Please quote as: Zierau, N., Engel, C., Söllner, M., and Leimeister, J. M. (2020) Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda. In 15th International Conference on Wirtschaftsinformatik (WI), pp 99-114.

Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda

Conference Paper · March 2020

DOI: 10.30844/wi_2020_a7-zierau

CITATIONS
3

READS
318

4 authors:

Naim Zierau
University of St.Gallen
6 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Christian Engel
University of St.Gallen
9 PUBLICATIONS **41** CITATIONS
SEE PROFILE

Matthias Söllner
Universität Kassel
173 PUBLICATIONS **905** CITATIONS
SEE PROFILE

Jan Marco Leimeister
University of St.Gallen
919 PUBLICATIONS **8,897** CITATIONS
SEE PROFILE

Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda

Naim Zierau[1], Christian Engel[1], Matthias Söllner[2], and Jan Marco Leimeister[3]

[1] University of St.Gallen, Institute of Information Management, St. Gallen, Switzerland;
{naim.zierau,christian.engel,matthias.soellner,
janmarco.leimeister}@unisg.ch
[2] University of Kassel, Information Systems and Systems Engineering, Kassel, Germany;
soellner@uni-kassel.de
[3] University of Kassel, Information Systems, Kassel, Germany;
leimeister@uni-kassel.de

Abstract. Smart Personal Assistants (SPA) fundamentally influence the way individuals perform tasks, use services and interact with organizations. They thus bear an immense economic and societal potential. However, a lack of trust - rooted in perceptions of uncertainty and risk - when interacting with intelligent computer agents can inhibit their adoption. In this paper, we conduct a systematic literature review to investigate the state of knowledge on trust in SPAs. Based on a concept-centric analysis of 50 papers, we derive three distinct research perspectives that constitute this nascent field: user interface-driven, interaction-driven, and explanation-driven trust in SPAs. Building on the results of our analysis, we develop a research agenda to spark and guide future research surrounding trust in SPAs. Ultimately, this paper intends to contribute to the body of knowledge of trust in artificial intelligence-based systems, specifically SPAs. It does so by proposing a novel framework mapping out their relationship.

Keywords: Trust, Smart Personal Assistant, Conversational Agent, Literature Review, Research Agenda

1 Introduction

In recent years, technologies based on Artificial Intelligence (AI) have matured and are increasingly permeating the professional and private lives of many people [1]. A key area of application represents the proliferation of Smart Personal Assistants (SPA) – computer agents that provide assistance by engaging with users via natural language. SPAs, sometimes also referred to as intelligent agents or conversational agents, are now applied in a wide area of usage scenarios [2]. These agents including Amazon`s *Echo*, Google`s *Google Assistant* and Apple`s *Siri* fundamentally influence the way in which individuals complete tasks, search for information, consume product and service offerings, and interact with organizations [3]. Thus, it is foreseen that SPAs will become

15th International Conference on Wirtschaftsinformatik,
March 08-11, 2020, Potsdam, Germany

a daily companion for a wide range of users – for instance, the number of worldwide users of these agents is predicted to rise to almost 1.8 billion in 2021, which reflects the social and economic impact SPAs exhibit on a global scale [4].

However, the potential of SPAs can only be realized if users feel that they can trust the new medium [5]. Since SPAs rely on AI, they face corresponding problems in terms of user acceptance. In a wide range of domains, such as finance or medicine, the acceptance towards the recommendations of AI-based systems, is below 60 percent [6]. These findings are in line with a recent survey by Forbes which has shown that 41.5 percent of U.S. consumers do not trust any AI-infused digital assistants [7]. These numbers highlight that trust is paramount for helping users overcome adaption resistance due to perceptions of risk and uncertainty [8]. At the same time, however, the unique characteristics of AI-infused systems including their opaqueness, embedded biases, and autonomous nature may not only make it difficult to garner trust, but afford risks based on non-reflective reliance [9]. Understanding the nature and antecedents of trust in intelligent agents such as SPAs as a combined unit of analysis is, therefore, a necessary requirement for both IS researchers and practitioners who aim to successfully design and deploy SPA-based systems. As we will show in this paper, there has been a steep increase in publications on the topic of trust in SPAs. For the progress of a young and emerging research field, it is important to present previous research coherently and transparently such that the important research streams are highlighted, and their interrelationships and theoretical basis are presented [10]. So far, to the best of our knowledge, an overview and conceptual structuring of the combined research field of trust and SPAs does not exist. This results in a lack of terminological clarity and theoretical integration of important concepts. Our Systematic Literature Review (SLR) addresses this gap by contributing to creating a common language and structuring the conceptual basis of trust in SPAs. Thus, we intend to answer the following research question:

Which research streams conceptually constitute the research field of trust in Smart Personal Assistants?

Overall, this SLR intends to contribute to the body of knowledge of trust in AI-based systems in general and trust in SPAs in particular by providing an integrated theoretical framework of the latter. This framework suggests that trust in SPAs can be distinguished into three distinct research perspectives. Based on this conceptual grounding, we propose a preliminary research agenda to trigger and guide future research in this nascent field.

