Demystification of Technology

Empowering Consumers to Access and Visualize Voice Interaction Data

Dominik Pins, Fatemeh Alizadeh, Alexander Boden, Sebastian Zilles, and Gunnar Stevens

Abstract *Voice assistants (VAs) in households are becoming increasingly commonplace, with many users expressing their appreciation of the devices' convenience. Nonetheless, a notable number of users have raised concerns that the devices are 'always listening', and that there is a lack of clear information from providers about the data collected and processed through their microphones. Adopting a socio-informatics research perspective, we used the living lab approach to work with users over three years to investigate their uncertainties regarding the data collected by VAs in everyday usage. Based on our findings from interviews, fieldwork, and participatory design workshops with 35 households, we developed the web tool "CheckMyVA" to support users to access and visualize their own VA data. This chapter presents the observations and findings of the three-year study by outlining the implemented features of the tool and reflecting on how its design can help improve data literacy and enable users to reflect on their long-term interactions with VAs, ultimately serving to 'demystify' the technology.*

1. Introduction and Background

Since their launch in 2015, voice assistants (VAs) for home use such as Google Assistant or Amazon's Alexa have been steadily gaining prevalence (Bohn 2016), with the global market estimated to exceed 200 million devices in 2023 (Laricchia 2023). While users appreciate the usefulness and convenience of VAs, the ability to control these devices by voice also serves as a gateway to a growing ecosystem of data-based services (Strüver 2023a). Initial studies have shown that users are often unaware of what data these devices capture and whether or how their data is stored (Abdi, Ramokapane, and Such 2019; Alepis and Patsakis 2017; Jakobi et al. 2020; Pins et al. 2020). One reason for this is the lack of opportunities provided to users to learn about, understand, or manage the data collected by companies (Jakobi et al. 2020; Pins et al. 2020; 2021).

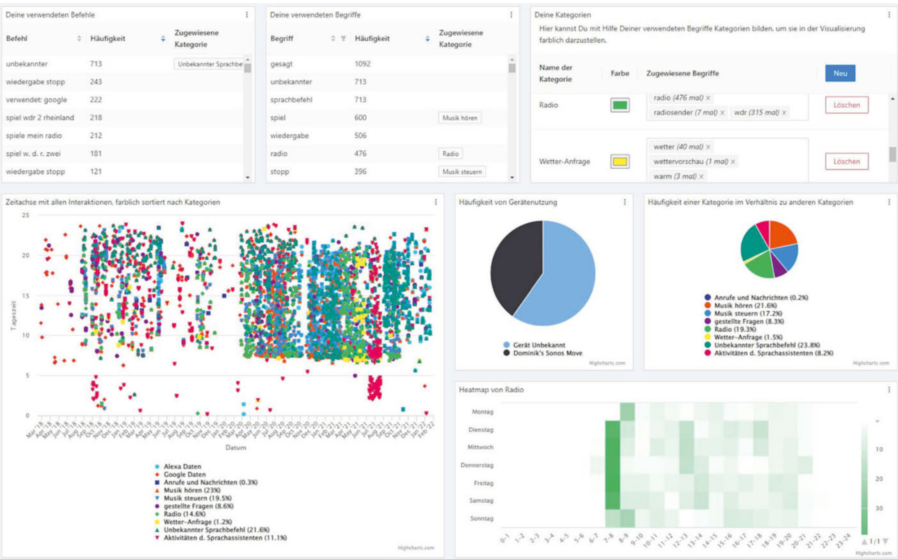
Figure 1: Extracts of raw data transcription files from data takeouts, received a) as a JSON file for Google Assistant and b) as a CVS file for Amazon Alexa

"header": "Google Assistant",
"title": "Lautstärke fünf gesagt",
"titleUrl": "https://www.google.com/search?q\
"time": "2020-06-09T17:21:42.533Z",
"products": ["Google Assistant"],
"locationInfos": [{
"name": "Ungefähre Gegend",
"url": "https://www.google.com/maps/@?api\u
"source": "Von meinem Gerät"
}]
},{
"header": "Google Assistant",
"title": "Spiele noch mal von vorne gesagt",
"titleUrl": "https://www.google.com/search?q\
"time": "2020-06-09T17:21:36.371Z",
"products": ["Google Assistant"],
"locationInfos": [{
"name": "Ungefähre Gegend",
"url": "https://www.google.com/maps/@?api\u
"source": "Von meinem Gerät"

68 2020-11-09T14:00:21.853Z,alexa wiedergabe starten,235ae1c97ab4101f61c072fbdae53cbe77c0f406.wav,No
69 2019-05-03T10:14:44.711Z,setze die zutaten auf die einkaufsliste,ce1a5aac205525b737001159c2a43bf2d758
70 2019-05-21T16:29:23.913Z,alexa weiter,735900cb0d2f416481e60633dfd04026b5c58d6f.wav,Not Applicable
71 2019-05-10T17:41:48.761Z,alexa öffne rewe,4539b3c488c13bf187085725aebca9a8f6f29621.wav,Sollen wir w
72 2019-05-10T17:49:25.599Z,alexa zubereitung,c2e61b1da4225219a5a38433a02d2689882ecea4.wav,"Hier der
73 2019-05-10T17:49:25.599Z,alexa zubereitung,c2e61b1da4225219a5a38433a02d2689882ecea4.wav,Soll ich da
74 2019-05-10T17:33:13.137Z,alexa zeige mir die zutaten,35c26561916f22e2e368efa2da4ce9a0f48b8b2f.wav,"I

When it comes to tracking what VAs have captured or processed, providers do offer options such as interaction logs, which can be accessed in users' account settings (see an analysis of the log data by Habscheid et al. (2021)), or, in the case of Amazon, users can ask Alexa directly why it performed in a certain way (Alizadeh, Pins, and Stevens 2023). However, studies have shown that while these options make it quite easy to access recent interactions, they do not offer an overall view of interactions over longer periods, nor are they suitable for conducting in-depth data work (Pins et al. 2020). For this reason, we leveraged the right to access data guaranteed by the General Data Protection Regulation (GDPR) in order to obtain raw interaction data from a longer period of time with which we could explore different visualization methods. Figure 1 shows how the interactions were presented in the data takeouts supplied by Google and Amazon respectively. The interaction data for Google Assistant (shown in JSON format in Figure 1) exhibit a uniform structure for each interaction. However, the individual labels at the beginning of each line are not self-descriptive: laypersons would not necessarily find them helpful to understand the subsequent information. Amazon provides the transcription of Alexa interactions as a CVS file, which includes the timestamp, the user command, the name of the audio file, and the response from Alexa for each interaction, listed line by line. As can be seen, both of these formats lack legibility, especially for laypersons, and interpreting them requires a deeper understanding of the data structure (Pins et al. 2021).

Figure 2: Dashboard for data visualization – (exemplary view)



Our aim was to examine, in a living lab study, how users of VAs integrate the devices into their daily lives and, in particular, how they deal with uncertainties regarding VAs' recordings of everyday life in their homes – whether intentional or accidentally activated, for example, by TV or human conversations. Our approach was guided by the understanding that the appropriation of technology is a social process, whereby artifacts are incorporated into one's everyday life (Draxler et al. 2012; Stevens, Pipek, and Wulf 2010; Wulf 2018); this incorporation influences behavior and can lead to new practices, thought patterns, and design approaches as reciprocal effects (Rohde et al. 2017).

This contribution reflects upon our development of CheckMyVA: a web tool intended to empower users of VAs from different providers by preparing and visualizing their interaction data. Figure 2 shows some of CheckMyVA's visualization options that allow users to view the recordings and corresponding transcriptions stored by VA providers, thereby demystifying what VAs are listening to and helping users to reflect on their usage behavior.

