Scrutability of Intelligent Personal Assistants

JOVAN JEROMELA, ADAPT Centre, Trinity College Dublin, Ireland

Intelligent personal assistants (IPAs) have become widely available, yet they remain primarily used for discrete, straightforward tasks. By contrast, both user studies and literature reviews indicate that IPAs of the future are to be personalised, proactive, and capable of performing elaborate undertakings. Such systems would have to be based on complex and dynamic user and context models. We believe that scrutability – i.e. the ability of the user to actively study and modify the models towards tuning personalisation – could emerge as an essential element of such a human-assistant interaction paradigm. Yet, to the best of our knowledge, no work so far has investigated how the principles of scrutability, as presented in [21], relate to the context and novel challenges raised by the proactive IPAs and how scrutability could facilitate effort-efficient control of the assistants. This paper introduces our vision of the confluence of the research fields of IPAs and scrutability, presents a diagram of the proposed interaction structure, and reanalyses data from user studies originally presented in [11, 39] to better understand user expectations regarding scrutability and proactivity of IPAs.

CCS Concepts: • **Human-centered computing** → **Personal digital assistants**; *User models*; Natural language interfaces; *User centered design.*

Additional Key Words and Phrases: scrutability, proactivity, intelligent personal assistants, personalization, user representation

ACM Reference Format:
Jovan Jeromela. 2022. Scrutability of Intelligent Personal Assistants. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3503252.3534355

1 RESEARCH PROBLEM

Virtual assistants, or intelligent personal assistants (IPAs), as they are frequently titled in literature [5], became omnipresent in recent years thanks to the now-common practice of integrating such services in personal computers, smartphones, smartwatches, smart TVs, and smart speakers [6, 18, 31]. Simultaneously, natural language user interfaces of IPAs advanced to enable both casual exchanges and fulfilling tasks such as controlling processes of the hosting device or of the internet of things (IoT), as surveyed in [12, 37]. Nevertheless, as noted by Clark et al. [7], the promise of true conversational interaction has not yet been delivered. While functionalities such as making reservations [28], scheduling meetings [3], and discerning and communicating germane pieces of information [35] make IPAs' prospective capabilities more akin to those of their human counterparts, Edwards et al. [15] concluded that IPAs might, in fact, disrupt multitasking. Moreover, it has been noted [11, 13, 15] that IPAs are currently mostly used for simple tasks such as setting alarms, playing music, setting reminders, and getting weather information. As these discrete, transaction-oriented conversations fall short of the envisioned potential [7, 11], human-assistant interaction remains an active area of research.

Further research in the field will likely investigate enabling a greater level of *proactivity* and *personalisation* [12, 18, 26, 35], which would enable the assistant to predict and perform actions most appropriate for the user's context and intent. Currently, the proactivity of assistants is mostly limited to presenting relevant *cards* (containing information

Manuscript submitted to ACM

such as calendar events, stocks prices, or road conditions) and pushing notifications [35, 36]. A majority of people in the study from [39] envisioned that future proactive assistants would also offer "well thought-through suggestions and recommendations to solve complex problems" and initiate conversations inspired by the knowledge of the user and her environment. The authors then conclude that such an assistant might rely on querying the user at opportune moments. This contrasts with the majority of contemporary solutions, which rely on explicit user activation [18, 30].

The announcement of Google Duplex [24], a robocalling system for making reservations that was to become a feature of the Google's IPA, included recordings of the system in action. One of the examples involved a case in which the initial desired booking date had already been reserved, so alternatives had to be found. Furthermore, [3] presented attempts of creating a virtual assistant capable of automating meeting scheduling. Prior to this, a personal meeting scheduling agent capable of bilateral and multilateral negotiations was proposed in [34]. These works indicate that future IPAs would likely *delegate for*, or represent, their users in conversations or even *negotiate* on their users' behalves with other assistants or humans.

The participants in a survey on the expectations of a perfect future assistant [39] anticipated the system would need to be readily aware of changes in the surrounding. Such assistants would ineluctably rely on complex and dynamic models, as they would necessitate representing not only the user but also her social relationships, mental and physical state, and the relevant broader context. Yet, even with today's comparatively less complex IPAs, the lack of trust has been recognised as one of the main reasons for not using virtual assistants [11, 25]. We can thus assume that human-assistant interaction in the future would include communicating how the IPA understood the user's desiderata and how it can minimise the likelihood of negative effects for the user and other humans the IPA interacts with on the user's behalf.

