

The Spectrum of Proactive Functioning in Digital Companions

Sanjiv S. Jha
University of St.Gallen
St. Gallen, Switzerland
sanjiv.jha@unisg.ch

Nicolò Ghielmini
University of St.Gallen
St. Gallen, Switzerland
nicolo.ghielmini@student.unisg.ch

Kimberly Garcia
University of St.Gallen
St. Gallen, Switzerland
kimberly.garcia@unisg.ch

Simon Mayer
University of St.Gallen
St. Gallen, Switzerland
simon.mayer@unisg.ch

Abstract

The future of proactive Digital Companions (DCs)—smart agents capable of assisting and protecting their users—lies in their ability to collaborate effectively with users, learn their preferences, and adjust their behavior according to the user's current state and their environment. To achieve this, DCs must carefully strike a balance between acting autonomously, while keeping users informed to minimize inconveniences, thereby enhancing user acceptance. In this article, we conduct a user survey to enrich the architecture of proactive personal DCs that explores the trade-off between full *autonomous functioning* and ensuring sufficient user control in various critical and non-critical scenarios. Our findings indicate that most of the participants lean towards having control over the actions of DCs, and will actively collaborate with those systems in everyday situations, in which decisions are not urgent. However, participants would not mind yielding their control to DCs in time-sensitive or urgent scenarios. Furthermore, in our survey, participants highlighted the importance of explaining such actions well. Given the results from our survey, we implemented a system prototype with the enriched architecture of an explainable and unobtrusive proactive DC for a smart home environment. The actions of this DC are not always fully autonomous and follow our survey findings.

CCS Concepts

• **Human-centered computing** → **Personal digital assistants**; **User studies**; • **Computing methodologies** → *Scene understanding*.

Keywords

Scene Understanding, Proactive, Video, Digital Assistants, Digital Companion



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

MUM '24, December 01–04, 2024, Stockholm, Sweden
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1283-8/24/12
https://doi.org/10.1145/3701571.3701606

ACM Reference Format:
Sanjiv S. Jha, Nicolò Ghielmini, Kimberly Garcia, and Simon Mayer. 2024. The Spectrum of Proactive Functioning in Digital Companions. In *International Conference on Mobile and Ubiquitous Multimedia (MUM '24), December 01–04, 2024, Stockholm, Sweden.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3701571.3701606

1 Introduction

Several studies in the last decades have explored the proactive behavior of Digital Companions (DC) [15] and digital assistant systems for helping users in a variety of situations, such as assisting them in do-it-yourself projects [6], reminders [21], and elderly care [9]. Considering these diverse application scenarios and the task-specific architectures of these assistance systems, Garcia et al. [4] proposed an architecture for DC systems that *autonomously* observe an environment and *proactively* assist or protect their users. These companions can be categorized according to the environment they are responsible for, into personal and expert companions. Both of these DC variants aim to assist and protect a user through *perceiving* the environment, *interpreting* the situation by contextualizing the perceived data, *deciding* on a set of actions that can support the user, and then *interacting* with the environment and the user to realize the selected support action.

Different definitions of *proactive* behavior have been presented over the last decades: Wooldridge et al. [19] define proactive agents as those "that do not simply react in response to their environment, they can exhibit goal-directed behavior by taking the initiative". According to Reicherts et al. [17], *proactive* behavior of a digital assistant system involves system-initiated interactions triggered by events related to users and their environment, contrasting to user-initiated actions, such as manually set reminders and alerts. Nothdurft et al. [12] define proactivity in technical systems as an autonomous, anticipatory system-initiated behavior, to act in advance of an anticipated future situation, rather than only reacting to it. For example, if a system continuously observes a user's body temperature and deduces that they are feeling cold, it might proactively raise the room temperature. Yorke-Smith et al. [20] highlight that even though a digital assistant system may autonomously observe a user and its environment to anticipate future events, it may choose to act (or not) on these events proactively depending on the user's context and preferences. Based on this definition, Garcia et al. [3] highlight two of the core requirements of a proactive DC—*transparency* and *unobtrusiveness*—that may lead it to be trustworthy and seamlessly integrated with users' daily activities. Nothdurft et al. [12] point

out that considering *if*, *how*, and *when* proactive behavior is needed may enhance user acceptance of digital assistant systems.