2. State of the Art

2.1. Privacy Concerns About the Use of VAs

VAs are valued highly for their convenience and for the captivating way they enable users to operate music, connected devices, and entire home systems by means of voice commands (Purington et al. 2017; Abdi, Ramokapane, and Such 2019; Brüggemeier et al. 2020). However, for many people, their usage is also associated with opacity, concern, and mistrust (Lau, Zimmerman, and Schaub 2018). Additionally, users have expressed disappointment that VAs do not always react and respond reliably, and more complex tasks are not always completed successfully (Bentley et al. 2018; Luger and Sellen 2016; Pins et al. 2020).

The reasons for these negative sentiments often lie in users' uncertainty about what exactly VAs 'understand' or record and how they process data (Luger and Sellen 2016; Malkin et al. 2019). Recent research has shown that most privacy concerns are associated with accidental activations (Schönherr et al. 2020; Malkin et al. 2019; Ford and Palmer 2019) along with anxiety about the presence of a device that is 'always listening' (Alepis and Patsakis 2017; Lau, Zimmerman, and Schaub 2018). However, disappointment and frustration were also expressed about providers' failure to provide appropriate support to deal with problems, such as by suggesting repair strategies to clarify why a VA acted in a certain way or to successfully resolve misleading interactions (Kiesel et al. 2019; Pins et al. 2020; Pins and Alizadeh 2021). Studies have shown that users are often unaware that they can view interaction-related data and review or delete them (Malkin et al. 2019; Pins et al. 2021; Sciuto et al. 2018).

As a result of these operational difficulties and privacy concerns, users tend to adapt their use behavior by trying to make their voice commands as trivial, uninteresting, or short as possible (Lau, Zimmerman, and Schaub 2018; Malkin et al. 2019; Pins et al. 2020). This behavior can also be explained by rational fatalism (Kerwin 2012) or resignation as an attempt to protect one's data from companies (Pins et al. 2021; Xie, Fowler-Dawson, and Tvauri 2019).

2.2. Usable Privacy for Greater Data Literacy

Advocates of 'usable privacy' argue for the need to design secure systems from the user's perspective (Cranor 2008) and to support consumers to manage their own data privacy (Adams and Sasse 1999). This includes aspects such as improving privacy awareness (Langheinrich 2002), making security tools usable (Whitten and Tygar 1999), and making privacy notices understandable (Angulo et al. 2012; Schaub et al. 2018).

Against the backdrop of increasingly comprehensive and complex data collection, current research in usable privacy focuses on adapting the data literacy concept (Zhang 2018). This concept, which originated in the educational sciences, has been defined in various ways (see Koltay (2015) for an overview). In summary, data literacy involves the ability to access, interpret, critically evaluate, manage, and process data, so that it can be transformed into actionable knowledge to make informed decisions (Calzada Prado and Marzal 2013; Koltay 2015; Mandinach and Gummer 2016).

In our contemporary data-driven economy and society, data literacy is not only a key skill for individuals, but also a prerequisite for informed data protection regulation. The GDPR right to access data has created an important technical basis for promoting data literacy by enabling individuals to access information stored about them. However, there is a lack of complementary measures to ensure that accessed data can be understood and effectively managed. To address this gap and promote data literacy among consumers, it is expedient to draw on methods such as information visualization (InfoViz) (Shneiderman 1996), data citizen science (Marr 2016) and data work (Tolmie et al. 2016), and combine them with playful data exploration techniques (Jakobi et al. 2017).

In the research field of digital consumer behavior, artificial intelligence and data science methods such as deep learning (Chapman and Feit 2019; Feldman and Sanger 2006; Igual Muñoz and Seguí Mesquida 2017) are increasingly used alongside classical statistical methods to identify relevant information in user data and to derive behavioral patterns. However, such methods are typically available only to companies and data scientists; there is a lack of usable solutions for consumers that enable automated analysis for different fields of application (Fischer et al. 2016).

InfoViz methods facilitate the visualization of time series, networks, and hierarchical data (Aigner et al. 2007; Ware 2013), which can reflect users' behaviors back to them (Castelli et al. 2017; Jakobi et al. 2017; Stevens et al. 2017).

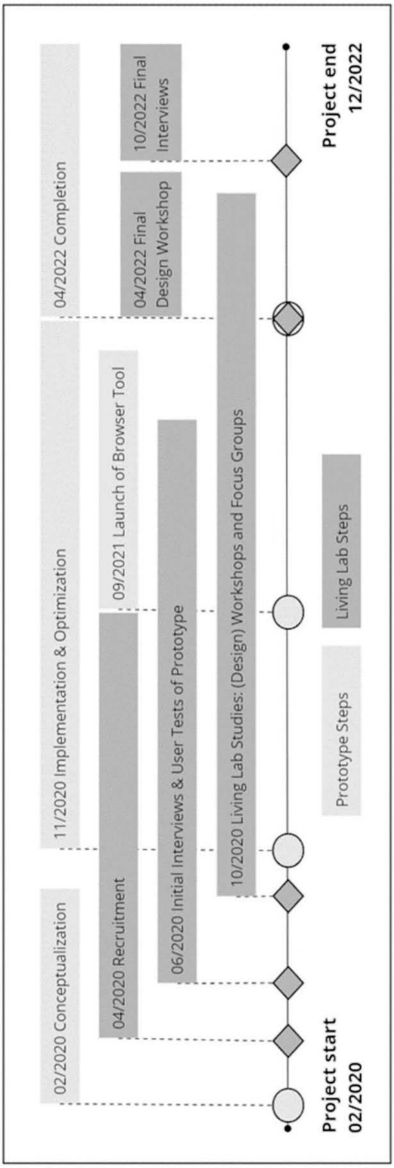
Initially, we were aware of just three further studies that had explicitly used log files of interactions to investigate the use of VAs (Malkin et al. 2019; Sciuto et al. 2018; Bentley et al. 2018). During our study, additional research examined interaction data to draw conclusions about human–VA interaction within the smart home ecosystem (Habscheid et al. 2021) and to assess privacy sensitivity and intimacy using data sharing scenarios (Gómez Ortega, Bourgeois, and Kortuem 2023).

While these studies primarily analyzed data for research purposes, our aim was to design a tool that could directly help consumers themselves to explore and understand their data, ultimately empowering them by improving their data literacy relating to VA use. The findings presented in this chapter build on a previous study that tested the process for requesting interaction data and evaluated an initial prototype for data visualization (Pins et al. 2021). Since then, we have completed the research project and are able to present the results of the iterative design process here.

3. Methodology

In this study, we adopted the 'living lab' approach to investigate ways to promote data sovereignty in the use of VAs. A living lab can be understood as a user-centered research methodology for sensing, prototyping, validating, and refining complex solutions in evolving real-life contexts (Eriksson and Kulkki 2005). Our procedure also incorporated the practice-orientated problem-solving strategy deployed in design case studies (Wulf et al. 2011). This approach takes into consideration the user, their (social) practices, institutional arrangements, and technological infrastructures, thereby exploring the design of innovative IT artifacts in situ (Wulf et al. 2015).

Figure 3: Project Timeline



Our aim was thus to study participants and their behavior in real-world settings, gaining insights into their use and understandings of voice interaction data. For the living lab study, we used mixed methods including interviews, fieldwork, and (design) workshops, in order to identify and validate users' needs and requirements. This iterative process enabled us to design, develop, and optimize a prototypical web tool (see Figure 3 for an overview of the research phases).

Parallel to the living lab study, we used several data donations from our participants to test the efficacy of various machine learning (ML) models to draw conclusions about users based on their data (digital consumer analytics), for example, to identify characteristics of users or their households. Unfortunately, the data set proved too small for the models to be trained precisely enough to be of practical use in the prototype.

Shortly after project launch in February 2020, we recruited households for the living lab via digital and social media.[1] By summer 2020, we were able to begin an initial needs assessment and evaluation of our first prototype with a sample of 12 households.

Over the course of the project, we worked with a total of 35 households. With each household we were in contact with a main participant who was the administrator of the VA and had access to the interaction data. These participants ranged in age from 18 to 56, with a mean age of 33. The sample included 24 males and 11 females, who lived in single and partner households, family households, and shared apartments. Sixteen households were 'beginners' who had never used a VA at home before joining the research project (for greater detail, see Table 1).