The remainder of this paper is structured as follows: First, we provide a brief summary of the theoretical background on both SPAs and trust. Subsequently, we conduct an SLR according to [11] and [12, 13] to provide an overview and structuration of the field of trust in SPAs. Furthermore, we present an integrated theoretical framework incorporating all research streams within the joint research field of trust in SPAs that we could conceptionally derive from literature. Finally, based on the state of the art in the particular research streams a preliminary research agenda is proposed.

2 Theoretical Foundation

To provide a foundation for our integrative review and our discussion of our research agenda, we first begin by defining key concepts including SPAs and trust.

2.1 Smart Personal Assistants

Research on SPAs is not a new field per se. However, it recently gained wide prominence in the broader public. In the past, these systems were almost exclusively studied as expert systems giving "intelligent advice" within a limited set of highly specified use cases [14]. However, due to the emergence of technologies associated with AI such as Machine Learning (ML), voice recognition, and natural language processing, new generations of SPAs have emerged such as Amazon`s *Echo*, Google`s *Google Assistant* and Apple`s *Siri*. They can now be applied in a wide range of use cases spanning from everyday tasks such as ordering consumer goods to more specialized tasks such as helping users track their expenses [15]. There are various terms describing these assistants – for example conversational agent, chatbot, virtual assistant, digital assistant -, who all are based on the idea of interacting with users via natural language (e.g., [2]). In this regard, in order to cover several types of systems, we refer to SPAs as AI-based systems embedded in personal technologies designed to assist users by interacting in a text or voice-based conversation [e.g., 16].

From a sociotechnical perspective and compared to other entities of IS, the novelty of SPAs lies in two major aspects, which potentially fundamentally affect user perceptions (i.e., trust): the way in which SPAs *interact* with users as well as the degree of *intelligence* they employ thereby [17]. Thus, they manifest the characteristics of two distinct but interrelated system classes - interactive and intelligent IS: on the one hand, based on anthropomorphic features interactions with SPAs are increasingly moving towards the level of interpersonal communications [18], including the establishment of emotional bonds [19]. At the same time, the pervasiveness of invasive technologies embedded in these systems as well as their autonomous nature raises questions of accountability and data security [5]. Moreover, the rising intelligence of SPAs comes with issues of interpretability of their behavior through users. This may explain why users still only reluctantly adapt and use these systems despite their potential [9].

2.2 Trust in Information Systems

One of the most important factors driving the adaption and use of complex and increasingly automated technical artifacts such as SPAs is trust. Traditionally, trust research in IS has been focusing on studying relationships among human beings that are mediated by an IS. However, due to developments such as increasing automation [9], IS have itself become an integral part of trust relationships in a wide area of usage scenarios. Automated systems are not only used to mediate trust relationships between human beings, but to support their users in achieving specific goals, thereby exhibiting agency on their behalf. Thus, these systems become themselves trustees in a trust relationship between the human user and a respective IS [20]. According to the trust definition of [21], users therefore need to exhibit willingness to be vulnerable to the

actions of an autonomous IS "*based on the expectation that the other [i.e., SPA] will perform a particular action important to the trustor [i.e., user], irrespective of the ability to monitor or control that other party [i.e., SPA]*" (Mayer et al. 1995, p. 712) [21].

In the past years, IS got increasingly interactive – specifically in regards to exhibiting anthropomorphic features – and intelligent – based on advancements in the domain of ML [17], which is why trust in these systems may not anymore entirely be explainable with current insights. On the one hand, SPAs are able to act and interact in an increasingly human way. Thus, the boundary between man and machine becomes increasingly blurred from a user perspective, which has important implications for theory and practice. Especially, the suitability of the theoretical basis on which trust in the system is studied becomes a relevant question. On the other hand, AI-infused systems raise the opaqueness and complexity for the user [1], therefore magnifying the issue of trust. It is argued that building trust is an essential means to address complexity and uncertainty because humans cannot have complete knowledge of most systems' inner processes. Additionally, as these systems continuously learn and adapt their behavior accordingly, there is an increased need to study trust in these systems from a longitudinal perspective [22]. In this work, we focus on SPAs as one concrete instantiation of AI-based IS, which we link to trust research as one combined unit of analysis based on a SLR.

3 Method

We conduct a SLR within the research field of trust in SPAs according to the principles and practices suggested by [11] and [12, 13]. Overall, the scope of the SLR can be structured along the dimensions of process, source, coverage, and techniques [13]: based on a *sequential search process* in four data bases, publications from IS literature and related fields such as business and human computer interaction as a *source* are identified. The literature search aims to reach a *representative coverage* of the distinct perspectives on the research field of trust in SPAs. Therefore, to establish the basis for the analysis and conceptualization, we used a *comprehensive* set of techniques (i.e., keyword search, backward search, and forward search). To reach a high level of reproducibility and transparency of our research, we describe in this section the single methodical steps that we undertook:

Selection of search strings: Aiming at covering literature that focuses on the combined unit of analysis of trust and SPAs, we select ("smart" OR "intelligent" OR "cognitive") AND ("assistant" OR "system") AND "trust" as the initial search string that we use as a starting point for the literature search process. The initial search string is constructed rather broad taking into account a variety of key word permutations to not neglect relevant research. It needs to be noted here that we apply the search string considering the particular variations the single keywords can exhibit such as singular and plural, and the use of hyphens or no hyphens. As we seek for papers that conduct research with a focus on trust and SPAs as a combined unit of analysis, the single parts

of the initial search string should appear in close proximity to each other in the papers. Thus, we choose to search in title, abstract, and keywords of publications.