In this article, we investigate DC proactivity further and argue that the need for proactive behavior should be assessed in each specific situation. If a proactive behavior is indeed needed, a DC should be able to adopt an appropriate level of proactivity (see Section 2) to seamlessly integrate into a user's environment while assisting; and it should provide explanations of its decisions to the user. This approach aims to achieve a favorable balance between DCs proactively assisting users and remaining unobtrusive. To this end, we conducted a user survey (see Section 3) to gather insights on the level of proactivity that users prefer in a DC across various scenarios, specifically considering the perceived *urgency to act* in each scenario. In addition, we have conceptualized, implemented, and validated a prototype of a *transparent* and *unobtrusive* proactive DC that is capable of understanding a situation a user is in and then can assist this user at an appropriate level of proactivity while explaining its decisions (Section 4). Our core contribution is a refined DC architecture that can adjust its proactive behavior to align with the user's proactivity preferences, the given urgency in a situation, and the impact of the actions.

2 Related Work

Many architectures for proactive assistant systems have been proposed in the last decade to make proactive systems less intrusive and more helpful. For instance, Machado et al. [9] propose a Proactive Actions Manager (PAM) for caregiving services. It uses Bayesian networks to learn from the patient's previous actions in response to various events and proactively avoids unwanted situations (e.g., when the patient does not take their medication). Miksik et al. [11] argue that digital voice assistants should consider user preferences and autonomously determine the optimal timing, frequency, format, and sequence of notifications through context awareness and semantic content modeling. This approach aims to capture the user's attention for proactive notifications and facilitate effective communication via the voice assistant's user interface without causing disturbance. Similarly, several other recent studies on human collaboration with digital assistants (cf. [1, 6, 22]) investigated ways to increase user acceptance, such as by anticipating appropriate timing [1] and refining (or augmenting) the digital assistant's dialogues with users [21]. To describe the extent of a system's intervention, Sheridan et al. [18] proposed system automation levels that were refined by Parasuraman et al. [14] as ranging from (1) *Offering no assistance* to (10) *Fully autonomous functioning in a system.* On this basis, Kraus et al. [7] introduced four levels of proactivity to help digital assistant systems moderate their autonomous behavior in action selection and execution, making them less intrusive to enhance user acceptance. The levels of proactivity according to Kraus et al. [7] are: (1) *None: The user must actively ask the system for support*; this level refers to reactive systems (e.g., Amazon Alexa, Microsoft Cortana). (2) *Notification: The system proactively notifies users of important events to let them decide on the next step to take*; for example, social media notifications on a smart device. (3) *Suggestion: The system proactively notifies users of important events and suggests possible actions including their respective explanations*; the user may then decide whether to act upon the suggestion. (4) *Intervention: The system detects a problem and takes full responsibility for acting autonomously. It provides the user with an explanation of the actions being executed*; for example, the crash detection function on smart devices (e.g. phones, watches) may proactively inform emergency services.

In a focus group discussion on designing a reference model of digital assistants, Meurisch et al. [10] found that—depending on the application area (e.g., mental, physical, activity, environment, social, and technology support)—users express more or less resistance towards high levels of proactivity. However, they also show that providing appropriate reasoning about the assistant's decision-making counteracts user frustration and loss of trust in the assistant [10]. While very few other studies, such as [8, 12, 16] have discussed the significance of providing reasoning of actions for regaining user trust that might be lost due to an unexpected intervention of the system, we emphasize that modern proactive DCs should prioritize estimating the context-based urgency for taking action and explain their reasoning behind those actions to users. This should have positive effects on the desired characteristics—*unobtrusiveness* and *transparency*—of a proactive DC highlighted by Garcia et al. [3]. Yet, today's architectures for proactive digital assistant systems and proactive DCs do not proactively adapt to different levels of proactivity while explaining their actions to users to seamlessly blend into the user's environment.