1 Due to the contact restrictions imposed by the simultaneous outbreak of the Covid-19 pandemic, no other recruitment strategies were practicable.

Table 1: Living lab household participants

#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H1	m	31	Software developer / consultant	Starter	2	Amazon Apple	2	high	high	very high	very high	very high
H2	m	35	Service technician (electrical)	User	3	Amazon	8	very high	very high	medium	very high	high
H3	Participation withdrawn											
H4	m	31	Public sector employee	User	5	Google	2	high	high	high	medium	high
H5	w	32	Market and social research specialist	Starter	3	Amazon Apple	2	high	high	high	medium	high
H6	m	24	Trainee in the higher forest service	User	2	Amazon	2	medium	medium	high	high	very high
H7	m	28	Electrical engineer	User	3	Google	5	high	high	very high	high	very high
H8	w	26	Research associate (HCI)	User	1	Amazon	1	little	medium	medium	little	medium
H9	m	56	Sales officer	User	2	Google	5	very high	very high	very high	very high	very high
H10	m	35	Industrial salesman, real estate industry	User	4	Amazon	4	high	medium	high	medium	high

#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H11	m	27	PhD student (Physics)	User	1	Google	3	high	very high	medium	high	medium
H12	m	18	Student	Starter	4	Google	2	N/A	N/A	N/A	N/A	N/A
H13	m	45	IT	Starter	4	Amazon	1	very high	very high	very high	high	very high
H14	w	32	N/A	Starter	1	Amazon	1	N/A	N/A	N/A	N/A	N/A
H15	w	27	Architect (residential)	Starter	2	Amazon	1	medium	very high	high	very high	medium
H16	w	40	Project manager	Starter	1	Amazon	1	medium	medium	medium	medium	medium
H17			Participation withdrawn									
H18	m	34	N/A	User	3	Amazon	4	medium	high	little	medium	little
H19	m	23	IT consultant	User	2	Amazon	3	medium	high	little	very high	little
H20	w	40	Research associate	Starter	2	Google	1	medium	high	high	medium	N/A

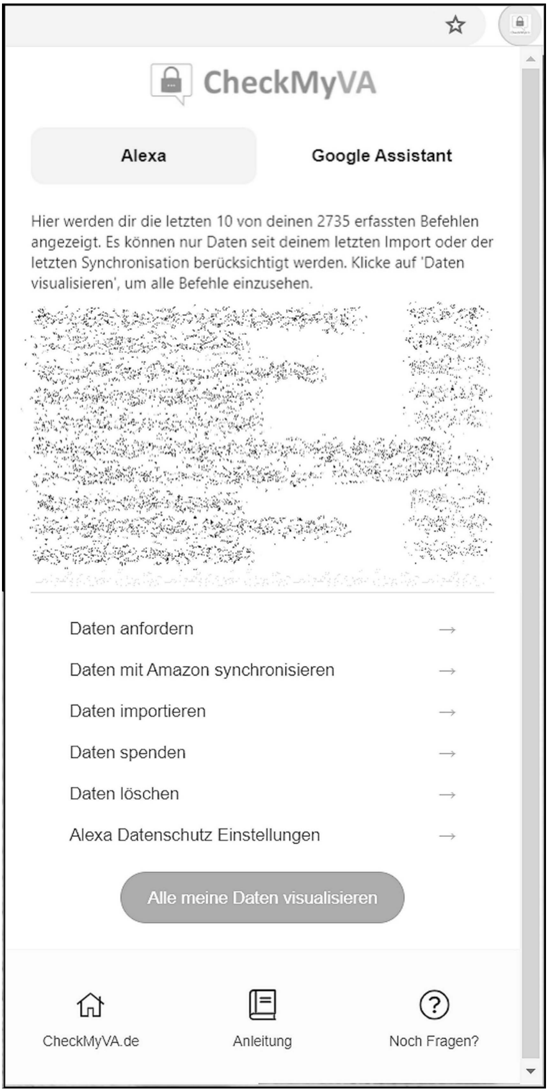
https://doi.org/10.14361/9783839472002-013 https://www.inlibra.com/de/agb - Open Access -

#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H21	m	29	Mechanical engineer (packaging industry)	Starter	3	Amazon	1	medium	high	very high	very high	medium
H22	m	27	N/A	User	2	Amazon	2	N/A	N/A	N/A	N/A	N/A
H23			Participation withdrawn									
H24	w	30	N/A	Starter	1	Google	2	N/A	N/A	N/A	N/A	N/A
H25	m	32	Scientific staff	User	2	Amazon	2	high	high	little	very high	high
H26	m	48	Scientific staff	User	2	Google	2	high	high	medium	high	medium
H27	m	21	ServiceNow-developer (ITSM)	Starter	1	Amazon	2	high	very high	medium	very high	high
H28	m	34	Research associate (Computer science)	User	1	Amazon Google	2	high	high	medium	high	medium
H29	w	N/A	Research associate	Starter	1	Amazon	2	medium	medium	high	medium	high
H30	w	27	Research associate	Starter	1	Google	1	medium	medium	high	little	medium

#	Gender	Age	Job (Field)	Level	Residents	VA-system	Devices	Interest in the topic				
								VAs	AI	Privacy	Smart Home	IT-security
H31	m	21	Student (Speech therapy)	Starter	2	Google Amazon	2	medium	medium	high	medium	medium
H32	m	34	Student	User	2	Google	2	high	high	high	high	medium
H33	m	N/A	Software developer	Starter	1	Google	1	little	high	high	No interest	high
H34	m	42	Sales for control	User	2	Google Amazon Siri	4	high	very high	high	high	high
H35	w	48	Freelancer (digital	Starter	1	Google Amazon	2	very high	very high	high	very high	very high

4. Findings and Implementation

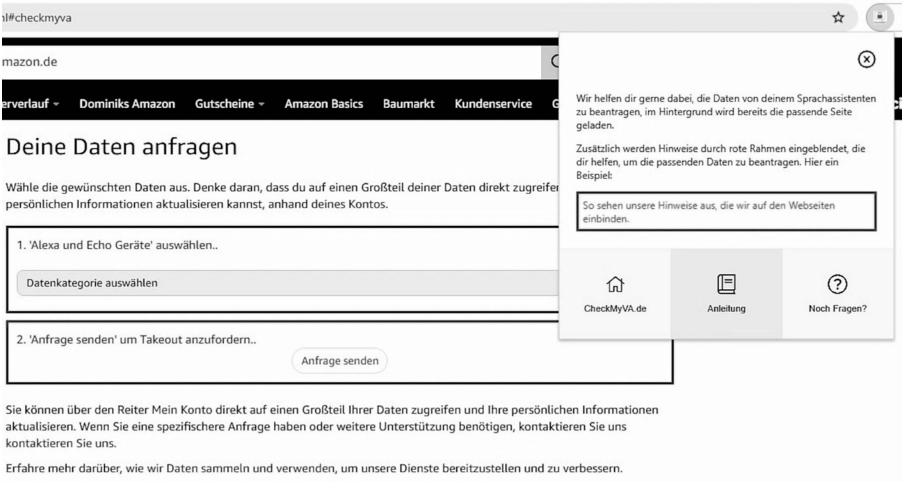
Figure 4: Main Menu of the CheckMyVA tool



The living lab study led us to design and produce CheckMyVA, a web tool that offers consumers two services: a data export wizard and a dashboard for data visualization, which can be accessed from a main menu (shown in Figure 4). The data export wizard directs users to VA providers' export websites and guides them with helpful dialogues through the often laborious and obscure export process. The dashboard enables consumers to display various data visualizations of the interaction data once they have accessed it. The tool is freely available as a browser extension for Google Chrome and Mozilla Firefox and can process data from Alexa and Google Assistant.[2]

4.1. Data Export Wizard

Figure 5: Guidelines for requesting a data takeout from Amazon



In our previous study, users had reported experiencing difficulties in finding user data, and that the process of retrieving it was very cumbersome and user-unfriendly (Pins et al. 2021). Hence, we created the data export wizard with the aim of supporting users through the process of exporting data from

2 For Google Chrome: https://chrome.google.com/webstore/detail/checkmyva-browser-erweite/kpllpbalbkdcdoklbnjlbbbeapfhoodp (18.05.2024) For Mozilla Firefox: https://addons.mozilla.org/de/firefox/addon/checkmyva/ (18.05.2024)

Google and Amazon. With a single click, users are directed to the appropriate export web pages and are guided by help dialogues in boxes highlighted in red (see Figure 5), making it easy for them to request data exports from their VAs.