Selection of databases: We apply a respective search in IS databases that contain a variety of IS journals and conferences to not restrict our search scope upfront and to cover more recent research as we assume that trust in SPAs is a young and emerging research field. Covering the latter aspect would not be assured by a journal-only-based literature search as processing journal reviews takes significantly longer than reviewing for conference proceedings. Consequently, we select five databases covering a wide range of Information Systems (IS) literature to assure a representative coverage of our literature search. The databases that we select are in particular the database of the Association for Information Systems (AISel), EBSCO, Sciencedirect, the database of the Association for Computing Machinery (ACM), and ProQuest.

Refinement of search strings: During our literature search process that starts with the search string described above, we iteratively refine and adapt the keywords to take into account the learnings from and our enhanced understanding of the field that emerges during our SLR. The final search string used in the SLR is ("smart" OR "intelligent" OR "cognitive" OR "conversational" OR "AI") AND ("assistant" OR "system" OR "agent" OR "application") AND "trust". It extends the initial search string by taking into account a larger set of synonyms used to describe SPAs and concepts related to the concept of "smartness" such as Artificial Intelligence (AI). The literature search conducted in title, abstract, and keywords of the publication texts, results in 1,168 hits, which still comprises duplicates and potentially irrelevant papers.

Selection of papers: During a first screening step, we focus on screening the title, abstract, keywords, and research domain of the papers and only consider papers that use English as their publication language. This process results in 45 papers, which undergo a detailed full-text screening that determines if a paper is finally considered relevant for deeper analysis. We label a paper as "relevant for further analysis" if it conducts research with a central focus on trust and SPAs as a combined unit of analysis. Consequently, during full-text screening, we remove literature that only marginally or trivially issues the intersected unit of analysis "trust and SPAs", such as for example paper that measured trust as one of many variables, but did not discuss this effect further. This leads to 35 relevant hits. After removing duplicates that stem from choosing a database-oriented literature search, 32 papers remain to be considered for further analyses. Table 1 provides an overview of the total hits and relevant search results structured along the particular databases to account for reproducibility and transparency of the SLR.

Table 1. Results of the Literature Search Process

	Databases									
	AISeL		**EBSCO**		**Sciencedirect**		**ACM**		**ProQuest**	
Search String	Hits	Relevant	Hits	Relevant	Hits	Relevant	Hits	Relevant	Hits	Relevant
("smart" OR "intelligent" OR "cognitive" OR "conversational" OR "AI") **AND** ("assistant" OR "system" OR "agent" OR "application") **AND** "trust"	79	8	142	4	480	4	82	14	385	5
Number papers selected for further analysis from 1168 screened papers	With Duplicates:		35	+	**3** Forward Search				=	**50**
	Without Duplicates:		**32**		**15** Backward Search					

To seize the benefits of a comprehensive set of search techniques, backward search (+15 additional papers) and forward search (+3 additional papers) are conducted on top of the literature retrieved from using the search string in the five databases above [12, 13, 23]. Finally, this results in an overall number of 50 papers that are analyzed and conceptualized in a detailed manner in this paper.

Paper Analysis and Conceptualization: We analyze the 50 papers identified to be relevant for this work from a concept-centric perspective. Thus, according to [11] a concept matrix is created based on the literature search results. Respectively, all papers are analyzed according to the focal concepts used to investigate the combined unit of analysis "trust in SPAs", according to the applied research method, and according to the contributions reached for theory and practice. This endeavor intends to conceptualize the distinct central research streams that constitute the combined research field of trust in SPAs, thus providing an integrated view on the latter. We use an iterative process guided by cross-validation discussions between two researchers, in which we analyze and aggregate the distinct concepts identified in the retrieved literature to higher-order, more abstract concepts that are merged into particular meta-perspectives on the research field of trust in SPAs. By iteratively cross-validating the conceptual insights that are abductively created form literature we aim for reproducible, transparent, and valid research results. However, we have to acknowledge that conceptualizing literature always contains a certain level of non-erasable subjectivity.

4 Results

Figure 1. shows that the number of identified publications has been steeply growing during the last years. The youngest paper is from 2019 and the oldest paper from 1999, when initial interest rose in light of the first expert systems being used in organizational contexts. The majority of papers has been published within the last two years, which supports our initial assumption that trust in SPAs represents an emerging research field.