3 Finding the User-Desired Level of Proactivity

To empirically validate our argument that estimating the context-based urgency, understanding users' preference for the level of proactivity, and explaining the reasons behind proactively taken actions may enhance the *unobtrusiveness* and *transparency* of the DC, we conducted a user survey among 19 participants. We defined seven diverse scenarios in which a DC might provide support to a user. For example, a DC capable of anticipating user purchases and identifying items a user might need to check off their shopping list (Scenario 1 detailed in Appendix A on the supplementary material); a DC for worker safety that identifies potential hazards if an employee enters a machine room (Scenario 3), and supports workers on the shop floor to improve efficiency and worker satisfaction in a car tire replacement business (Scenario 7). These scenarios were presented to 19 participants at our institution. 17 participants (84%) were between 18 and 35 years old and the rest were below 45 years of age. 18 participants (95%) had at least a bachelor's degree and one was a high school graduate. After reading each scenario, the participants were asked to: (1) rate the impact that the decision or action would have on their lives considering health, social status, and economic aspects (five-point Likert scale ranging from 1: low impact to 5: high impact), (2) rate the level of urgency that this scenario requires (five-point Likert scale ranging from 1: not urgent to 5: highly urgent), (3) select their desired level of proactivity in the scenario according to Kraus et al. [7]: *None*, *Notification*, *Suggestion*, and *Intervention*, and (4) provide the reason for selecting that level of proactivity (open text field).

Our survey results (see Figure 1) show that the greater the level of urgency to act, the stronger the preference for a higher level of proactivity is; for instance, in Scenario 3, where an employee

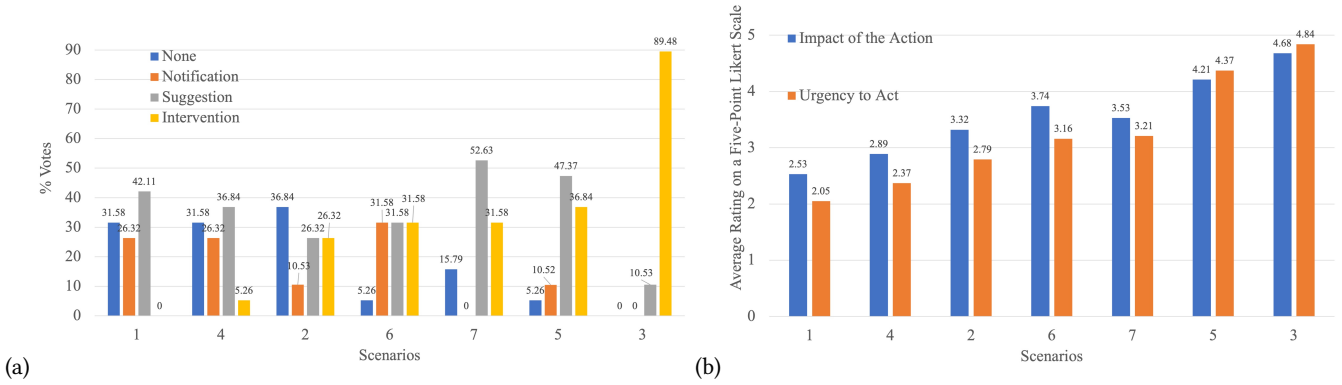


Figure 1: Survey results on the desired level of proactivity in a DC across seven scenarios. (a) Average preferences for each proactivity level (cf. [7]) scenario. (b) Average ratings for each scenario illustrate the perceived sense of urgency and the need for action.

unintentionally enters a dangerous machine room, exposing himself to danger, all participants assign high or very high urgency and impact values (average ratings 4.84 and 4.68, respectively) and 89.48% of the participants select *Intervention* as the preferred proactivity level. The provided arguments in the open text fields discuss that, since humans may not have sufficient time to think and act efficiently in critical situations, the system's rapid processing of large amounts of data and quick decision-making would allow for a desired intervention. Other than for Scenario 3, the preferred level of proactivity is *Suggestion* since this allows the user to remain in charge of deciding whether to act or not. However, in scenario 2, the rating distribution indicates a lack of consensus on the preferred level of proactivity. In this scenario, the participants were asked to choose between letting a DC help their 7-year-old who did not know what to wear to school and helping their kid themselves despite an urgent meeting. They rated the scenario as only slightly impactful and with a slightly low urgency. Regarding the choice of proactivity level in this scenario, a majority of seven participants opted for *None*, five for a *Suggestion*, five for an *Intervention* by the DC, and two preferred a *Notification*. Across scenarios, participants' explanations for choosing a preferred level of proactivity indicate that people have concerns about the system's obtrusiveness when a proactive system is installed in a personal space which leads to a lowering of the preferred level of proactivity. For example, one of the participants in Scenario 1 selected *None* as the preferred level of proactivity and wrote "The other options can get annoying fast. Think about all the spam emails, and recommendations on Instagram" whereas another participant for the same scenario chose *Notification* and wrote "In a relaxed atmosphere a reminder would be sufficient".