Once the user has received the data takeout, the wizard processes and reads the data locally in the background. The stored data is then made available through an interface between the web tool and the dashboard. This ensures that data remain secure in the browser without needing to be uploaded to other services. Users do not have to unpack data archives or search for and open the relevant files themselves.

Obtaining a data takeout from Alexa can take from several days to several weeks, and to obtain the latest interaction data, a new takeout request must be made each time. To address this issue, we explored alternative methods to make the latest interaction data available to participants more quickly. We successfully implemented a system that enables interactions to be synchronized with our dashboard in real time. This real-time approach was well-received by participants. To make the process even simpler, we implemented another function that synchronizes data each time the browser is started.[3] In addition to data request and synchronization, we added the following features to the wizard (see also Figure 4):

- **Import data**: Users can import locally-stored interaction data.
- **Delete data**: Users can delete the data stored in the browser and the prototype.
- **Privacy settings**: After viewing the data, some users wanted to check their settings. For this reason, we added a link that directs users to the privacy settings in their Google or Amazon accounts so that they can make quick adjustments.
- **Data donation**: This feature allows users to transmit their data stored in the prototype (transcribed voice commands, responses, timestamps, devices, etc.)[4] to an internal server for further research within our research project, including user evaluation and training of ML models. Users must explicitly opt in to this procedure.

---

3 For Google Assistant data, this took a few minutes to a few hours, which participants considered acceptable. However, we have not found a way to synchronize the data in a similar way to Amazon.

4 Due to the large size of audio files, and because the dashboard could not process audio data anyway, we limited the data donation to textual data only.

4.2. Data Visualization Dashboard

The initial prototype for data visualization featured a timeline that helped participants gain an overview of their interactions over a longer period of time (Pins et al. 2021). This visualization was evaluated by participants as very useful and informative. Furthermore, the categorization of interactions according to specific terms enables the data points to be structured along the timeline in relevant ways, helping users to identify frequent or typical usage times and situations. Step-by-step categorization also facilitates the identification of further interaction patterns and of unusual or unexpected interactions or recordings. For instance, participants often expressed their surprise at discovering unexpected activities at night, or mentioned that they became aware that in viewing the visualization they were surveilling the interactions of other household members at times during which they themselves had been absent, e.g., when they had been at work (Pins et al. 2021).

Based on the results of the ongoing iterative process, we continued to optimize and extend the initial prototype. Like the data export wizard, the dashboard is implemented as a web application. It can access the user's data via the data export wizard automatically and offline, performing like a native desktop application. A screenshot of the final version of the visualization dashboard is shown in Figure 2.

In the process of preparing data for visualization, transcription errors (such as miscoded umlauts or punctuation marks) and VA command words ("Google", "Alexa" or "said") are removed to facilitate the visualization. Once the preparation is complete, users can create categories based upon individual command words using the Boolean operations (AND or OR). Each category can be assigned a color and a theme (see Figure 6). Additionally, we created a catalog of predefined categories that users can select from and customize.

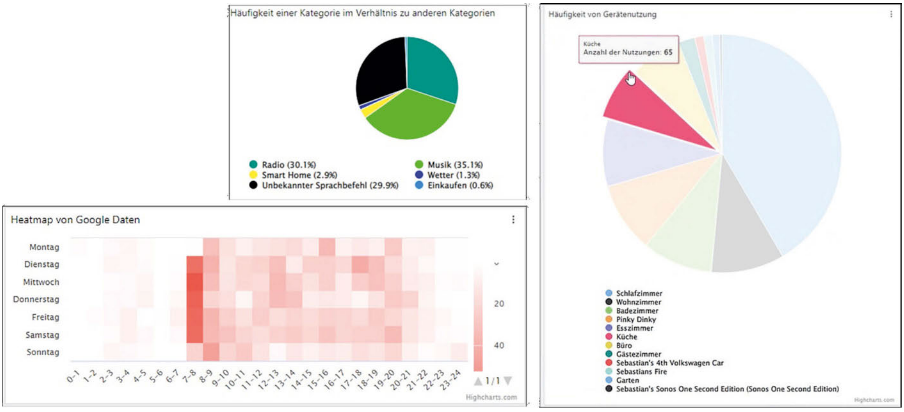
The categorized data thus forms the starting point for different visualizations and analyses. A scatter chart (timeline) displays each command as a dot in the color of the defined category (see Figure 2; bottom left of screen). This visualization shows the frequency of commands per category and thus helps users identify behavioral patterns associated with frequently-used categories. Users can zoom in on specific areas by dragging a frame over them with the mouse. Finally, by clicking on a category in the legend, all the corresponding dots on the timeline can be shown or hidden.

Figure 6: Widget for creating new categories based on terms.

Over the course of the project, we added various widgets based on users' needs and interests. Widgets are tiles with data visualizations that can be freely arranged on the dashboard, allowing users to customize their views and prioritize the information most important to them. Each widget also offers the option to display data from an individually-defined time period. The following widgets were implemented:

- **Word list**: Following evaluation of the initial prototype, we modified the sorting options for the list of words spoken so that they could be presented either in order of frequency or alphabetically. The list also shows which categories a term has been assigned to.
- **Command list**: To meet participants' requests for a list of spoken commands, we added another list with the same presentation options as the word list. It helps users to identify speech patterns.
- **Usage occasions and their frequency**: A pie chart and bar chart showing the relative distribution of categories (see Figure 7, top left).
- **Device usage**: A pie chart showing the relative distribution of devices used (see Figure 7, right). This enables users to check the frequency of device use and draw conclusions about the associated rooms in the home.
- **Occasions of use per day and time**: A heat map that shows the number of commands in a given category aggregated into hours per week. Each field is displayed in varying intensity of the category color depending on the frequency of use (see Figure 7, bottom left). This helps users identify typical usage times per category.

Figure 7: Widgets for data visualization: Relative frequency of each category within the total data set (top left), heatmap with clusters of interactions of a selected category (bottom left), and relative frequency of device usage by (assigned) device name (right).



We also conceptualized some additional widget designs in participatory design workshops in which participants expressed their needs and interests. The limited project timeframe prevented these widgets from being implemented into the tool, but participants' request for them nonetheless constitutes a significant research outcome. The following three design concepts for widgets would help users to gain a better understanding of VAs' data processing procedures:

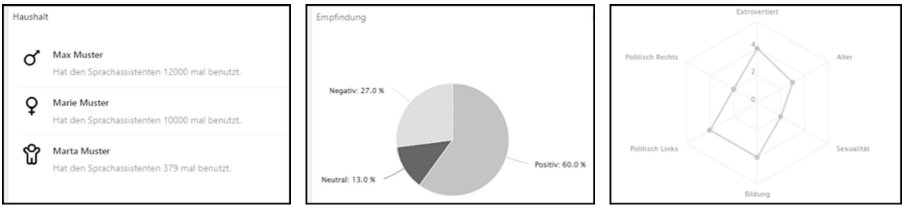
- **Speech analysis**: A widget for categorizing and detailing commands in order to correct interaction/pronunciation differences and recognize changes in interaction behavior.
- **Data flows**: A widget to show how (and with whom) data is shared, identifying critical or personal data and providing user action options.
- **Memories**: A widget for saving interaction data as material that can evoke memories of appointments, special occasions, or situations; supported by images or sound if these are available or linked to the data.

