. Notably, the results point towards the increasing importance of personalization and improving cooperation between the user and the agent by improving user understanding of the agent's actions while also improving the agent's learning from the users' actions and interactions. This work forms the next step towards realizing the potential of virtual assistants to handle communication for their users effectively.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under award CNS–1814866. We would also like to thank our participants for their contribution to this study.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3173574.3174156

[7] https://signal.org/en/

[2] Daniel Avrahami and Scott E. Hudson. 2006. Responsiveness in Instant Messaging: Predictive Models Supporting Inter-Personal Communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 731–740. https://doi.org/10.1145/1124772.1124881

[3] Brian P Bailey, Joseph A Konstan, and John V Carlis. 2001. The Effects of Interruptions on Task Performance, Annoyance, and Anxiety in the User Interface.. In *Human-Computer Interaction: IFIP TC13 International Conference on Human-Computer Interaction, July 9-13* (Tokyo, Japan) *(INTERACT '01)*. IOS Press, 593–601. https://doi.org/10.1145/3025453.3025946

[4] João Balsa, Isa Félix, Ana Paula Cláudio, Maria Beatriz Carmo, Isabel Costa e Silva, Ana Guerreiro, Maria Guedes, Adriana Henriques, and Mara Pereira Guerreiro. 2020. Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. *Journal of Medical Systems* 44, 7 (2020), 1–12.

[5] Linas Baltrunas, Karen Church, Alexandros Karatzoglou, and Nuria Oliver. 2015. Frappe: Understanding the Usage and Perception of Mobile App Recommendations In-The-Wild. https://doi.org/10.48550/ARXIV.1505.03014

[6] James "Bo" Begole, Nicholas E. Matsakis, and John C. Tang. 2004. Lilsys: Sensing Unavailability. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 511–514. https://doi.org/10.1145/1031607.1031691

[7] V. Braun and V. Clarke. 2013. *Successful Qualitative Research: A Practical Guide for Beginners*. SAGE Publications. https://books.google.com/books?id=nYMQAgAAQBAJ

[8] Andreas Buchenscheit, Bastian Könings, Andreas Neubert, Florian Schaub, Matthias Schneider, and Frank Kargl. 2014. Privacy Implications of Presence Sharing in Mobile Messaging Applications. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia* (Melbourne, Victoria, Australia) *(MUM '14)*. Association for Computing Machinery, New York, NY, USA, 20–29. https://doi.org/10.1145/2677972.2677980

[9] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[10] Hyunsung Cho, Jinyoung Oh, Juho Kim, and Sung-Ju Lee. 2020. I Share, You Care: Private Status Sharing and Sender-Controlled Notifications in Mobile Instant Messaging. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 34 (may 2020), 25 pages. https://doi.org/10.1145/3392839

[11] Karen Church and Rodrigo de Oliveira. 2013. What's up with Whatsapp? Comparing Mobile Instant Messaging Behaviors with Traditional SMS. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Munich, Germany) *(MobileHCI '13)*. Association for Computing Machinery, New York, NY, USA, 352–361. https://doi.org/10.1145/2493190.2493225

[12] Drew Clinkenbeard, Jennifer Clinkenbeard, Guillaume Faddoul, Heejung Kang, Sean Mayes, Alp Toygar, and Samir Chatterjee. 2014. What's Your 2%? A Pilot Study for Encouraging Physical Activity Using Persuasive Video and Social Media. In *Persuasive Technology*, Anna Spagnolli, Luca Chittaro, and Luciano Gamberini (Eds.). Springer International Publishing, Cham, 43–55.

[13] Justin Cranshaw, Emad Elwany, Todd Newman, Rafal Kocielnik, Bowen Yu, Sandeep Soni, Jaime Teevan, and Andrés Monroy-Hernández. 2017. Calendar.Help: Designing a Workflow-Based Scheduling Agent with Humans in the Loop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2382–2393. https://doi.org/10.1145/3025453.3025780

[14] Mary L Cummings. 2017. Automation bias in intelligent time critical decision support systems. In *Decision Making in Aviation*. Routledge, 289–294.

[15] Ed Cutrell, Mary Czerwinski, and Eric Horvitz. 2001. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In *INTERACT 2001*. IOS Press, 263–269. https://www.microsoft.com/en-us/research/publication/notification-disruption-and-memory-effects-of-messaging-interruptions-on-memory-and-performance/

[16] Allan de Barcelos Silva, Marcio Miguel Gomes, Cristiano André da Costa, Rodrigo da Rosa Righi, Jorge Luis Victoria Barbosa, Gustavo Pessin, Geert De Doncker, and Gustavo Federizzi. 2020. Intelligent personal assistants: A systematic literature review. *Expert Systems with Applications* 147 (2020), 113193.