Our survey results motivate our argument that a proactive DC system could use the urgency of a situation to act together with a user's proactivity preferences to select the most appropriate proactivity level, supporting the system's overall unobtrusiveness (e.g., participants expressed that "Choice between Suggestion and Intervention depends on degree of danger" and "I prefer to be able to confirm."). Such a system should further explain its actions to users to mitigate privacy and transparency concerns (e.g., a participant wrote "I would prefer a notification so that I can review why the system called the police", another one wrote "we would need to make sure the data is not collected/deleted right away and is not abused to control employees").

4 Proactive Personal DC Architecture and System Implementation

Given our survey results, we implemented a prototype of a DC that showcases the significance of contextual understanding and user preferences to infer action urgency and select an appropriate proactivity level for unobtrusively assisting users while providing explanations.

4.1 Proactive Personal DC Architecture

Our prototype follows the architecture of DCs proposed by Garcia et al. [4]; consisting of four phases to assist its users (see Figure 2). In the *Perceive* phase, a DC may observe the environment using multiple sensors (connected devices) such as cameras, microphones, and environmental sensors. A DC then *Interprets* the observed data using a knowledge base, which can be the common sense of an LLM to contextualize the perceived data. Then, the DC *Decides* on an appropriate set of actions to assist or protect a user, and produces explanations of those actions based on the information from the *Interpret* phase. In the *Interact* phase, a DC may actuate connected devices available in the environment to execute actions autonomously or collaborate with the user to confirm an action before it is taken.

Garcia et al. [4] distinguish between personal and expert DCs, stating that a personal DC takes care of a user, while an expert DC is in charge of a physical or virtual environment. We propose to expand the architecture for personal DCs to prioritize their unobtrusive behaviors and promote transparency. Hence, as Figure 2 shows, the urgency of a situation is assessed in the *Decide* phase based on the user's proactivity preferences (learned over time and curated by the users considering their health, social status, and economic aspects) and the contextual understanding of the current situation. These preferences may specify the proactivity levels of a

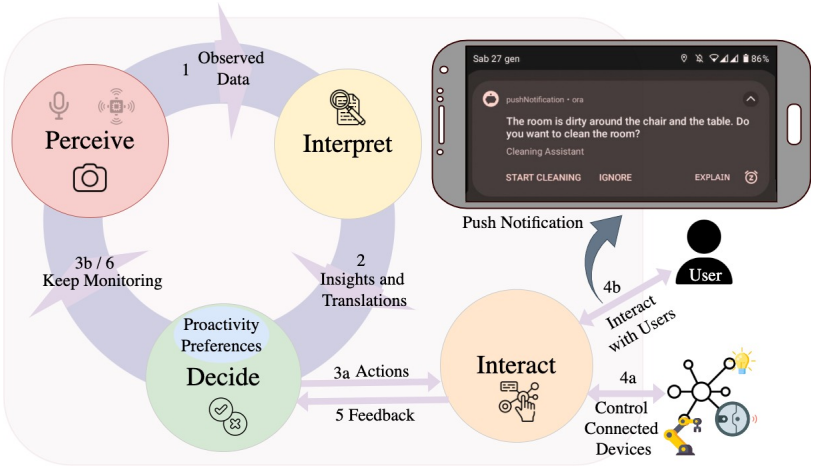


Figure 2: Phases of the system architecture of our proactive DC system. This augments the Decide phase of a personal DC [4] by adapting its proactivity level based on user preferences and interpretation of the context. In an example implementation scenario, the push notification suggests to start cleaning. The user may decline or accept the *suggestion* and also ask the DC to explain its actions.

user's choice of interaction in typical and critical scenarios. Thus, in the *Decide* phase, a DC computes a set of actions considering the urgency of the situation. Then, in the *Interact* phase, these actions are either autonomously executed and the user is notified about it, or a dialogue is established with the user to confirm the execution of an action; in which case, feedback to the *Decide* phase is provided for future decision-making.