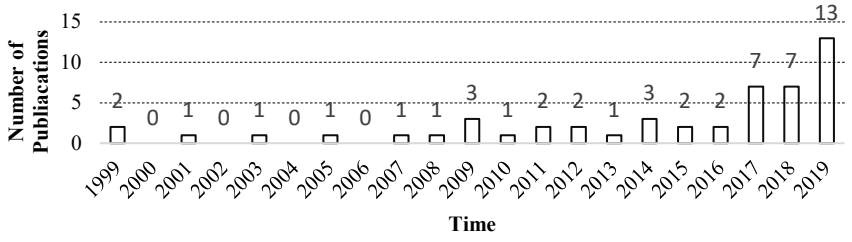

Figure 1. Number of Publications over Time

Based on a concept-centric analysis of identified papers, we were able to identify three main research perspectives, which constitute the research field of trust in SPAs: User Interface-Driven Trust (UIDT), Interaction-Driven Trust (IDT), and Explanation-

Driven Trust (EDT). The research perspective with the most papers was IDT (26) followed by EDT (18) and UIDT (16).

Table 2. Definition and conceptual boundaries of the three perspectives on trust in SPAs

	Definition	**Conceptual Boundaries**
User Interface-Driven Trust (UIDT) in SPAs	UIDT in SPAs research deals with static design features of SPAs such as haptics and audio-visuals towards enhancing trust in the latter.	The scope of UIDT in SPAs research encompasses the effect of all components that make up the static appearance of SPAs on trust, including its audio, visual and haptic representation. This does not include the actions performed by the SPA.
Interaction-Driven Trust (IDT) in SPAs	IDT in SPAs research addresses the design of interactions between the user and the SPA on a timeline to form trust in the latter.	The scope of IDT in SPAs research encompasses the effect of the actual interaction behavior of SPAs with the user (and vice versa) on trust, including verbal and non-verbal behavior. This does not include the static representation of the SPA.
Explanation-Driven Trust (EDT) in SPAs	EDT in SPAs research focuses on creating trust in SPAs by varying the degrees of understandability and transparency provided to the user.	The scope of EDT in SPAs encompasses the effect of information and self-disclosure on trust in the SPA. This does not include the static representation of the SPA.

Overall, we strive for a precise and unambiguous description of the different research perspectives, in order to allow for a robust categorization of identified publications. Therefore, as Table 2 shows, we formulated concise definitions for the derived research streams and their respective conceptual boundaries. Furthermore, we highlight their interrelations according to the principles of construct clarity in order to achieve clear differentiations between the streams and to maintain internal consistency within the streams [24].

To make this conceptualization of trust in SPAs more graspable, we would like to discuss the different perspectives using the example of a Smart Home Assistant. Those agents can for example be used to obtain a wide range of information, to order products or control other smart home devices. UIDT addresses questions such as should the assistant speak with a male or female voice to the user. IDT, on the other hand, focuses on questions as to whether the assistant should communicate in a friendly or more emotionally neutral way. Finally, EDT is concerned with question as to whether and how the assistant should explain recommendations or actions takes such as automatically ordering a product.

In the following sub sections, we present the conceptually derived research streams that constitute the combined research field of trust in SPAs by elaborating on the particular research perspectives and sub research streams of the latter.

4.1 User Interface-Driven Trust in Smart Personal Assistants

We refer to *User Interface-Driven Trust (UIDT) in SPAs* as trust emerging from the static design features of SPAs such as haptics and audio-visuals. Thereby, we found authors to mainly study the effects of visual and auditory design features.

One sub-research stream deals with the effect of visual design features, which refers to interface-transmitted cues that can be perceived visually [16]. A majority of studies within this sub-stream focuses on the effect of SPA embodiment on user trust [e,g., 6, 7]. For instance, it was found that humans perceive trust relationships with humans differently than with avatars such that humans are better in predicting the trustworthiness of humans than of avatars. However, the trustworthiness learning rate is similar, whether interacting with humans or avatars [27]. In regards to the effect of embodiment, there are somehow mixed results. In a study conducted with children, they rated the most visually embodied character as the most trustworthy SPA in a game scenario [28]. Moreover, it was found in another study that embodied SPAs are associated with greater trust resilience, a higher resistance to breakdowns in trust, and that these effects were magnified by greater uncertainty. However, once the different SPAs incorporated human-like trust repair behavior the effect was largely erased [26]. Moreover, in a survey ranking different trust mechanisms it was found that visual appearance is the least important for users of robo-advisory services [29]. These findings suggest that the effect of appearance is highly dependent on the context. This is supported by another study showing that gender fit between the avatar and the user may present an important antecedent for trust formation [25]. Another study reports that male avatars are being experienced as more trustworthy as female avatar in interview scenarios [25]. Finally, authors investigated the effects of demeanor and could for instance show that a smiling SPA increased trustworthiness [30].

Another sub research stream in the dimension of UIDT in SPAs is related to the effect of auditory design features on user trust, which refer to cues that can be perceived audibly [16]. Most authors, studied the effects of response modality, such as the questions if and which kind of voice is perceived as more trustworthy compared to using a chat interfaces [e.g., 12–14]. Generally, identified studies show that human-(like) voice carries important cues that evoke perceptions of social presence resulting in higher levels of perceived trustworthiness. Thereby, there are a number of SPA speech properties such as pitch contour and flanging increments that classify speech along a machine-to-human spectrum [33]. Along this spectrum, several authors found that human speech had higher ratings of trust than machine-like speech, which also translated into higher compliance with the SPAs recommendations [32, 34]. Text-to-Speech (TTS) voice leads to lower levels of social presence and was perceived as less trustworthy than text. [32]. Moreover, authors found that expression modality seems only to be relevant for initial trust formation, since it was found that vocal pitch only influenced trust perception early in the interaction process [30].