[17] María del Carmen Rodríguez-Hernández and Sergio Ilarri. 2021. AI-based mobile context-aware recommender systems from an information management perspective: Progress and directions. *Knowledge-Based Systems* 215 (2021), 106740.

[18] Paul Dourish. 2004. What we talk about when we talk about context. *Personal and ubiquitous computing* 8, 1 (2004), 19–30.

[19] Denzil Ferreira, Vassilis Kostakos, and Anind K Dey. 2015. AWARE: mobile context instrumentation framework. *Frontiers in ICT* 2 (2015), 6.

[20] Rebecca Grinter and Margery Eldridge. 2003. Wan2tlk? Everyday Text Messaging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 441–448. https://doi.org/10.1145/642611.642688

[21] Ted Grover, Kael Rowan, Jina Suh, Daniel McDuff, and Mary Czerwinski. 2020. Design and Evaluation of Intelligent Agent Prototypes for Assistance with Focus and Productivity at Work. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 390–400. https://doi.org/10.1145/3377325.3377507

[22] Jeff Hancock, Jeremy Birnholtz, Natalya Bazarova, Jamie Guillory, Josh Perlin, and Barrett Amos. 2009. Butler Lies: Awareness, Deception and Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 517–526. https://doi.org/10.1145/1518701.1518782

[23] Steven R Haynes, Mark A Cohen, and Frank E Ritter. 2009. Designs for explaining intelligent agents. *International Journal of Human-Computer Studies* 67, 1 (2009), 90–110.

[24] Roberto Hoyle, Srijita Das, Apu Kapadia, Adam J. Lee, and Kami Vaniea. 2017. Was My Message Read? Privacy and Signaling on Facebook Messenger. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3838–3842. https://doi.org/10.1145/3025453.3025925

[25] Cheng-Kang Hsieh, Hongsuda Tangmunarunkit, Faisal Alquaddoomi, John Jenkins, Jinha Kang, Cameron Ketcham, Brent Longstaff, Joshua Selsky, Betta Dawson, Dallas Swendeman, Deborah Estrin, and Nithya Ramanathan. 2013. Lifestreams: A Modular Sense-Making Toolset for Identifying Important Patterns from Everyday Life. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems* (Roma, Italy) *(SenSys '13)*. Association for Computing Machinery, New York, NY, USA, Article 5, 13 pages. https://doi.org/10.1145/2517351.2517368

[26] Pranut Jain, Rosta Farzan, and Adam J. Lee. 2019. Adaptive Modelling of Attentiveness to Messaging: A Hybrid Approach. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP '19)*. Association for Computing Machinery, New York, NY, USA, 261–270. https://doi.org/10.1145/3320435.3320461

[27] Pranut Jain, Rosta Farzan, and Adam J. Lee. 2019. Are You There? Identifying Unavailability in Mobile Messaging. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312893

[28] Pranut Jain, Rosta Farzan, and Adam J Lee. 2021. Context-based Automated Responses of Unavailability in Mobile Messaging. *Computer Supported Cooperative Work (CSCW)* (2021), 1–43.

[29] Jason D. Johnson, Julian Sanchez, Arthur D. Fisk, and Wendy A. Rogers. 2004. Type of Automation Failure: The Effects on Trust and Reliance in Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 48, 18 (2004), 2163–2167. https://doi.org/10.1177/154193120404801807 arXiv:https://doi.org/10.1177/154193120404801807

[30] E. Kimani, K. Rowan, D. McDuff, M. Czerwinski, and G. Mark. 2019. A Conversational Agent in Support of Productivity and Wellbeing at Work. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–7. https://doi.org/10.1109/ACII.2019.8925488

[31] Johannes Knittel, Alireza Sahami Shirazi, Niels Henze, and Albrecht Schmidt. 2013. Utilizing Contextual Information for Mobile Communication. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 1371–1376. https://doi.org/10.1145/2468356.2468601

[32] Stefan Kopp, Mara Brandt, Hendrik Buschmeier, Katharina Cyra, Farina Freigang, Nicole Krämer, Franz Kummert, Christiane Opfermann, Karola Pitsch, Lars Schillingmann, Carolin Straßmann, Eduard Wall, and Ramin Yaghoubzadeh. 2018. Conversational Assistants for Elderly Users – The Importance of Socially Cooperative Dialogue. In *Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications co-located with the Federated AI Meeting* (Stockholm, Sweden), Elizabeth André, Timothy Bickmore, Stefanos Vrochidis, and Leo Wanner (Eds.), Vol. 2338. RWTH, 10–17. https://nbn-resolving.org/urn:nbn:de:0070-pub-29201667,https://pub.uni-bielefeld.de/record/2920166