Finally, we conducted internal tests with ML models to explore how the data could be used in digital consumer analytics. The main goal was to identify profiles of users or their households. The users as well as our research team were interested whether the data could convey information about household size, age, gender, or a speaker's mood when interacting with the VA. We asked participants for data donations to test various ML models.

Due to the small size of the donated data set we were unable to train the models precisely enough to achieve conclusive results. Nevertheless, to give users a sense of what information could potentially be extrapolated from the data, we generated mock-ups based on the available data. These mock-ups present insights in the following widgets:

- **Household widget**: A list of all VA users, distinguishing individual voices and creating profiles that record their age, gender, and frequency of use of the VA (see Figure 8, left).
- **Sentiment widget**: A pie chart showing how often a particular command is executed with a positive, neutral, or negative intonation (see Figure 8, center).
- **Politeness widget**: Emojis indicating how politely users speak to the VA.
- **Health widget**: A scatter chart showing how often a user is sick, based upon audible symptoms like coughing, sneezing, hoarseness, or fatigue.
- **Background noise widget**: A scatter chart showing the frequency of certain background sounds and any incorrect activations they may have caused. For example, it indicates how often media (TV, radio, music), other conversations, or other sounds are present in the background.
- **Advertising widget**: A word cloud visualization of the brand names mentioned in voice commands.
- **Profiling widget**: A spider chart ranking inferred personality traits (see Figure 8, right).

Figures 8-10: Widgets for the household/user profile: Amount of use per household member (left), inferred personality traits of a user (center), and inferred positive, neutral, or negative mood when articulating a voice command (right).



5. Discussion

5.1. Data Work Promotes Data Awareness and Literacy

In our living lab study, participants expressed great interest to try out the CheckMyVA tool for the first time, and reported that using it made them feel reassured (Pins et al. 2021). Over the course of the study, however, only a small number of participants continued to use the tool regularly on their own initiative. In final interviews, the following reasons for using the tool were mentioned:

- Curiosity about what new interactions had been detected or stored by the VA.
- Coming across the tool icon by chance while using the browser.
- Checking for funny answers given by the VA.
- Checking for interactions including insults by others (and deleting them).

While the first two reasons indicate curiosity or the 'accidental rediscovery' of the tool, the last two are motivated by the desire to review unusual situations and interactions. This might explain why the majority of participants did not use the tool again; they may not have expected any new insights or unusual interactions, and therefore felt no subsequent need to explore the data. When asked in which situations they thought the tool might be helpful, several participants mentioned reviewing unexpected or incorrect responses. This suggests that after an initial 'awareness' check, users' interest in the data shifts over time, with the most attention concerning deviant activities. Such a shift has also been identified in other studies with different data work contexts (Castelli, Stevens, and Jakobi 2019; Jakobi et al. 2018).

5.2. Towards Better Support in Requesting Data (According to Article 15 of the GDPR)

The study has shown that the procedures of requesting data collected by VAs are neither simple nor easily comprehensible from the user's perspective (Pins et al. 2021). Tools like our prototype that can guide users through the data request process thus make a valuable contribution to increasing data literacy and users' knowledge. Easily locatable and accessible guidance on how to view or request data from each provider can help users overcome barriers to addressing the issue of data collection, thereby increasing their competence to use products, services, and systems, as well as to manage their collected data.

It also became apparent that different corporations deal with the volume and format of users' data in very different ways (Cena et al. 2016; Shafagh and Hithnawi 2017; Pins et al. 2022). Even between the two VA systems considered in this project, approaches vary significantly. Initially, we had planned to include interaction data from Siri (Apple), but that proved to not be possible due to their pseudonymization process, which prevents access to usable interaction data. Additionally, the ongoing development of these systems appears very inconstant. For instance, in response to public criticism, Google suspended the automatic storage of audio recordings for a while. Since then, it changed its policy so that Google Assistant users can currently opt in to anonymized data storage to improve speech and audio recognition, which may involve human review.[5]

While such pseudonymization (or anonymization) practices are to be welcomed from a data protection perspective, their effectiveness remains questionable if conclusions can still be drawn from the content of audio or transcript data, even when it has been separated from user profile data. Amazon also allows Alexa users the option to disallow the storage of interaction data, but this requires deliberate deactivation by the user – if the default settings are not adjusted, users implicitly consent to data storage.

Policy makers should work to generate more guidelines for the storage of user data and should make corporations accountable for providing easily-accessible, relevant information about the collection and processing of users' data, especially regarding the companies' evaluation and analysis processes for consumer analytics purposes. For example, information should be provided on how sensitive information is handled when recordings are subject to human review.

5.3. Towards Demystification: Visualization and Sense-making of Data

In accordance with Article 15 of the GDPR, companies are obliged to provide consumers with information about their personal data collected by the company, and to transmit it upon request in a machine-readable format for the purpose of data portability in accordance with Article 20 (European Parliament

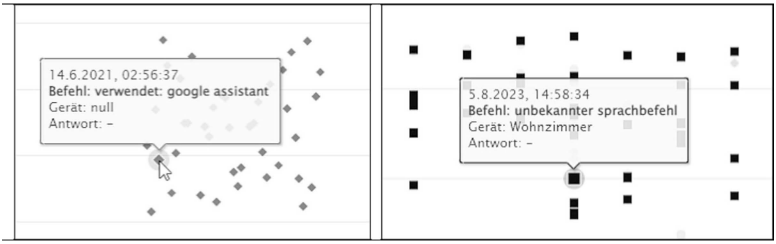
5 https://www.cnet.com/home/smart-home/googles-privacy-controls-on-recordings-changes-what-that-means-for-your-google-home/

and the Council 2018). The personal data of our participants that we requested for this project was raw data, which consumers without technical knowledge would barely understand (see Figure 1). Previous studies have shown that users expect a more human-readable format (Alizadeh et al. 2019; Pins et al. 2021). This highlights the need for solutions like our visualization dashboard, which enables users to make sense of their data and to better understand what it can show about their usage behavior (Castelli et al. 2017; Jakobi et al. 2017; Stevens and Bossauer 2017). While some information from the raw data takeout was clear and actionable, a major challenge for us was to identify significant insights that could potentially be derived from the data in order to draw conclusions about users and their behavior. These relevancies are not clearly evident within the raw data sets, which makes it difficult for users to understand the profiles created about them, and there are no end-user options that would enable them to create their own analyses.

In order to meet our aim to empower VA users to understand and interpret the data that companies collect from them, we needed a sample data set with which we could demonstrate in an exemplary way to users the potential analytical capabilities of companies. Our ability to do this was limited by the small size of the data set that we were able to obtain voluntarily from the few households that were willing to donate their data. We believe that companies should make it more transparent how a user's profile is compiled and what criteria are used to generate such profiles, so that users can understand and adjust settings accordingly if they so wish. This transparency could balance the information and power disparity between the user and provider, without requiring corporations to disclose their algorithms, but nonetheless helping to clarify or 'demystify' the opacity of technologies like VAs. Indeed, the few households in our study that agreed to donate us their data only did so once they understood what it included, suggesting that transparency might influence users' decisions about sharing data, especially when they feel uncertain about how the data could be analyzed and interpreted.

During the study and data analysis, certain inconsistencies in the data takeouts became apparent. For example, Amazon provides information about the device used for each interaction in the accounts' interaction log, which can be found in the account settings. However, we could not find corresponding information in the data takeouts. This suggests that some data correlations are not included in the takeouts, even though some connections between transcripts and recordings are traceable. Similarly, with Apple, it cannot be ruled out that personal data may still be found in the data records that are stored pseudonymously. Previous studies (Malkin et al. 2019; Pins et al. 2021), found that consumers were surprised to learn that voice commands were stored long-term. Figure 9 illustrates two activities of the VA shown on the dashboard that were included in the data set that users were surprised to discover had been stored, especially as such activities had an unclear purpose or occurred at unusual times.

Figure 9: Unusual activities of the VA without intelligible clarification.



Providers should therefore be held to account to make their data storage 'transparent' in the sense that users should be able to understand which elements of the data are interlinked for companies' analytical purposes (without firms having to disclose their algorithms or methods used). Companies should also be required to delete data that no longer serves a purpose.