4.2 Interaction-Driven Trust in Smart Personal Assistants

The identified research stream of *Interaction-Driven Trust (IDT) in SPAs* addresses the design of interactions between the user and the SPA on a timeline to form trust in the latter. Thereby, the authors generally discuss different verbal and non-verbal features that make up the behavior of the SPA employing a processual perspective:

The first sub research stream (i.e., verbal features) can be divided into contributions that focus on content and on conversational style [16]. Thereby, we found that most publications address the latter. In regards to content, initial work identified the use of small talk in the SPAs interaction with the user as an effective strategy for building trust [14]. Moreover, being critical to the users wishes was mentioned as another trust-building strategy [38]. Studies addressing conversational styles in general found a positive effect of relational strategies. Thus, socially oriented SPAs interjecting an informal and friendly conversational style lead to enhanced perceptions of interactivity and trust in the system [35]. In comparison, task-oriented SPAs without any deliberate social-emotional capabilities were trusted less. This effect seems to remain stable over time, since even after weeks of interacting with a SPA, relational agents were perceived as significantly more trustworthy than task-oriented agents [36]. Thereby, SPA responsiveness was found to be a major antecedent [37]. One study highlights the importance for SPAs to learn from human-human interactions on how to build trust, but recommends to not just mimic it. Instead, human-SPA conversations may need to be treated as a new genre of interaction as trustworthiness was discussed exclusively in utilitarian terms by interviewees in the same study – responses related to security, privacy, and transparency over emotional trust [39].

The second sub research stream (i.e., non-verbal features) investigated primarily the effect of the degree of autonomy (i.e., proactivity) exhibited by the SPA. Thus, it was shown that autonomy may lead to a more human-like appearance evoking a feeling of social presence and therefore inducing trust [40]. However, another study suggests that there is a need for a fit between the SPA's autonomy and user preferences to maintain trust. In a smart home environment it was shown that the users level of comfort with the SPA's level of autonomy was depended on personality types and task characteristics [41]. To give users a perception of control different mechanisms are proposed – personalization-driven control, task-driven control, and mechanism that allow direct control [42]. Specifically, providing users with different alternatives has proven to provide users with a sense of control, which increases trust in the SPA [43].

4.3 Explanation-Driven Trust in Smart Personal Assistants

Explanation-Driven Trust (EDT) in SPAs refers to varying the degrees of understandability and transparency provided to the user. Within the analyzed papers, we could identify two sets of strategies that were applied to increase transparency and understandability – explanations and interactive machine learning [44].

In the first sub-research stream, authors investigate the effect of different types of explanations for increasing trust in SPAs. Generally self-disclosure was deemed an effective measure to signal trustworthiness to users [40–43]. In this regard, it was

suggested that the SPA should frequently communicate what kind of data it needs to generate value for the user [42]. Apart from one study reporting that only users with low task familiarity are susceptible for explanations provided by SPAs [45], most of the studies see generally a positive effect in respect to trust formation [e.g., 34] by raising transparency and focusing the users focus on system ability [22]. Even providing placebic explanations was found to raise trust in the SPA [46]. However, informative explanations were more effective in building trust. Thereby, it was shown that even laypersons are able to understand the basic logic behind ML-models. In this regard, rule-based and keyword-based ML-models ranked high in understandability, while similarity-based ML models were harder to grasp for users resulting in less trust into the actions of the SPA [47]. An often-addressed topic is the analysis of explanation types for improving the intelligibility of SPAs. In this regard, it was proposed that explanations based on justifications following a structured argumentation approach and addressing the reasoning of the system's behavior would evoke higher perceptions of trustworthiness [35-36]. Finally, the effect of explanations on user trust was investigated from a longitudinal perspective. It was shown that explanations increase trust levels in the short term, but have no effect in the long term. However, without any explanation shown trust levels in regards to the SPA degenerated steadily [22].

The second sub-research stream, smaller in numbers than the previous one, is concerned with the ability of SPAs to increase system transparency by enabling users to influence its recommendations (i.e., interactive ML) [44]. Thus, users were given the opportunity to set constraint thresholds or to change algorithm weights [43]. Thereby, it was shown that providing the user with the opportunity to test the SPA is one of the most effective means to increase trust [29].

5 Discussion and Development of a Research Agenda

In this section, we aim to discuss the contributions of our SLR and propose a preliminary research agenda that provides first promising points for future research on trust in SPAs and illustrates how they can be positioned based on our conceptualization.