4.2 Prototype Implementation for a Smart Home Scenario

To validate the architecture proposed in Section 4.1, we implemented a prototype of a personal DC in a smart home scenario. It perceives the environment and can use connected devices, such as a robot vacuum cleaner, to assist its users. In our implementation, the robot vacuum cleaner is simulated using a *Lidar Bot AGV Mini Carkit Mecanum Wheels*.[1]

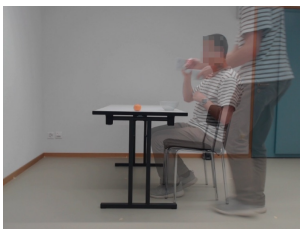
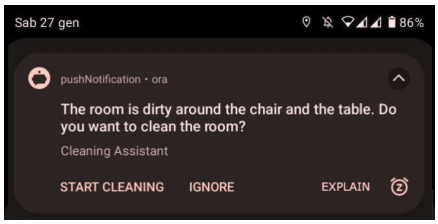
Following the phases of the proposed proactive personal DC architecture, in the *Perceive* phase, our DC uses an RGB camera module to capture a sequence of video frames that are processed using a scene graph generation algorithm—RelTR [2]. RelTR detects objects and the relationships among these objects in the scene. Hence, its output is a Scene Graph (SG) composed of semantic triples (subject-predicate-object) that denote the class name of pairs of detected objects and their relationship (e.g., `cup on table`) and the 2D-pixel positions (bounding boxes) of the objects. Given that our DC can assist its user in the real world, for example, by cleaning the floor, the positions of each identified object in the scene (2D coordinates derived using bounding boxes) are converted into (real-world) 3D coordinates. To find these 3D coordinates from an object's position in an observed 2D image frame, we use the inverse pinhole camera model[2] that describes the relationships between 3D points and their projection on a 2D plane. Thus, the DC determines the 3D coordinates of the objects inside an area of interest (e.g., on the floor see Figure 3 (c)). Our DC then takes the SG generated from subsequent frames captured every ten seconds to build a temporal structure that we refer to as a *temporal graph* to track objects and their relationships over time[3]. This temporal graph with the 3D coordinates of each identified object is then verbalized (i.e., converted into a text form) to create a scene description of each second.

In the *Interpret* phase, the generated scene description is fed to an LLM—in our implementation, GPT-4 [13]—to contextualize further the perceived information, and to understand the evolution of the scene by leveraging the LLM's common-sense knowledge. Then, the DC *Decides* on the actions to assist a user while addressing key concerns, including the estimation of the urgency of the situation and the user's context-specific preferences for proactivity levels, as identified in our survey (see Section 3). To do so, it uses a knowledge base that specifies the proactivity preferences of the users. This knowledge base can be represented as a JSON object or expressed in Resource Description Framework (RDF) as part of a Knowledge Graph. Thus, given the contextualized data from the *Interpret* phase, the proactivity preferences of a specific user, and the specification of what each proactivity level means, our DC identifies *if*, *how*, and *when* it should execute an action (e.g., cleaning an area) to be unobtrusive. Finally, in the *Interact* phase, the DC executes the computed tasks by communicating with connected devices (e.g., the vacuum cleaner), and it is capable of explaining its actions to the users via a push notification on the user's smartphone (see Figure 3 (b)). By explaining its actions, our prototype addresses another key concern expressed by our survey participants: "transparency" (see Section 3).

[1] https://shop.m5stack.com/products/new-lidar-bot-agv-mini-carkit-mecanum-wheels

[2] https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html

[3] A detailed description of this approach will appear soon as [5]

(a)  (b) 

(c) 

Figure 3: The robot starts to follow approximately the same path in (c) as the person's path in (a) after he permits the DC to clean through the push notification in (b) since the LLM interprets the shown activities as a possible source of dirt and spillage.