[33] Albrecht Kurze, Andreas Bischof, Sören Totzauer, Michael Storz, Maximilian Eibl, Margot Brereton, and Arne Berger. 2020. *Guess the Data: Data Work to Understand How People Make Sense of and Use Simple Sensor Data from Homes*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376273

[34] Christine Lee. 2003. How does instant messaging affect interaction between the genders. *Stanford, CA: The Mercury Project for Instant Messaging Studies at Stanford University. Retrieved August* 11 (2003), 2006.

[35] Zilu Liang, Bernd Ploderer, Wanyu Liu, Yukiko Nagata, James Bailey, Lars Kulik, and Yuxuan Li. 2016. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Personal and Ubiquitous Computing* 20, 6 (2016), 985–1000.

[36] Yuting Liao, Jessica Vitak, Priya Kumar, Michael Zimmer, and Katherine Kritikos. 2019. Understanding the Role of Privacy and Trust in Intelligent Personal Assistant Adoption. In *Information in Contemporary Society*, Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin, and Bonnie Nardi (Eds.). Springer International Publishing, Cham, 102–113.

[37] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2, 1 (2020), 2522–5839.

[38] Deborah Lupton, Sarah Pink, Christine Heyes LaBond, and Shanti Sumartojo. 2018. Digital Traces in Context: Personal Data Contexts, Data Sense, and Self-Tracking Cycling. *International Journal of Communications Special Issue on Digital Traces in Context, Vol 12, 647-665* (2018).

[39] Lisa M. Mai, Rainer Freudenthaler, Frank M. Schneider, and Peter Vorderer. 2015. "I know you've seen it!" Individual and social factors for users' chatting behavior on Facebook. *Computers in Human Behavior* 49 (2015), 296 – 302. https://doi.org/10.1016/j.chb.2015.01.074

[40] Mehdi Mekni, Zakaria Baani, and Dalia Sulieman. 2020. A Smart Virtual Assistant for Students. In *Proceedings of the 3rd International Conference on Applications of Intelligent Systems* (Las Palmas de Gran Canaria, Spain) *(APPIS 2020)*. Association for Computing Machinery, New York, NY, USA, Article 15, 6 pages. https://doi.org/10.1145/3378184.3378199

[41] Mohammad Naiseh, Nan Jiang, Jianbing Ma, and Raian Ali. 2020. Personalising Explainable Recommendations: Literature and Conceptualisation. In *Trends and Innovations in Information Systems and Technologies*, Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis, Sandra Costanzo, Irena Orovic, and Fernando Moreira (Eds.). Springer International Publishing, Cham, 518–533.

[42] Harri Oinas-Kukkonen and Marja Harjumaa. 2009. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* 24, 1 (2009), 28.

[43] Bauyrzhan Ospan, Nawaz Khan, Juan Augusto, Mario Quinde, and Kenzhegali Nurgaliyev. 2018. Context aware virtual assistant with case-based conflict resolution in multi-user smart home environment. In *2018 international conference on computing and network communications (coconet)*. IEEE, 36–44.

[44] Motoyuki Ozeki, Shunichi Maeda, Kanako Obata, and Yuichi Nakamura. 2008. Virtual Assistant: An Artificial Agent for Enhancing Content Acquisition: How Ambient Media Elicit Information from Humans. In *Proceedings of the 1st ACM International Workshop on Semantic Ambient Media Experiences* (Vancouver, British Columbia, Canada) *(SAME '08)*. Association for Computing Machinery, New York, NY, USA, 75–82. https://doi.org/10.1145/1461912.1461927

[45] Motoyuki Ozeki, Shunichi Maeda, Kanako Obata, and Yuichi Nakamura. 2009. Virtual assistant: enhancing content acquisition by eliciting information from humans. *Multimedia Tools and Applications* 44, 3 (2009), 433–448.

[46] Lindsay C Page and Hunter Gehlbach. 2017. How an artificially intelligent virtual assistant helps students navigate the road to college. *AERA Open* 3, 4 (2017), 2332858417749220.

[47] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[48] Sobah Abbas Petersen, Jörg Cassens, Anders Kofod-Petersen, and Monica Divitini. 2008. To be or not to be aware: Reducing interruptions in pervasive awareness systems. In *2008 The Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. IEEE, 327–332.