5.4. Raising Awareness of the Technological Infrastructure in Which the VA is Embedded

For most participants, our study provided their first ever chance to view and engage with the data collected by their VAs. On the one hand, they said they felt reassured, because they had gained more clarity about what data the VA was collecting and how they could exert control over its transmission. In particular, it became clear that the majority of the data and usage situations (e.g., setting timers or playing music) that the participants learned about were not considered risky, concerning, or sensitive. This enabled them to act more self-determinedly when talking or acting near to a smart speaker at home. But on the other hand, viewing the data raised new questions, as they had expected to be able to obtain more information directly from the (raw) data received about the extent to which data was exchanged between various services. Instead, they initially found themselves confronted with a folder directory comprising incomprehensible data records that first had to be 'decoded' (Pins et al. 2021).

Research on VA systems should never consider them in isolation, but always in the context of the environment and linked services within which they are embedded (Strüver 2023a). Consumers express particular uncertainty regarding the extent of corporations' access to and exchange of data (Huang, Obada-Obieh, and Beznosov 2020; Luger and Sellen 2016; Malkin et al. 2019). Recent research and our study indicate the importance of viewing the home holistically, as a network of different players, in order to understand various links and activities in context (Strüver 2023b; Häußling 2017). For example, further research could distinguish between smart home products and services used (or their manufacturers) to provide more differentiated information about their general usage or integration in everyday practices.

A holistic view of the infrastructures or platform systems (Plantin et al. 2018) would also help consumers to create transparent and trustworthy environments for themselves, which is particularly important for private and intimate areas like the home. Recent studies have furthered understanding of the basic intentions behind data collection/processing (Sadowski 2020; Strüver 2023a; 2023b). Our approach also focuses on showing users what the storage of interaction data can mean for them, their household, and their usage behavior. Further research should link these aspects more closely to help users better understand how their data is affected by corporations' intentions. To conduct such research effectively would require a larger data set than was available to us for this study.

Current data work practices offered by companies usually only address the account owner/administrator (Meng, Kekülluoğlu, and Vaniea 2021). Therefore, a more holistic view of the home (technology) ecosystem is needed to achieve a multi-user-centric design, creating more productive, convenient, and inclusive IoT environments for other household members, visitors, etc. (Strüver 2023b). This approach would allow more people to gain insights into the interaction data and learn what the VA has captured about them and their households (Meng, Kekülluoğlu, and Vaniea 2021; Strüver 2023b; Waldecker, Hector, and Hoffmann 2023).

5.5. Limitations and Reflections

The scope of this study was limited by the sample. First, we engaged primarily with the administrators of the devices who had direct access to the interaction data via their user accounts; hence we focused mainly on their needs. As our study showed, and as other studies have demonstrated in greater depth (Thakkar et al. 2022; Sun et al. 2021), other residents in a household are also affected by a VA's data collection – but they were not included in the study. These individuals should also be able to view the interaction data to see what the VA provider or account holder can see about them. Our tool provides an initial indication of how this could be achieved by making the dashboard accessible to other household members, for example, via a shared device (e.g., a tablet or PC) with the tool's browser application installed.

Second, by only working with administrators, our sample could have been affected by a demographic imbalance in terms of age and gender. Previous studies (Geeng and Roesner 2019; Pins et al. 2020; 2021; Shin, Park, and Lee 2018) suggest that administrators tend to be male and tech-savvy, which may influence their interest in using VAs. However, this study did not aim to be representative but rather to support consumers who use a VA. Nonetheless, other user groups might express different needs and interests relating to data access from that we were not able to take into account in our study.

Another limitation arises from the dynamic nature of data usage and the services available to consumers at any specific time, which is constantly changing in response to ongoing developments, public criticism, and policy changes. Hence, replicating this study at a later time might yield different results.

6. Conclusion

In this chapter, we presented the features of a web tool created with the aim to empower VA users by increasing their data literacy, and reflected on the tool's development. This involved conducting a three-year living lab study to investigate VA use and data work practices, identifying what users need in order to better understand how VAs collect and process interaction data. The tool includes a data export wizard that guides users through the process of requesting interaction data as well as assisting them in viewing and managing privacy settings. It also offers a dashboard that allows the data to be structured and visualized in different ways (e.g., according to user-defined categories) to help users better understand and reflect on their usage.

Previous studies have shown that users often express uncertainty and skepticism about what their VA is listening to and storing; similar sentiments were voiced by our participants. Our tool addresses this by demystifying VA systems for users, enabling them to explore their own behavioral patterns through visualizations and to recognize unconscious or accidental activations. Ultimately, the tool helped participants to assess what data the VA collects and what it can reveal about a person or household.

Acknowledgements

The project was supported by funds of the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program.

References

Abdi, N, K M Ramokapane, and J M Such. 2019. "More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants." In *Proceedings of the Fifteenth Symposium on Usable Privacy and Security*, 15:17. Santa Clara: USENIX Association. https://www.usenix.org/system/files/soups2019-abdi.pdf.

Adams, Anne, and Martina Angela Sasse. 1999. "Users Are Not the Enemy." *Communications of the ACM* 42 (12): 40–46. https://doi.org/10.1145/322796.322806.

Aigner, Wolfgang, Silvia Miksch, Wolfgang Müller, H. Schumann, and Christian Tominski. 2007. "Visualizing Time-Oriented Data – A Systematic View." *Computers & Graphics* no. 31 (June): 401–09. https://doi.org/10.1016/j.cag.2007.01.030.

Alepis, Efthimios, and Constantinos Patsakis. 2017. "Monkey Says, Monkey Does: Security and Privacy on Voice Assistants." *IEEE Access* 5: 17841–51. https://doi.org/10.1109/ACCESS.2017.2747626.

Alizadeh, Fatemeh, Timo Jakobi, Jens Boldt, and Gunnar Stevens. 2019. "GDPR-Reality Check on the Right to Access Data: Claiming and Investigating Personally Identifiable Data from Companies." In *Proceedings of Mensch Und Computer 2019 on – MuC'19*, 811–14. Hamburg: ACM Press. https://doi.org/10.1145/3340764.3344913.

Alizadeh, Fatemeh, Dominik Pins, and Gunnar Stevens. 2023. "User-Friendly Explanatory Dialogues." In *Mensch Und Computer 2023– Workshopband,* 3. Rapperswil (SG): Gesellschaft für Informatik e.Vin P. Fröhlich & V. Cobus. https://doi.org/10.18420/muc2023-ws16-120.

Angulo, Julio, Simone Fischer-Hübner, Erik Wästlund, and Tobias Pulls. 2012. "Towards Usable Privacy Policy Display and Management." Edited by Nathan Clarke and Steven Furnell. *Information Management & Computer Security* 20 (1): 4–17. https://doi.org/10.1108/09685221211219155.

Bentley, Frank, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottrjdge. 2018. "Understanding the Long-Term Use of Smart Speaker Assistants." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (3): 1–24. https://doi.org/10.1145/3264901.

Bohn, Dieter. 2016. "Google Home: A Speaker to Finally Take on the Amazon Echo." *The Verge* (blog). May 18, 2016. https://www.theverge.com/2016/5/18/11688376/google-home-speaker-announced-virtual-assistant-io-2016.

Brüggemeier, Birgit, M. Breiter, Miriam Kurz, and Johanna Schiwy. 2020. "User Experience of Alexa, Siri and Google Assistant When Controlling Music – Comparison of Four Questionnaires." In *HCI International 2020 – Late Breaking Papers: User Experience Design and Case*, 600–18. Online. https://doi.org/10.1007/978-3-030-60114-0_40.

Calzada Prado, Javier, and Miguel Ángel Marzal. 2013. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents." *Libri* 63 (2): 123–34. https://doi.org/10.1515/libri-2013-0010.