Therefore, in order to make the future *proactive and delegative IPAs* enticing to users, it is imperative to devise a mechanism through which the users could understand what information the system collects and how it is then processed to provide befitting personalised services. Our guiding premise is that a manner to achieve this is by making the intelligent personal assistants *scrutable*. Scrutability is a term representing the ability of a user to understand, interact with, adapt, and study or *scrutinise* a personalised model [21]. This concept was introduced by Judy Kay in her doctoral dissertation [20]. In [21], it is stated that the word was chosen over related concepts in literature such as *open* or *transparent* because it emphasises the need for the user to actively engage with the system and the embedded user model. We want to examine what challenges arise when scrutability is considered from the aspect of IPAs. Moreover, we want to investigate whether interacting with an assistant that is scrutable by design would lead to a greater sense of trust and rapport in comparison to the contemporary opaque IPAs.

Already with the present-day assistants with comparatively fewer capabilities, not trusting the system with complex tasks is a major issue recognised in [11, 15]. Scrutability might help mitigate this concern too, by enabling the user to verify the assistant's understanding and by allowing the IPA to confirm its classification of the user's intent. Yet, as reported in [39], users expect the perfect assistant to be smart and to possess a high degree of knowledge about them. This would indicate that the IPA would need to be effort-efficient – i.e. employing the IPA must principally be less time-consuming than completing the task without it – while also maintaining reliability by minimising the probability of performing a misidentified action. Ergo a confidence model ought to be constructed, which could appraise when an explicit confirmation is appropriate and thus balance the ease of use with the risk of negative consequences of wrong assumptions. Finding this balance within the context of the scrutable IPA is a further point of investigation.