As our SLR shows, the three identified research streams enable a distinct perspective on studying trust in SPAs, which also relates to the theoretical lenses that are applied. While UIDT and IDT are linked primarily to enhance trust by creating a sense of social presence, EDT mainly relies on perceived transparency. Additionally, each of the research streams can be distinguished by its time perspective. While UIDT is important to build initial trust as the interface represents the first point of interaction, IDT and EDT additionally imply a longitudinal perspective. However, this conceptualization only serves as a starting point for the elaboration of the different theoretical lenses, which can be applied to study trust in SPAs. This is important to allow researchers to apply a more nuanced perspective when studying trust in SPAs.

Building on the insights that have been gained through this SLR and by linking our conceptualization to our theoretical background, we propose the following preliminary research agenda as presented in *Table 3*. Thereby, both possible questions that arise from the research streams and some overarching research questions are addressed:

Table 3. Preliminary research agenda on user trust in SPAs

	Research Opportunities	**Corresponding Research Questions**
UIDT	Haptics are associated with user perceptions (i.e., trust) [16]; however, this effect has not yet been sufficiently addressed in the context of SPAs.	How do haptics (e.g., temperature, tactile touch) influence trust?
	Interfaces are usually characterized by multiple features; however, the effect of specific feature combinations has not yet been adequately explored.	How do specific feature combinations (i.e., based on different system archetypes [15]) compare in regards to user trust?
IDT	As our SLR has shown, content may significantly influence trust, but still represents an area that has not yet been sufficiently addressed.	How do different content features (i.e., praise [16]) of a conversation affect trust?
EDT	As new types of explanations emerge based on technologic advances [2], there is a need to study their efficiency in regards to creating trust.	How do different types and instantiations of explanations (i.e., attribution-, example- or model-based explanations affect trust?
	Although interactive ML has been identified as a major source of trust [29], there are still few insights on the effect of different types of interactive ML [44] on trust.	How do different types of interactive ML affect trust?
Overarching	There are few insights on the relative importance of the identified types of trust (e.g., our conceptualized research streams).	How do different types of trust (e.g., UIDT, IDT, EIDT) compare in regards to their effectiveness for building trust?
	Different applications of SPAs may have different presuppositions for trust formation [18]; however, the impact of context remains under-researched.	How do contextual factors (e.g., user group, time) influence the effect of different SPA elements and characteristics on trust?
	There is a lack of design-oriented studies [18] and successful SPA designs that foster trust.	How to leverage theory to design more trustworthy SPAs?

All in all, eight research questions emerged both on the level of the individual research streams as well as some overarching research questions, which bears to the relevance of the topic of trust in SPAs. Across all research streams, we identified research opportunities in regards to the influence of specific features. Moreover, as the use context of SPAs may entail different presuppositions for trust formation, especially based on their high degree of adaptivity [1], we propose to increasingly investigate the influence of contextual factors. Especially, the influence of time has not yet been sufficiently addressed. A good starting point could be to analyze which design features

are important for building initial trust in SPAs and which are more important to uphold trust in the long-term. Additionally, we need to create insights across different domains on how to design SPAs to increase their trustworthiness (e.g., [18]). Finally, from a methodological perspective, as most of the identified studies are based on laboratory experiments, we recommend to increasingly use field experiments to ensure external validity and, thus, to be able to provide stronger insights for practitioners.

In sum, these emerging agenda points may serve as a first foundation for studying trust in SPAs.

6 Limitations

Although we attempted to analyze the identified literature on trust in SPAs as rigorously as possible, there are a number of limitations to this SLR. First, of course the scope of our SLR is not fully exhaustive. However, in order to reach representativity, we chose to conduct a database-oriented search instead of a journal-based search. This enabled us to also consider conference proceedings, which include recent publications, which are especially important when the analyzed research field is still young and emerging such as research on trust in SPAs. We restricted the keyword search to title, abstract, and keywords, since we aimed for identifying publications where the keywords appeared in close proximity to each other for the combined unit of analysis was trust in SPAs. Therefore, it can be argued that the initial screening is limited in scope, but we thoroughly analyzed identified publications based on a concept-centric approach. Moreover, we conducted a rigorous back and forward search. Finally, we did not consider the downstream effects of trust in SPAs (e.g., possible negative effects of too much trust [49]), which may represent an interesting research avenue for future literature reviews in this area.

7 Conclusion

In this paper, we conducted a Systematic Literature Review (SLR) on trust in Smart Personal Assistants (SPA). SPAs bear an immense potential for economic and societal impact, since they fundamentally change the way how individuals perform tasks, use services and interact with organizations. However, they are only hesitantly adopted by users amongst others due to a lack of trust. Building on a concept-centric analysis of 50 publications, we derived three main research perspectives, which constitute the research field of trust in SPAs: User Interface-Driven Trust (UIDT), Interaction-Driven Trust (IDT), and Explanation-Driven Trust (EDT). UIDT deals with static design features of SPAs such as haptics and audio-visuals towards enhancing trust in the latter. IDT refers to trust that is elicited by the design of events between the user and the SPA on a timeline. While UIDT and IDT in SPAs focus on creating a sense of social presence, which has been identified as an important antecedent of trust, EDT in SPAs aims at creating a sense of transparency by varying the degrees of understandability and transparency provided to the user. Based on the results of our analysis, we propose a preliminary research agenda to spark and guide future research in this nascent field. We

intend to contribute to the body of knowledge of trust in Artificial Intelligence (AI)-based systems by proposing an integrated theoretical framework and associated research agenda for studying trust in SPAs. This framework may serve as a starting point for the elaboration on different theoretical lenses which can be applied to study trust in SPAs from a more nuanced perspective. Ultimately, this may also enable practitioners to build more trustworthy SPAs as we introduce new terminology that facilitates the sharing of design knowledge on the effects of different features on trust.