4.3 Prototype Validation

Our DC prototype starts working by *perceiving* its environment through its camera module. An overlay of multiple captured frames is illustrated in Figure 3 (a), where a man walks to a table and enjoys breakfast from a bowl. While perceiving the environment, the DC creates a scene description of the captured frames every 10 seconds and computes the 3D coordinates of the objects (converted from 2D-pixel positions). Since the DC has access to its users' proactivity preferences stored in a knowledge base, it knows their preferences for cleanliness in different scenarios (we classify them into three categories—Critical, Standard, and Reminders—based on the urgency of the user's required actions). Then, our DC leverages the common sense knowledge of GPT-4 to *interpret* the scene description phase. To do so, the DC uses the following prompt: *"Considering what people are explicitly doing in the scene, what activities are associated with making the floor dirty?"* to which the LLM responds with a list of activities recognized such as walking, eating, drinking, and handling potentially dirty objects. Our DC considers those recognized activities that potentially make the floor dirty and assesses the urgency of action based on the user's proactivity preferences (e.g., *Intervention* in critical situations, *Suggestion* in a standard situation, *Notification* for Reminders) to *decide* whether the floor needs to be cleaned and select the appropriate proactivity level; it thereby increases *unobstrusiveness* and conforms to our survey findings (see Section 3). To achieve this, our DC, in its *Decide* phase again utilizes GPT-4 by asking: *"Considering the given user's proactivity preferences, infer the urgency of cleaning the room. Would you recommend cleaning the floor or not? Which proactivity level would you recommend? Explain"*. In response, GPT-4 recommends using *Suggestion* as the proactivity level and explains its reasoning (this is communicated to the user in the *Interact* phase of our DC to promote *transparency*): *"The man (man_8caa) holding a bowl (bowl_e558) and later placing it on the table(table_7953). If the bowl is not properly cleaned or if there are any spills or drips from the bowl, it can make the floor dirty. While the need for cleaning is apparent, it falls under the standard preference category for the user. The user may decide when to do so '*. Because GPT-4 recommends to *clean*, our DC then asks it to list the 3D coordinates of the dirty areas in the room for deriving a path for the connected vacuum cleaner to follow. Our DC uses the resulting list of 3D coordinates of all of the detected dirty areas, such as *"the area beneath shoe_1cb8: (37, 206, 0)* and *the area beneath chair_e56a: (176, 195, 0)* to derive a path for the vacuum cleaner to clean along.

Since our DC decided on *Suggestion* as the proactivity level for this scenario, in its *Interact* phase, it asks the user to confirm cleaning the space. Thus, the DC sends a push notification (see Figure 3 (b)) to the user's smartphone where the user can ask for an explanation for the suggested action (by selecting `EXPLAIN`), addressing the user's transparency concerns identified in our survey results. If the user declines (selecting `IGNORE`), the DC returns to monitoring the environment. Otherwise, when the user wants to `START CLEANING`, our prototype sends the derived path (using the 3D coordinates of dirty areas) to the connected vacuum cleaner, which then starts cleaning by following the path (see Figure 3 (c)). Thus, it showcases the DC's flexibility in achieving unobtrusiveness by acting according to the user's scenario-based preference for proactivity level. In either of the cases, our DC keeps the user's response as feedback to improve its understanding of the user preferences and better assist them in the future.

5 Conclusions and Future Work

In this article, we investigate proactivity in DCs; after surveying related work, we decided to take urgency, the estimated impact of an action on a user's health, social status, and economic aspects, as well as transparency, and unobtrusiveness as the main characteristics of proactive personal DCs. We considered Kraus [7] four levels of proactivity to conduct a survey in which we asked 19 participants to provide feedback about the level of proactivity that they would be comfortable with in seven different scenarios. This survey revealed that the preferred level of proactivity is *Suggestion*. Based on our findings, we then proposed to enrich a DC architecture to include levels of proactivity. To continue our work, we plan to expand our participant pool in the survey. Moreover, we will evaluate our proposed architecture regarding its effectiveness in several other real-world scenarios such as in workplaces, outdoors, and in personal care. In this implementation, we used an RGB camera module to estimate 3D coordinates for the vacuum cleaner to follow. However, we plan to enhance the accuracy of our DC in the future using an RGB-D camera.

Acknowledgments

This work has been supported by the BMK, BMAW, and FFG COMET K1 Center "Pro²Future" (Products and Production Systems of the Future), Contract No. 881844, within the E-Minds project.