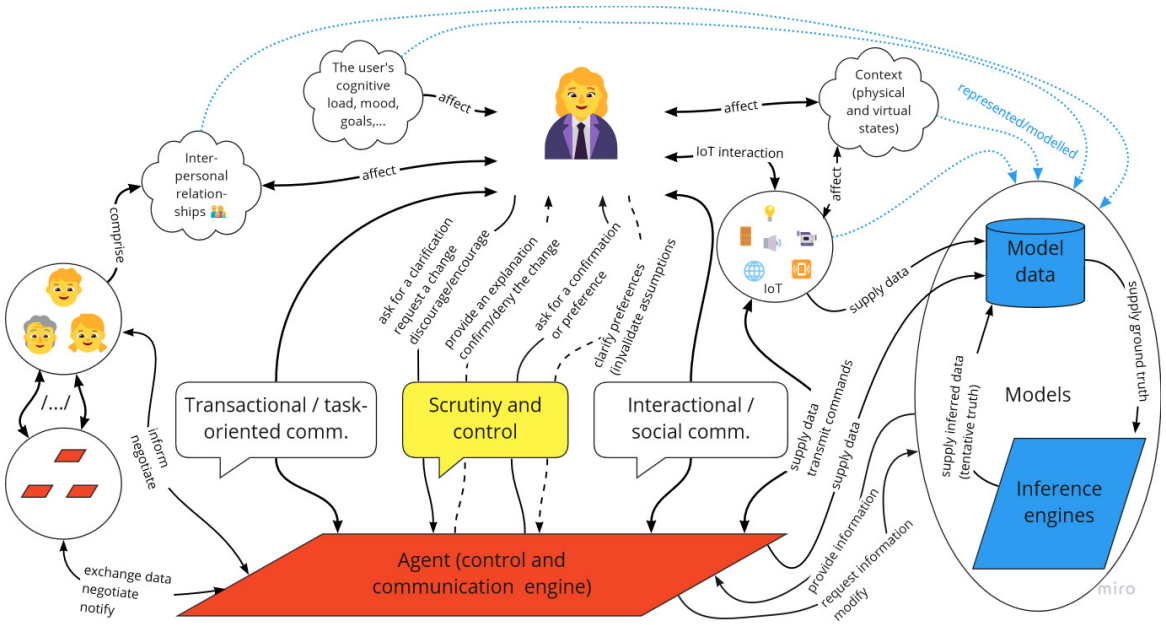

Fig. 1. Diagram of the envisioned IPA interaction design

2 RELATED WORK

2.1 Proactive and Delegative Intelligent Personal Assistants

The designation *intelligent personal assistant* became prevalent in the literature, as it declares that the core of its functionality is personalisation powered by artificial intelligence, as elaborated in [5]. Other adjectives often used [5, 12, 32, 37] to identify these assistants include virtual, digital, smart, and voice-enabled. Similarly, instead of – or in addition to – the word assistant, the concepts agent and voice interface are sometimes used. In our case, however, we strictly differentiate between the three, as indicated in figure 1. We see the *agent* as the central control and communication engine of the IPA, where the data from various models (cf. subsection 2.2) is utilised to provide services and power the *voice interface*. Following the classification from [7], the voice interface can facilitate transactional (centred around a specific goal) and interactional (ranging from small talk to longer dialogues) communication with the user. In addition, we see communication for the purposes of scrutiny and model control as a novel interaction type. As indicated in [7], even though the purposes of these classes differ, they can overlap in natural conversations. Many IPAs are capable of performing actions in the physical world by e.g. controlling the IoT devices [4, 5]. As elaborated in subsection 2.3, we expect our IPA to be able to affect its physical surrounding by interacting with the IoT. This is already possible with many widely deployed and popular commercial IPAs such as [12, 37], as well as on other systems based on open-source technology [4] or through purpose-built assistants [38].

In their systematic review [12], the authors state that "future research could deal with the proactivity and personalization challenges when developing intelligent personal assistants". Sarikaya in [32] divides IPA support into proactive (anticipatory, system-initiated) and reactive (user-initiated) assistance. He then references [17] and notes that proactive assistance operates on a proactivity continuum ranging from the user having to do the task manually; through having the assistant suggest actions to the user; and to the full autonomy of the assistant to decide and perform actions. To

achieve the latter stages on that continuum, the assistant has to model helpfulness of actions and user desires [32, 33]. A commercial example of a proactive system is XiaoIce [41] which was originally developed by Microsoft as an empathetic chatbot on Chinese social media. The system is proactively suggesting reminders for actions deemed relevant and it may nudge the user to change behaviour estimated to be disruptive. In 2012, Yorke-Smith et al. [40] presented a generic framework for proactive goal generation and deliberation. The framework was then implemented in CALO, an assistant for knowledge workers [16]. The main parts of the framework are a task manager (through which the proactivity is realised), a to-do manager that aids the user in tracking her tasks, and a calendar manager, that is also capable of delegating for the user in meeting negotiations via email. Enabling the assistant to communicate on the user's behalf for the purposes of arranging meetings was also investigated in [3]. Yet, in the subsequent years, there seems [12, 18] to have been little work on further allowing assistants to directly represent – or delegate for – their users in interactions in natural language with other people. One notable exception is Google Duplex whose base functionality – scheduling appointments at hairdressers' and restaurants – was showcased in [24]. Yet the system was controversial from the launch due to ethical concerns [28] and has to date only seen a limited distribution [27]. Nevertheless, we believe that future IPAs will be able to delegate for their users in correspondence (text messages, calls) with other people and other assistants, as represented on the left-hand side of the diagram.

2.2 Multi-Model, Metadata Driven Approach to Personalisation

Our envisioned IPA would be highly personalised and thus underpinned by modelling the user (representing her preferences, goals, state), interpersonal relations, context, and connected IoT devices (both in terms of their sensors and possible actions). This view largely overlaps with the six areas of digital personal assistance recognised in [26]. Employing the multi-model, metadata driven approach to providing personalisation of adaptive systems, as proposed and evaluated in [8–10] within the context of eLearning, allows for a greater degree of module reusability and control flexibility. As illustrated in figure 1, these models would save both the received evidence as well as inferences from evidence interpretation. This conceptualisation is akin to that of ALIGN system architecture presented in [29], which comprises a cyclical process of updating models through adaptive interaction with the user. Unlike ALIGN, however, we envision that the user would also be able to explicitly scrutinise and modify the models.