References

1. Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., Söllner, M.: AI-Based Digital Assistants. Bus. Inf. Syst. Eng. (2019).
2. Rzepka, C., Berger, B.: User Interaction with AI-enabled Systems: A systematic review of IS research. In: Pro. 39th Int. Conf. Inf. Syst. (2018).
3. McLean, G., Osei-Frimpong, K.: Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. Comput. Human Behav. 99, 28–37 (2019).
4. Tractica: Worldwide Usage of Virtual Digital Assistant https://de.statista.com/statistik/daten/studie/620321/umfrage/ nutzung-von-virtuellen-digitalen-assistenten-weltweit/ (Accessed: 16.08.2019).
5. Cowan, B.R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., Bandeira, N.: "What can i help you with?": Infrequent users' experiences of intelligent personal assistants. In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (2017).
6. Jensen, M.L., Lowry, P.B., Burgoon, J.K., Jay, F., Jensen, M.L., Lowry, P.B., Burgoon, J.K., Jr, J.F.N.: Technology Dominance in Complex Decision Making : The Case of Aided Credibility Assessment Technology Dominance in Complex Decision Making : The Case of Aided Credibility Assessment. J. Manag. Inf. Syst. 27, 175–202 (2010).
7. Krogue, K.: Artificial Intelligence is Here to Stay, but Constumer Trust is a Must for AI in Business. Forbes. https://www.forbes.com/sites/kenkrogue/2017/09/11/artificial-intelligence-is-here-to-stay-but-consumer-trust-is-a-must-for-ai-in-business/#2cc10a2b776e (Accessed: 16.08.2019).
8. Mc Knight, D.H., Choudhury, V., Kacmar, C.: Developing And Validating Trust Measure for E-Commerce: An Integrative Typology. Inf. Syst. Res. 13, 334–359 (2002).
9. Lee, J.D., See, K.A.: Trust in Automation: Designing for Appropriate Reliance. Hum. Factors. 46, 50–80 (2004).
10. Torraco, R.J.: Writing Integrative Literature Reviews: Guidelines and Examples. Hum. Resour. Dev. Rev. 4, 356–367 (2005). https://doi.org/10.1177/1534484305278283.
11. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future : Writing a literature review. MIS Q. 26, 13–23 (2002).
12. vom Brocke, J., Simons, A., Niehavens, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Docmeunting the Literature Search Process. In: 17th Eur. Conf. Inf. Syst.. pp. 2206–2217 (2009).
13. vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R.: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. Commun. Assoc. Inf. Syst. 37, 205–224 (2015).
14. Gregor, S., Benbasat, I.: Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. MIS Q. 23, 497 (2006).

15. Knote, R., Janson, A., Söllner, M., Leimeister, J.M.: Classifying Smart Personal Assistants: An Empirical Cluster Analysis. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. (2019).
16. Feine, J., Gnewuch, U., Morana, S., Maedche, A.: A Taxonomy of Social Cues for Conversational Agents. Int. J. Hum. Comput. Stud. (2019).
17. Maedche, A., Morana, S., Schacht, S., Werth, D., Krumeich, J.: Advanced user assistance systems. Bus. Inf. Syst. Eng. 58, 367–370 (2016).
18. Pfeuffer, N., Benlian, A., Gimpel, H., Hinz, O.: Anthropomorphic Information Systems. Bus. Inf. Syst. Eng. 61, 523–533 (2019).
19. Purington, A., Taft, J.G., Sannon, S., Bazarova, N.N., Taylor, S.H.: "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. In: Extended Abstracts of the 2017 CHI Conference on Human Factors in Computing Systems. (2017).
20. Söllner, M., Hoffmann, A., Leimeister, J.M.: Why different trust relationships matter for information systems users. Eur. J. Inf. Syst. 25, 274–287 (2016).
21. Mayer, R.C., Davis, J.H., Schoorman, D.F.: An Integrative Model of Organizational Trust. Acad. Manag. Rev. 20, 709–734 (1995).
22. Holliday, D., Wilson, S., Stumpf, S.: User Trust in Intelligent Systems: A Journey over Time. In: Proceedings of the 21st International Conference on Intelligent User Interfaces. pp. 164–168 (2016). https://doi.org/10.1016/j.ijid.2016.10.011.
23. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. - Manag. Inf. Syst. 26, 3 (2002).
24. Suddaby, R.: Editor's Comments : Construct Clarity in Theories of. Acad. Manag. Rev. 35, 346–357 (2010).
25. Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., Patton, M.W.: Embodied Conversational Agent-Based Kiosk for Automated Interviewing. J. Manag. Inf. Syst. 28, 17–48 (2011).
26. de Visser, E.J., Monfort, S.S., McKendrick, R., Smith, M.A.B., McKnight, P.E., Krueger, F., Parasuraman, R.: Almost human: Anthropomorphism increases trust resilience in cognitive agents. J. Exp. Psychol. Appl. 22, 331–349 (2016).
27. Riedl, R., Mohr, P.N.C., Kenning, P.H., Davis, F.D., Heekeren, H.R.: Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory. J. Manag. Inf. Syst. 30, 83–114 (2014).
28. Druga, S., Williams, R., Breazeal, C., Resnick, M.: "Hey Google is it OK if I eat you?" In: Proceedings of the 2017 Conference on Interaction Design and Children. (2017).
29. Mesbah, N., Olt, C.M., Tauchert, C., Buxmann, P.: Promoting Trust in AI-based Expert Systems. In: Proceedings of the 25th Americas Conference on Information Systems (2019).
30. Elkins, A.C., Derrick, D.C.: The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents. Gr. Decis. Negot. 22, 897–913 (2013).
31. Schroeder, J., Schroeder, M.: Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018).
32. Qiu, L., Benbasat, I.: Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. J. Manag. Inf. Syst. 25, 145–182 (2009).
33. Muralidharan, L., de Visser, E.J., Parasuraman, R.: The effects of pitch contour and flanging on trust in speaking cognitive agents. In: Extended Abstracts of the 2014 CHI Conference Human Factors in Computing Systems - CHI '14. (2014).