References

[1] Shaftab Ahmed, Mohammad Ilyas, and M. Yasin Akhtar Raja. 2019. Smart Living: Ubiquitous Services Powered by Ambient Intelligence (AmI). In *2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT)*. IEEE, Charlotte, NC, USA, 043–048. https://doi.org/10.1109/HONET.2019.8908054

[2] Yuren Cong, Michael Ying Yang, and Bodo Rosenhahn. 2023. RelTR: Relation Transformer for Scene Graph Generation. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 9 (sep 2023), 11169–11183. https://doi.org/10.1109/TPAMI.2023.3268066

[3] Kimberly Garcia, Simon Mayer, Alessandro Ricci, and Andrei Ciortea. 2020. Proactive Digital Companions in Pervasive Hypermedia Environments. In *2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC)*. IEEE, Atlanta, GA, USA, 54–59. https://doi.org/10.1109/CIC50333.2020.00017

[4] Kimberly Garcia, Jonathan Vontobel, and Simon Mayer. 2024. A Digital Companion Architecture for Ambient Intelligence. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–26.

[5] Sanjiv S Jha, Kimberly Garcia, Yasmine Antille, Marc Solèr, Simon Padua, and Simon Mayer. 2024. Temporal Scene Understanding using Contextually Unique Identification. Accepted at the 36th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2024).

[6] Matthias Kraus, Marvin Schiller, Gregor Behnke, Pascal Bercher, Michael Dorna, Michael Dambier, Birte Glimm, Susanne Biundo, and Wolfgang Minker. 2020. "Was that Successful?" On Integrating Proactive Meta-Dialogue in a DIY-Assistant using Multimodal Cues. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (Virtual Event, Netherlands) *(ICMI '20)*. Association for Computing Machinery, New York, NY, USA, 585–594. https://doi.org/10.1145/3382507.3418818

[7] Matthias Kraus, Nicolas Wagner, Zoraida Callejas, and Wolfgang Minker. 2021. The Role of Trust in Proactive Conversational Assistants. *IEEE Access* 9 (2021), 112821–112836. https://doi.org/10.1109/ACCESS.2021.3103893

[8] Yuanchun Li, Hao Wen, Weijun Wang, Xiangyu Li, Yizhen Yuan, Guohong Liu, Jiacheng Liu, Wenxing Xu, Xiang Wang, Yi Sun, et al. 2024. Personal LLM Agents: Insights and Survey about the Capability, Efficiency and Security. *arXiv:2401.05459* (2024).

[9] Alencar Machado, Daniel Lichtnow, Ana Marilza Pernas, Leandro Krug Wives, and José Palazzo Moreira de Oliveira. 2014. A Reactive and Proactive Approach for Ambient Intelligence. In *Proceedings of the 16th International Conference on Enterprise Information Systems - Volume 2* (Lisbon, Portugal) *(ICEIS 2014)*. SCITEPRESS - Science and Technology Publications, Lda, Setubal, PRT, 501–512. https://doi.org/10.5220/0004884205010512

[10] Christian Meurisch, Maria-Dorina Ionescu, Benedikt Schmidt, and Max Mühlhäuser. 2017. Reference Model of Next-Generation Proactive Personal Assistant: Integrating Proactive Behavior. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 149–152. https://doi.org/10.1145/3123024.3123145

[11] Ondrej Miksik, I Munasinghe, J Asensio-Cubero, S Reddy Bethi, ST Huang, S Zylfo, Xuechen Liu, T Nica, A Mitrocsak, S Mezza, et al. 2020. Building Proactive Voice Assistants: When and How (not) to Interact. *arXiv:2005.01322* (2020).

[12] Florian Nothdurft, Stefan Ultes, and Wolfgang Minker. 2014. Finding Appropriate Interaction Strategies for Proactive Dialogue Systems—An Open Quest. In *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication*, Vol. 110. Citeseer, LiU Electronic Press, Tartu, Estonia, 73–80.

[13] OpenAI. 2024. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL] https://arxiv.org/abs/2303.08774

[14] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 2000. A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30, 3 (2000), 286–297.

[15] Malcolm Peltu and Yorick Wilks. 2008. Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues. In *OII Forum Discussion Paper No. 14*. Elsevier BV, Oxford, United Kingdom.

[16] Pei-Luen Patrick Rau, Ye Li, and Jun Liu. 2013. Effects of a Social Robot's Autonomy and Group Orientation on Human Decision-Making. *Advances in Human-Computer Interaction* 2013, 1 (2013), 263721.