2.3 Scrutability

As indicated in figure 1, we see control of the assistant through explicit verification and scrutiny of its beliefs and actions as central to the future user-assistant interaction. Kay and Kummerfeld [21] define principles of designing scrutable user modelling and personalisation, which arise from solving problems relating to *privacy*, *invisibility* (i.e. the fact that it is often impossible to determine if and to what extent the interface differs from that of another user due to personalisation), *errors in user model* (which are exacerbated by the fact that there is often no explicit way for the user to fix them), *wasted user models* (i.e. the data that is captured by various devices but not utilised meaningfully for personalisation), and *the lack of control* of the user's personal information that is integrated into models. IoTum – presented in [23] and evolved from the *um* toolkit [19] and the *Personis* user model [22] – is a framework for user modelling for IoT applications. Application can *tell* evidence to IoTum, *ask* it for interpretations, or *listen* for triggering actions. Decisively, IoTum facilitates scrutability and accountability by maintaining provenance information. Yet, as depicted in figure 1, our envisioned IPA would not only exchange data with applications in a similar manner, but would also (proactively) control the IoT devices and correspond with other people or their IPAs. This would necessitate extending the framework to allow for scrutability and control of the broader effects of the IPA's models and actions.

3 RESEARCH QUESTION, OBJECTIVES, AND GOALS

After the author's first semester, the research question reads as follows: *How can principles of scrutability be extended to facilitate effort-efficient control of the proactive and delegative intelligent personal assistants?* The question thus explicitly assumes that IPAs in the future would be proactive – anticipating the user's intention and initiating or continuing conversation accordingly – as well as delegative, i.e. capable of representing the user in conversations and negotiations with other assistants and humans. Such assistants would likely rely on complex and dynamic models. Our initial contribution would be to investigate how scrutability, i.e. allowing users to study and directly adapt aspects of these models, could aid the control of the IPAs and if it would lead to a greater sense of trust and rapport. This investigation would then result in a theoretical framework for scrutable proactive delegative IPAs and an apt prototype. The envisioned IPA would have to reason about the provenance of modelled information and its applicability in similar situations. Namely, the assistant must possess a confidence model that balances the trade-off between potential social and financial costs of undesired actions caused by a misunderstanding on the one hand and distracting or frustrating the user by asking additional questions on the other. Consequently, investigating the *effort-efficiency* of controlling scrutable IPAs and its effect on the cognitive load would be a further point of investigation. We hold that there are both nuances and significant differences that have to be considered when deliberating about extending principles of scrutability (cf. section 2.3) to IPAs. These differences stem from questions concerning *proactivity* (When/why to initiate interaction? How do the user model and the context relate to the interruption design?), *delegation* (When/how to delegate for the user?), and the *broader effect* (How does the user model affect others?). Defining a set of design principles for scrutable assistants is thus the principal goal of the proposed dissertation. Lastly, as the user can always simply complete the intended action without the IPA, we want to investigate the effects of scrutability on the appeal of IPA usage.

4 RESEARCH APPROACH, METHODS, AND RATIONALE

To tackle the challenges formulated in the research question above, our work follows the Design-Based Research (DBR) approach [14]. This approach had originally been devised for education research [2], but has since been applied in the areas of human-centred computing as well [1]. Returning to previous phases when this is meaningful due to the knowledge gained in later phases (cf. [14]) is integral to the DBR, which is very advantageous given the partially unpredictable nature of interventions in human-centred research [1].

As the focus of the dissertation is human-assistant interaction, user studies are to continuously guide and evaluate our work. We plan to organise semi-structured interviews as well as focus groups akin to those in [11, 25] to better understand expectations regarding scrutable assistants. We would also organise fill-in-the-gap questionnaires (where the participants imagine appropriate responses for given scenarios), similarly as in [39]. In contrast to these studies, our experiments would have a top-down approach and would fixate on the scrutability rather than on general experiences or suppositions. Thereafter, we would create a prototype that would focus on simulating proactive and delegative interaction (rather than on actually performing such services) that participants can evaluate in usage tests akin to those in [21]. Three aspects of *scrutability by design* would be investigated: 1) how scrutability of the assistant's personalisation models, their substantiates, and consequences is realised (what can be questioned, what explanations can be produced, how are the follow-up questions supported); 2) the appropriateness of time and manner to verify current beliefs about the user and the desired actions, with special attention given to information provenance as well as to the potential broader effect of requested actions; and 3) proactive facets through which the assistant initiates a conversation with the user or even performs an action on the user's behalf with no explicit prior request.