Castelli, Nico, Corinna Ogonowski, Timo Jakobi, Martin Stein, Gunnar Stevens, and Volker Wulf. 2017. "What Happened in My Home?: An End-User Development Approach for Smart Home Data Visualization." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 853–66. Denver: ACM. https://doi.org/10.1145/3025453.3025485.

Castelli, Nico, Gunnar Stevens, and Timo Jakobi. 2019. "Information Visualization at Home: A Literature Survey of Consumption Feedback Design." *Literature Survey* 16 (16). Bonn: International Institute for Socio-Informatics. https://pub.h-brs.de/frontdoor/index/index/docId/4505.

Cena, Federica, Silvia Likavec, Amon Rapp, and Alessandro Marcengo. 2016. "An Ontology for Quantified Self: Capturing the Concepts behind the Numbers." In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 602–04. Heidelberg: ACM. https://doi.org/10.1145/2968219.2968329.

Chapman, Chris, and Elea Nc Donald Feit. 2019. *R For Marketing Research and Analytics*. Use R! online (e-Book): Springer Cham. https://link.springer.com/book/10.1007/978-3-030-14316-9.

Cranor, Lorrie F. 2008. "A Framework for Reasoning About the Human in the Loop." In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, 1–15. USA: USENIX Association.

Draxler, Sebastian, Gunnar Stevens, Martin Stein, Alexander Boden, and David Randall. 2012. "Supporting the Social Context of Technology Appropriation: On a Synthesis of Sharing Tools and Tool Knowledge." In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems – CHI '12*, 2835. Austin: ACM Press. https://doi.org/10.1145/2207676.2208687.

Eriksson, Mats, and Seija Kulkki. 2005. "State-of-the-Art in Utilizing Living Labs Approach to User-Centric ICT Innovation – A European Approach." *State-of-the-Art in Utilizing Living Labs Approach to User-Centric ICT Innovation* 15 (January).

European Parliament and the Council. 2018. *General Data Protection Regulation (GDPR)*. https://gdpr-info.eu/.

Feldman, Ronen, and James Sanger. 2006. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511546914.

Fischer, Joel E., Andy Crabtree, Tom Rodden, James A. Colley, Enrico Costanza, Michael O. Jewell, and Sarvapali D. Ramchurn. 2016. "'Just Whack It on until It Gets Hot': Working with IoT Data in the Home." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5933–44. San Jose: ACM. https://doi.org/10.1145/2858036.2858518.

Ford, Marcia, and William Palmer. 2019. "Alexa, Are You Listening to Me? An Analysis of Alexa Voice Service Network Traffic." *Personal and Ubiquitous Computing* 23 (1): 67–79. https://doi.org/10.1007/s00779-018-1174-x.

Geeng, Christine, and Franziska Roesner. 2019. "Who's In Control?: Interactions In Multi-User Smart Homes." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems – CHI '19*, 1–13. Glasgow: ACM Press. https://doi.org/10.1145/3290605.3300498.

Gómez Ortega, Alejandra, Jacky Bourgeois, and Gerd Kortuem. 2023. "What Is Sensitive About (Sensitive) Data? Characterizing Sensitivity and Intimacy with Google Assistant Users." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16. Hamburg: ACM. https://doi.org/10.1145/3544548.3581164.

Habscheid, Stephan, Tim Moritz Hector, Christine Hrncal, and David Waldecker. 2021. "Intelligente Persönliche Assistenten (IPA) mit Voice User Interfaces (VUI) als 'Beteiligte' in häuslicher Alltagsinteraktion. Welchen Aufschluss geben die Protokolldaten der Assistenzsysteme?" *Journal für Medienlinguistik* 4 (1): 16–53. https://doi.org/10.21248/jfml.2021.44.

Häußling, Roger. 2017. "Relationale Technik- und Organisationssoziologie." *Soziologie: Forum der Deutschen Gesellschaft für Soziologie*, Symposium: Was ist Netzwerkforschung, 61 (1): 47–50.

Huang, Yue, Borke Obada-Obieh, and Konstantin (Kosta) Beznosov. 2020. "Amazon vs. My Brother: How Users of Shared Smart Speakers Perceive and Cope with Privacy Risks." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. Honolulu: ACM. https://doi.org/10.1145/3313831.3376529.

Igual Muñoz, Laura, and Santi Seguí Mesquida. 2017. *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications*. Undergraduate Topics in Computer Science. Cham: Springer.

Jakobi, Timo, Edna Kropp, Gunnar Stevens, and Mats Schmal. 2017. "Providing Smartphone Data Visualizations to Support Privacy Literacy." In *EuroUSEC 2017 Workshop*, 2. Paris: Online. https://www.ndss-symposium.org/eurousec-2017/.

Jakobi, Timo, Gunnar Stevens, Nico Castelli, Corinna Ogonowski, Florian Schaub, Nils Vindice, Dave Randall, Peter Tolmie, and Volker Wulf. 2018. "Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (4): 1–28. https://doi.org/10.1145/3287049.

Jakobi, Timo, Gunnar Stevens, Maximilian von Grafenstein, Dominik Pins, and Alexander Boden. 2020. "User-Friendly Formulation of Data Processing Purposes of Voice Assistants: A User Perspective on the Principle of Purpose Limitation." In *Proceedings of the Conference on Mensch Und Computer*, 361–72. Magdeburg: ACM. https://doi.org/10.1145/3404983.3405588.

Kerwin, Jason T. 2012. "'Rational Fatalism': Non-Monotonic Choices in Response to Risk." *Working Paper*, no. WGAPE Spring 2012, 70.

Kiesel, Johannes, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. 2019. "Clarifying False Memories in Voice-Based Search." In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 331–35. Glasgow: ACM. https://doi.org/10.1145/3295750.3298961.

Koltay, Tibor. 2015. "Data Literacy: In Search of a Name and Identity." *Journal of Documentation* 71 (2): 401–15. https://doi.org/10.1108/JD-02-2014-0026.

Langheinrich, Marc. 2002. "A Privacy Awareness System for Ubiquitous Computing Environments." In *UbiComp 2002: Ubiquitous Computing*, edited by Gaetano Borriello and Lars Erik Holmquist, 237–45. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-45809-3_19.

Laricchia, Federica. 2023. "Smart Speakers – Statistics & Facts." Dossier. Consumer Electronics. Statista. https://www.statista.com/topics/4748/smart-speakers/.

Lau, Josephine, Benjamin Zimmerman, and Florian Schaub. 2018. "Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers." *Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW): 1–31. https://doi.org/10.1145/3274371.

Luger, Ewa, and Abigail Sellen. 2016. "'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems – CHI '16*, 5286–5297. Santa Clara: ACM Press. https://doi.org/10.1145/2858036.2858288.

Malkin, Nathan, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. "Privacy Attitudes of Smart Speaker Users." *Proceedings on Privacy Enhancing Technologies* 2019 (4): 250–71. https://doi.org/10.2478/popets-2019-0068.

Mandinach, Ellen Beth, and Edith S. Gummer. 2016. *Data Literacy for Educators: Making It Count in Teacher Preparation and Practice*. Technology, Education-Connections. New York: Teachers College Press.

Marr, Bernard. 2016. "How The Citizen Data Scientist Will Democratize Big Data." Forbes. April 2016. https://www.forbes.com/sites/bernardmarr/2016/04/01/how-the-citizen-data-scientist-will-democratize-big-data/.

Meng, Nicole, Dilara Keküllüoğlu, and Kami Vaniea. 2021. "Owning and Sharing: Privacy Perceptions of Smart Speaker Users." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW1): 1–29. https://doi.org/10.1145/3449119.

Pins, Dominik, and Fatemeh Alizadeh. 2021. "Without Being Asked: Identifying Use-Cases for Explanations in Interaction with Voice Assistants." In *CHI Workshop-Paper*, 8. Yokohama: N/A. https://doi.org/10.13140/RG.2.2.18764.33923.

Pins, Dominik, Alexander Boden, Britta Essing, and Gunnar Stevens. 2020. "'Miss Understandable': a study on how users appropriate voice assistants and deal with misunderstandings." In *MuC '20: Proceedings of the Conference on Mensch und Computer*, 349–59. Magdeburg: Association for Computing Machinery. https://doi.org/10.1145/3404983.3405511.