[17] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I Interrupt? Diverging Opinions on Proactive Smart Speakers. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 34, 10 pages. https://doi.org/10.1145/3469595.3469629

[18] Thomas B Sheridan, William L Verplank, and TL Brooks. 1978. Human/Computer Control of Undersea Teleoperators. In *NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control*. NASA, Massachusetts, United States.

[19] Michael Wooldridge and Nicholas R Jennings. 1995. Intelligent Agents: Theory and Practice. *The knowledge engineering review* 10, 2 (1995), 115–152.

[20] 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.

[21] Nima Zargham, Leon Reicherts, Vino Avanesi, Yvonne Rogers, and Rainer Malaka. 2023. Tickling Proactivity: Exploring the Use of Humor in Proactive Voice Assistants. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia* (Vienna, Austria) *(MUM '23)*. Association for Computing Machinery, New York, NY, USA, 294–320. https://doi.org/10.1145/3626705.3627777

[22] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 3, 14 pages. https://doi.org/10.1145/3543829.3543834

A List of Scenarios

This section describes the seven scenarios that we presented to our survey participants to answer the four questions mentioned in the article.

A.1 Scenario 1

Scenario: It is a Saturday afternoon and you are out shopping with friends. You are in a nice fashion shop in the city center, you have already carefully chosen a pair of elegant trousers and a pair of shoes. The finishing touch is a bow tie, which you need to buy to complete the outfit. In this relaxed atmosphere, you run the risk of forgetting to buy one.

System: A sophisticated system discreetly observes customers selecting clothes, employing predictive algorithms to anticipate potential purchases and identifying items they might need. This technology enhances the customers' shopping experience by offering timely recommendations and minimizing the risk of oversight, ensuring a seamless and personalized customer journey.

A.2 Scenario 2

Scenario: You are a manager working from home. It is the end of the year and you are in a work meeting discussing next year's budget, which requires your full concentration. At the same time, your 7-year-old son is standing in front of the wardrobe and does not know what to wear to school.

System: A smart home system observes the activities in your home. It uses predictive algorithms to recognize people's behaviors. This technology goes beyond mere monitoring. It is able to anticipate the potential needs of occupants, improving safety and comfort in the home.

A.3 Scenario 3

Scenario: You are the production manager of a furniture company. You orchestrate the use of machines and human resources. Today, just before an important production deadline, the assembly line is humming and you are in the warehouse managing logistics. But suddenly an employee unintentionally enters a dangerous machine room, exposing himself to danger.

System: A smart manufacturing system, connected to various devices and machines in the factory, carefully monitors activity within the building and uses advanced algorithms to recognize people's actions, understand their needs, and identify potential hazards. Based on this, the system improves people's efficiency and proactively enhances safety.

A.4 Scenario 4

Scenario: You are a passionate skateboarder who is preparing for a ride on a beautiful Friday evening. Safety is very important to you, so you usually gear up carefully before each session and put on your helmet, knee pads, elbow pads, and wrist guards. However, there is always a risk of forgetting to put something on.

System: A smart system in your home is capable of recognizing people's behaviors and uses predictive algorithms. This system is capable of anticipating the potential needs and habits of residents, improving the safety, convenience, and comfort of people. To enhance delivering assistance, the system includes a mobile home robot.

A.5 Scenario 5

Scenario: You are on a peaceful holiday far away from Switzerland, enjoying the sun and the calming sound of the sea. The atmosphere is perfect for relaxing and unwinding. At the same time, however, someone unexpectedly enters your home in Switzerland.

System: A smart system recognizes people's activities. The system is capable of anticipating potential dangers and acting upon them to improve safety.

A.6 Scenario 6

Scenario: You are relaxing at home on the sofa watching a film like you do every evening. At 9.00 pm you have to take your daily medicine.

System: A smart home system is capable of understanding people's behaviors and habits based on previous observations. The system includes a mobile robot that can freely operate inside the house.

A.7 Scenario 7

Scenario: You are a garage owner. Winter is on its way and the car tire-changing business is booming. Accordingly, you are currently changing tires all the time and are often on the lookout for the right tools.

System: A smart system perceives the activities in the garage. It uses predictive algorithms to recognize people's activities and it is capable of anticipating potential dangers and worker needs. The system is capable of improving workers' safety and increasing efficiency.