5 PROGRESS TO DATE

During the first semester of the PhD, we focused on defining the research area and the research question. An extensive survey of related work has been undertaken. A concrete result of this work is the diagram of the envision assistant interaction design, presented in section 2. In addition, after contacting the corresponding authors of [11] and [39], we analysed raw data from their experiments. Rather than reproducing their approach and extracting common themes, we wanted to investigate to what extent the participants envisioned personalisation, proactivity, delegation, and elements of scrutability. Tables with the most relevant statements and lines of dialogue and further elaboration on the findings can be seen in a git repository at https://gitlab.com/Jovan_NS/envisioned-ipa-scrutability. For example, one of the scenarios in [39] prompted the participants to write down the dialogue with their assistant concerning visiting a local cinema. Participant *155* imagined the IPA to ask the user for a seat position preference, whereas in *105* the IPA just booked seats and noted that they are in the user's "preferred seating area". Therefore, the participants in [39] envisioned that the need for explicit verification would depend on prior interactions. A similar proposition was voiced by an infrequent IPA user in focus group three from [11] who had stated: *I would like Siri to ask me back a question, to clarify something. /…/ I would have liked a question-answer to be more of a dialogue.* Likewise, users in both studies envisioned proactivity (both in terms of initiating conversations and proposing activities) and delegation (negotiating meetings, writing letters and text messages on the user's behalf). Some users in [39] also envisioned elements of scrutability such as studying why a certain suggestion had been given or a proactive action performed. The IPA's awareness of social relations (also depicted in figure 1) is implied in many open scenario dialogues from [39], in which the IPA is capable of proposing, discussing, and elaborating on gift ideas for the user's significant others.

6 DISSERTATION STATUS, EXPECTED NEXT STEPS, AND LONG-TERM GOALS

Jovan Jeromela is a first-year PhD student partaking in a four-year full-time PhD research programme under the supervision of Professor Owen Conlan at Trinity College Dublin. The expected time of graduation is thus the Summer of 2025. With the initial literature survey and the preliminary interaction diagram completed (cf. section 2), the focus of the work at the moment is to theoretically define a framework for the intersection of scrutability [21] and intelligent personal assistants [12]. From the framework and the initial insights gathered from data from [11, 39] (cf. section 5), mock use-cases and scenarios will be devised to serve as the base for semi-structured interviews and a fill-in-the-gap study (cf. section 4). Based on the gained knowledge, a prototype of a scrutable assistant will be designed and implemented. Evaluation results of this prototype interaction studies as well as the prior interviews will then lead to the central contribution of the dissertation: concrete guidelines for the design of scrutable intelligent personal assistants. In addition, evaluation criteria would be defined for assessing how principles of scrutability were integrated with a given intelligent personal assistant's interaction design.

There are thus multiple related goals and envisioned contributions of the dissertation. Firstly, we propose a novel IPA interaction paradigm for the envisioned proactive and delegative IPAs based on scrutability. We also want to provide a formal framework for such assistants that accounts for existing and envisioned issues. Moreover, evaluation criteria to determine the extent to which these issues are alleviated is to be proposed. A further contribution would be to develop a fitting confidence model for assumption verification that considers information provenance and conceivable action consequences. Another point of interest is discovering potential long-term effects of scrutable human-assistant interaction and its effects on the cognitive load, multitasking, sense of trust and rapport, and appeal for IPA use. Lastly, a practical contribution of the thesis is to be an interactive prototype of a scrutable, proactive, and delegative IPA.

ACKNOWLEDGMENTS

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6223.

This work was supported by the ADAPT Centre, funded by Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106; 13/RC/2106P2) and co-funded by the European Regional Development Fund.

We thank Benjamin R. Cowan and Sarah Theres Völkel for their support and permission to access and reanalyse their datasets originally presented in [11] and [39] respectively.

REFERENCES

[1] Christos Amyrotos. 2021. Adaptive Visualizations for Enhanced Data Understanding and Interpretation. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 291–297.