34. Yu, Q., Nguyen, T., Prakkamakul, S., Salehi, N.: “I Almost Fell in Love with a Machine”: Speaking with Computers Affects Self-disclosure. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. (2019).
35. Chattaraman, V., Kwon, W.S., Gilbert, J.E., Ross, K.: Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. Comput. Human Behav. 90, 315–330 (2019). https://doi.org/10.1016/j.chb.2018.08.048.
36. Bickmore, T.W., Picard, R.W.: Establishing and Maintaining Long-Term Human-Computer Relationships. ACM Trans. Comput. Interact. 12, 293–327 (2005).
37. Wong-Villacres, M., Evans, H., Schechter, D., DiSalvo, B., Kumar, N.: Consejero automatico: Chatbots for Supporting Latino Parents’ Educational Engagement Marisol. In: Proc. Tenth Int. Conf. Inf. and Com. Techn. and Dev.(2019).
38. Vaccaro, K., Agarwalla, T., Shivakumar, S., Kumar, R.: Designing the Future of Personal Fashion. In: Proceedings of the 2018 CHI Conference Human Factors in Computing Systems. (2018).
39. Clark, L., Munteanu, C., Wade, V., Cowan, B.R., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C.: What Makes a Good Conversation? In: Proceedings of the 2019 CHI Conference Human Factors in Computing Systems (2019).
40. Lee, J.G., Kim, K.J., Lee, S., Shin, D.H.: Can Autonomous Vehicles Be Safe and Trustworthy? Effects of Appearance and Autonomy of Unmanned Driving Systems. Int. J. Hum. Comput. Interact. 31, 682–691 (2015). h
41. Hammer, S., Wißner, M., André, E.: Trust-based decision-making for smart and adaptive environments. User Model. User-adapt. Interact. 25, 267–293 (2015).
42. Cesta, A., D’aloisi, D.: Mixed-Initiative Issues in an Agent-Based Meeting Scheduler. Comput. Model. Mix. Interact. 229–262 (1999).
43. Cummings, M.L., Buchin, M., Carrigan, G., Donmez, B.: Supporting intelligent and trustworthy maritime path planning decisions. Int. J. Hum. Comput. Stud. 68, 616–626 (2010).
44. Meza Martínez, M.A., Nadj, M., Maedche, A.: Towards an Integrative Theoretical Framework of Interactive Machine Learning Systems. Proc. 27th Eur. Conf. Inf. Syst. (2019).
45. Schaffer, J., O’donovan, J., Michaelis, J., Raglin, A., Höllerer, T.: I Can Do Better Than Your AI: Expertise and Explanations ACM Reference Format. In: Proceedings of the 24th ACM International Conference on Intelligent User Interfaces. pp. 240–251 (2019).
46. Eiband, M., Buschek, D., Kremer, A., Hussmann, H.: The Impact of Placebic Explanations on Trust in Intelligent Systems. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19. (2019).
47. Stumpf, S., Rajaram, V., Li, L., Wong, W.K., Burnett, M., Dietterich, T., Sullivan, E., Herlocker, J.: Interacting meaningfully with machine learning systems: Three experiments. Int. J. Hum. Comput. Stud. 67, 639–662 (2009).
48. Lim, B.Y., Dey, A.K., Avrahami, D.: Why and Why Not Explanations Improve the Intelligibility of Context-Aware Intelligent Systems. In: Proceedings of the 2009 CHI Conference Human Factors in Computing Systems. pp. 2119–2128 (2009).
49. Benlian, A., Klumpe, J., Hinz, O.: Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. Inf. Syst. J. (2019).