Pins, Dominik, Timo Jakobi, Alexander Boden, Fatemeh Alizadeh, and Volker Wulf. 2021. "Alexa, We Need to Talk: A Data Literacy Approach on Voice Assistants." In *Designing Interactive Systems Conference 2021*, 495–507. New York: Association for Computing Machinery. https://doi.org/10.1145/3461778.3462001.

Pins, Dominik, Timo Jakobi, Gunnar Stevens, Fatemeh Alizadeh, and Jana Krüger. 2022. "Finding, Getting and Understanding: The User Journey for the GDPR'S Right to Access." *Behaviour & Information Technology* 41 (10): 2174–2200. https://doi.org/10.1080/0144929X.2022.2074894.

Plantin, Jean-Christophe, Carl Lagoze, Paul N Edwards, and Christian Sandvig. 2018. "Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook." *New Media & Society* 20 (1): 293–10. https://doi.org/10.1177/1461444816661553.

Purington, Amanda, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. "Alexa Is My New BFF." In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems – CHI EA '17*, edited by Gloria Mark, Susan Fussell, Cliff Lampe, m.c schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor, 2853–59. New York: ACM Press. https://doi.org/10.1145/3027063.3053246.

Rohde, Markus, Peter Brödner, Gunnar Stevens, Matthias Betz, and Volker Wulf. 2017. "Grounded Design – a Praxeological IS Research Perspective." *Journal of Information Technology* 32 (2): 163–79. https://doi.org/10.1057/jit.2016.5.

Sadowski, Jathan. 2020. *Too Smart: How Digital Capitalism Is Extracting Data, Controlling Our Lives, and Taking Over the World*. The MIT Press. https://doi.org/10.7551/mitpress/12240.001.0001.

Schaub, Florian, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2018. "A Design Space for Effective Privacy Notices*." In *The Cambridge Handbook of Consumer Privacy*, edited by Evan Selinger, Jules Polonetsky, and Omer Tene, 365–93. Online: Cambridge University Press. https://doi.org/10.1017/9781316831960.021.

Schönherr, Lea, Maximilian Golla, Thorsten Eisenhofer, Jan Wiele, Dorothea Kolossa, and Thorsten Holz. 2020. "Unacceptable, Where Is My Privacy? Exploring Accidental Triggers of Smart Speakers." *arXiv:* 1 (1): 15. https://doi.org/10.48550/arXiv.2008.00508.

Sciuto, Alex, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "'Hey Alexa, What's Up?': A Mixed-Methods Studies of In-Home Conversational Agent Usage." In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 – DIS '18*, 857–68. Hong Kong: ACM Press. https://doi.org/10.1145/3196709.3196772.

Shafagh, Hossein, and Anwar Hithnawi. 2017. "Privacy-Preserving Quantified Self: Secure Sharing and Processing of Encrypted Small Data." In *Proceedings of the Workshop on Mobility in the Evolving Internet Architecture – MobiArch '17*, 25–30. Los Angeles: ACM Press. https://doi.org/10.1145/3097620.3097625.

Shin, Jungwoo, Yuri Park, and Daeho Lee. 2018. "Who Will Be Smart Home Users? An Analysis of Adoption and Diffusion of Smart Homes." *Technological Forecasting and Social Change* no. 134 (September): 246–53. https://doi.org/10.1016/j.techfore.2018.06.029.

Shneiderman, Ben. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, 336. USA: IEEE Computer Society.

Stevens, Gunnar, and Paul Bossauer. 2017. "Dealing with Personal Data in the Age of Big Data Economies." *Zeitschrift Für Geistiges Eigentum* 9 (3): 266. https://doi.org/10.1628/186723717X15069451170856.

Stevens, Gunnar, Paul Bossauer, Timo Jakobi, and Christina Pakusch. 2017. "Second Dashboard: Information Demands in a Connected Car." In *Tagungsband MuC 2017*, 215–26. Regensburg: Gesellschaft für Informatik e.V. https://doi.org/10.18420/muc2017-mci-0141.

Stevens, Gunnar, Volkmar Pipek, and Volker Wulf. 2010. "Appropriation Infrastructure: Mediating Appropriation and Production Work." *Journal of Organizational and End User Computing (JOEUC)* 22 (2): 58–81. https://doi.org/10.4018/joeuc.2010040104.

Strüver, Niklas. 2023a. "Frustration Free: How Alexa Orchestrates the Development of the Smart Home." *Digital Culture & Society* 9 (1): 99–124. https://doi.org/10.14361/dcs-2023-0106.

Strüver, Niklas 2023b. "Wieso eigentlich Alexa? Konzeptualisierung eines Sprachassistenten als Infrastruktur und Plattform im soziotechnischen Ökosystem Amazons." *kommunikation@gesellschaft* 24 (1). https://doi.org/10.15460/kommges.2023.24.1.1194.

Sun, Kaiwen, Yixin Zou, Jenny Radesky, Christopher Brooks, and Florian Schaub. 2021. "Child Safety in the Smart Home: Parents' Perceptions, Needs, and Mitigation Strategies." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–41. https://doi.org/10.1145/3479858.

Thakkar, Parth Kirankumar, Shijing He, Shiyu Xu, Danny Yuxing Huang, and Yaxing Yao. 2022. "'It Would Probably Turn into a Social Faux-Pas': Users' and Bystanders' Preferences of Privacy Awareness Mechanisms in Smart Homes." In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–13. New York: Association for Computing Machinery. https://doi.org/10.1145/3491102.3502137.

Tolmie, Peter, Andy Crabtree, Tom Rodden, James A Colley, and Ewa A Luger. 2016. "'This Has to Be the Cats' – Personal Data Legibility in Networked Sensing Systems." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing – CSCW '16*, 490–501. San Francisco: ACM Press. https://doi.org/10.1145/2818048.2819992.

Waldecker, David, Tim Hector, and Dagmar Hoffmann. 2023. "Intelligent Personal Assistants in Practice. Situational Agencies and the Multiple Forms of Cooperation without Consensus." *Convergence: The International Journal of Research into New Media Technologies* 0 (0): 1–17. https://doi.org/10.1177/13548565231189584.

Ware, Colin. 2013. *Information Visualization: Perception for Design*. Third edition. Interactive Technologies. Waltham: Morgan Kaufmann.

Whitten, Alma, and J D Tygar. 1999. "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0." In *Proceedings of the 8th USENIX Security Symposium*, 169–84. Washington, D.C.: USENIX Association.

Wulf, Volker, 2018. *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*. First edition. Oxford: Oxford University Press.

Wulf, Volker, Claudia Müller, Volkmar Pipek, David Randall, Markus Rohde, and Gunnar Stevens. 2015. "Practice-Based Computing: Empirically Grounded Conceptualizations Derived from Design Case Studies." In *Designing Socially Embedded Technologies in the Real-World*, edited by Volker Wulf, Kjeld Schmidt, and David Randall, 111–50. Computer Supported Cooperative Work. London: Springer. https://doi.org/10.1007/978-1-4471-6720-4_7.

Wulf, Volker, Markus Rohde, Volkmar Pipek, and Gunnar Stevens. 2011. "Engaging with Practices: Design Case Studies as a Research Framework in CSCW." In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 505–512. New York: Association for Computing Machinery. https://doi.org/10.1145/1958824.1958902.

Xie, Wenjing, Amy Fowler-Dawson, and Anita Tvauri. 2019. "Revealing the Relationship between Rational Fatalism and the Online Privacy Paradox." *Behaviour & Information Technology* 38 (7): 742–59. https://doi.org/10.1080/0144929X.2018.1552717.

Zhang, Dongpo. 2018. "Big Data Security and Privacy Protection." In *Proceedings of the 8th International Conference on Management and Computer Science (ICMCS 2018)*, 4. Shenyang: Atlantis Press. https://doi.org/10.2991/icmcs-18.2018.56.