[2] Terry Anderson and Julie Shattuck. 2012. Design-Based Research. *Educational Researcher* 41 (2012), 16–25. https://doi.org/10.3102/0013189X11428813

[3] Pauline M. Berry, Thierry Donneau-Golencer, Khang Duong, Melinda Gervasio, Bart Peintner, and Neil Yorke-Smith. 2017. Evaluating intelligent knowledge systems: experiences with a user-adaptive assistant agent. *Knowledge and Information Systems* 52, 2 (2017), 379–409. https://doi.org/10.1007/s10115-016-1011-3

[4] Giovanni Campagna, Rakesh Ramesh, Silei Xu, Michael Fischer, and Monica S. Lam. 2017. Almond: The Architecture of an Open, Crowdsourced, Privacy-Preserving, Programmable Virtual Assistant. , 341–350 pages. https://doi.org/10.1145/3038912.3052562

[5] H. Chung, M. Iorga, J. Voas, and S. Lee. 2017. "Alexa, Can I Trust You?". *Computer* 50, 9 (2017), 100–104. https://doi.org/10.1109/MC.2017.3571053

[6] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, and Benjamin R Cowan. 2019. The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers* 31, 4 (2019), 349–371. https://doi.org/10.1093/iwc/iwz016

[7] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[8] Owen Conlan. 2005. *The Multi-Model, Metadata Driven Approach to Personalised eLearning Services*. Doctoral dissertation. School of Computer Science and Statistics, Trinity College Dublin.

[9] Owen Conlan, Vincent Wade, Catherine Bruen, and Mark Gargan. 2002. Multi-model, Metadata Driven Approach to Adaptive Hypermedia Services for Personalized eLearning *(Adaptive Hypermedia and Adaptive Web-Based Systems)*. Springer Berlin Heidelberg, 100–111.

[10] Owen Conlan and Vincent P. Wade. 2004. Evaluation of APeLS – An Adaptive eLearning Service Based on the Multi-model, Metadata-Driven Approach *(Adaptive Hypermedia and Adaptive Web-Based Systems)*. Springer Berlin Heidelberg, 291–295.

[11] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What can i help you with?": infrequent users' experiences of intelligent personal assistants. , Article 43 pages. https://doi.org/10.1145/3098279.3098539

[12] 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. https://doi.org/10.1016/j.eswa.2020.113193

[13] Mateusz Dubiel, Martin Halvey, and Leif Azzopardi. 2018. A Survey Investigating Usage of Virtual Personal Assistants. *arXiv pre-print server* (2018). https://arxiv.org/abs/1807.04606

[14] Matthew W Easterday, Daniel Rees Lewis, and Elizabeth M Gerber. 2014. *Design-based research process: Problems, phases, and applications*. Boulder, CO: International Society of the Learning Sciences.

[15] Justin Edwards, He Liu, Tianyu Zhou, Sandy J. J. Gould, Leigh Clark, Philip Doyle, and Benjamin R. Cowan. 2019. Multitasking with Alexa: how using intelligent personal assistants impacts language-based primary task performance. , Article 4 pages. https://doi.org/10.1145/3342775.3342785

[16] SRI International. 2020. 75 Years of Innovation: CALO (Cognitive Assistant that Learns and Organizes). , A research project that developed an enduring personalized cognitive assistant pages. https://medium.com/dish/75-years-of-innovation-calo-cognitive-assistant-that-learns-and-organizes-9f176f2de05d Last access: 2020-03-10.

[17] Charles L Isbell and Jeffrey S Pierce. 2005. An IP continuum for adaptive interface design. In *Proc. of HCI International*. 10.

[18] Eduardo Islas-Cota, J. Octavio Gutierrez-Garcia, Christian O. Acosta, and Luis-Felipe Rodríguez. 2022. A systematic review of intelligent assistants. *Future Generation Computer Systems* 128 (2022), 45–62. https://doi.org/10.1016/j.future.2021.09.035

[19] Judy Kay. 1994. The um toolkit for cooperative user modelling. *User Modeling and User-Adapted Interaction* 4, 3 (1994), 149–196. https://doi.org/10.1007/BF01100243

[20] Judy Kay. 1998. *A scrutable user modelling shell for user-adapted interaction*. Doctoral dissertation. Basser Department of Computer Science, Faculty of Science, University of Sydney.