[49] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 91 (sep 2017), 25 pages. https://doi.org/10.1145/3130956

[50] Martin Pielot, Rodrigo de Oliveira, Haewoon Kwak, and Nuria Oliver. 2014. Didn't You See My Message? Predicting Attentiveness to Mobile Instant Messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3319–3328. https://doi.org/10.1145/2556288.2556973

[51] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 825–836. https://doi.org/10.1145/2750858.2804252

[52] Ana Paula Retore and Leonelo Dell Anhol Almeida. 2019. Understanding Appropriation Through End-User Tailoring in Communication Systems: A Case Study on Slack and WhatsApp. In *Social Computing and Social Media. Design, Human Behavior and Analytics*, Gabriele Meiselwitz (Ed.). Springer International Publishing, Cham, 245–264.

[53] Stephanie Rosenthal, Anind K. Dey, and Manuela Veloso. 2011. Using Decision-Theoretic Experience Sampling to Build Personalized Mobile Phone Interruption Models. In *Proceedings of the 9th International Conference on Pervasive Computing* (San Francisco, USA) *(Pervasive'11)*. Springer-Verlag, Berlin, Heidelberg, 170–187.

[54] Antti Salovaara, Antti Lindqvist, Tero Hasu, and Jonna Häkkilä. 2011. The Phone Rings but the User Doesn't Answer: Unavailability in Mobile Communication. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 503–512. https://doi.org/10.1145/2037373.2037448

[55] Maxim Stepanov. 2020. Some stats about voice assistants. https://uxdesign.cc/some-stats-about-voice-assistants-1c292476584. Accessed: 2021-07-20.

[56] John C. Tang. 2007. Approaching and Leave-Taking: Negotiating Contact in Computer-Mediated Communication. *ACM Trans. Comput.-Hum. Interact.* 14, 1 (may 2007), 5–es. https://doi.org/10.1145/1229855.1229860

[57] Lauren Thomson, Adam J. Lee, and Rosta Farzan. 2018. Ephemeral Communication and Communication Places. In *Transforming Digital Worlds*, Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett (Eds.). Springer International Publishing, Cham, 132–138.

[58] Nava Tintarev and Judith Masthoff. 2011. *Designing and Evaluating Explanations for Recommender Systems*. Springer US, Boston, MA, 479–510. https://doi.org/10.1007/978-0-387-85820-3_15

[59] Christiana Tsiourti, Maher Ben Moussa, João Quintas, Ben Loke, Inge Jochem, Joana Albuquerque Lopes, and Dimitri Konstantas. 2018. A Virtual Assistive Companion for Older Adults: Design Implications for a Real-World Application. In *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*, Yaxin Bi, Supriya Kapoor, and Rahul Bhatia (Eds.). Springer International Publishing, Cham, 1014–1033.

[60] Joshua R. Tyler and John C. Tang. 2003. When Can I Expect an Email Response? A Study of Rhythms in Email Usage. In *ECSCW 2003*, Kari Kuutti, Eija Helena Karsten, Geraldine Fitzpatrick, Paul Dourish, and Kjeld Schmidt (Eds.). Springer Netherlands, Dordrecht, 239–258.

[61] Amy Voida, Wendy C. Newstetter, and Elizabeth D. Mynatt. 2002. When Conventions Collide: The Tensions of Instant Messaging Attributed. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 187–194. https://doi.org/10.1145/503376.503410

[62] Steve Whittaker, Vaiva Kalnikaite, Victoria Hollis, and Andrew Guydish. 2016. 'Don't Waste My Time': Use of Time Information Improves Focus. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1729–1738. https://doi.org/10.1145/2858036.2858193

[63] Michael Winikoff, Jocelyn Cranefield, Jane Li, Cathal Doyle, and Alexander Richter. 2021. The Advent of Digital Productivity Assistants: The Case of Microsoft MyAnalytics.. In *HICSS*. 1–10.

[64] Ting-Wei Wu, Yu-Ling Chien, Hao-Ping Lee, and Yung-Ju Chang. 2021. IM Receptivity and Presentation-Type Preferences among Users of a Mobile App with Automated Receptivity-Status Adjustment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 640, 14 pages. https://doi.org/10.1145/3411764.3445209

[65] Neil Yorke-Smith, Shahin Saadati, Karen L Myers, and David N Morley. 2012. The design of a proactive personal agent for task management. *International Journal on Artificial Intelligence Tools* 21, 01 (2012), 1250004.

[66] Fengpeng Yuan, Xianyi Gao, and Janne Lindqvist. 2017. How Busy Are You?: Predicting the Interruptibility Intensity of Mobile Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 5346–5360. https://doi.org/10.1145/3025453.3025946