[21] Judy Kay and Bob Kummerfeld. 2013. Creating Personalized Systems That People Can Scrutinize and Control: Drivers, Principles and Experience. *ACM Trans. Interact. Intell. Syst.* 2, 4, Article 24 (jan 2013), 42 pages. https://doi.org/10.1145/2395123.2395129

[22] Judy Kay, Bob Kummerfeld, and Piers Lauder. 2002. Personis: A Server for User Models *(Adaptive Hypermedia and Adaptive Web-Based Systems)*. Springer Berlin Heidelberg, 203–212.

[23] Bob Kummerfeld and Judy Kay. 2017. User Modeling for the Internet of Things. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, 367–368. https://doi.org/10.1145/3079628.3079658

[24] Yaniv Leviathan and Yossi Matias. 2018. Google Duplex: An AI system for accomplishing real-world tasks over the phone. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html Last access: 2021-11-14.

[25] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. , 5286–5297 pages. https://doi.org/10.1145/2858036.2858288

[26] Christian Meurisch, Maria-Dorina Ionescu, Benedikt Schmidt, and Max Mühlhäuser. 2017. Reference model of next-generation digital 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*. 149–152.

[27] Nikola Mrkšić. 2020. Our Voice Assistant Spoke to Google Duplex. Here's What Happened.... https://www.polyai.com/our-voice-assistant-spoke-to-google-duplex-heres-what-happened/ Last access: 2021-11-21.

[28] Daniel E O'Leary. 2019. GOOGLE'S Duplex: Pretending to be human. *Intelligent Systems in Accounting, Finance and Management* 26, 1 (2019), 46–53. https://doi.org/10.1002/isaf.1443

[29] Neil Peirce, Owen Conlan, and Vincent Wade. 2008. Adaptive educational games: Providing non-invasive personalised learning experiences. In *2008 second IEEE international conference on digital game and intelligent toy enhanced learning*. IEEE, 28–35.

[30] Ricardo Ponciano, Sebastião Pais, and João Casal. 2015. Using Accuracy Analysis to Find the Best Classifier for Intelligent Personal Assistants. *Procedia Computer Science* 52 (2015), 310–317. https://doi.org/10.1016/j.procs.2015.05.090

[31] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. *Voice Interfaces in Everyday Life*. Association for Computing Machinery, Paper 640. https://doi.org/10.1145/3173574.3174214

[32] R. Sarikaya. 2017. The Technology Behind Personal Digital Assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine* 34, 1 (2017), 67–81. https://doi.org/10.1109/MSP.2016.2617341

[33] Silvia Schiaffino and Analía Amandi. 2004. User – interface agent interaction: personalization issues. *International Journal of Human-Computer Studies* 60, 1 (2004), 129–148. https://doi.org/10.1016/j.ijhcs.2003.09.003

[34] Elhadi M Shakshuki and SM Mozammal Hossain. 2014. A personal meeting scheduling agent. *Personal and ubiquitous computing* 18, 4 (2014), 909–922.

[35] Yu Sun, Nicholas Jing Yuan, Yingzi Wang, Xing Xie, Kieran McDonald, and Rui Zhang. 2016. Contextual Intent Tracking for Personal Assistants. , 273–282 pages. https://doi.org/10.1145/2939672.2939676

[36] Yu Sun, Nicholas Jing Yuan, Xing Xie, Kieran McDonald, and Rui Zhang. 2017. Collaborative Intent Prediction with Real-Time Contextual Data. *ACM Trans. Inf. Syst.* 35, 4 (2017), Article 30. https://doi.org/10.1145/3041659

[37] Amrita S Tulshan and Sudhir Namdeorao Dhage. 2018. Survey on virtual assistant: Google assistant, siri, cortana, alexa. In *International symposium on signal processing and intelligent recognition systems*. Springer, 190–201.

[38] Piyush Vashistha, Juginder Pal Singh, Pranav Jain, and Jitendra Kumar. 2019. Raspberry Pi based voice-operated personal assistant (Neobot). In *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 974–978.

[39] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. *Eliciting and Analysing Users' Envisioned Dialogues with Perfect Voice Assistants*. Association for Computing Machinery, Article 254. https://doi.org/10.1145/3411764.3445536

[40] 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), 30. https://doi.org/10.1142/S0218213012500042

[41] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. *Computational Linguistics* 46, 1 (2020), 53–93. https://doi.org/10.1162/coli_a_00368