Voice Assistants in Private Homes. Introduction to the Volume

Stephan Habscheid, Dagmar Hoffmann, Tim Hector, and David Waldecker

1. The Emergence of Voice Assistants

In 2011, together with the new iPhone 4S, Apple launched a voice assistant called "Siri"[1], which it claimed could understand questions and commands in spoken language (initially in English, German, and French) and respond to them as a human conversation partner would (Huq 2011). The announcement was met with great fascination: here was a talking technology, the first "intelligent personal assistance" system (IPA) to promise to make everyday life easier. With IPAs, it looked like a very popular, prototypical motif of science fiction was finally to become reality (Stresing 2011)[2]. Just a few years after "Siri", Amazon followed suit with "Alexa" (2015) and Google with its "Google Assistant" (2016) (cf. Dürscheid 2023), and by the end of the decade, the systems had become increasingly established in private households (Statista 2021). As well as in smartphone apps, voice assistants have been finding their way into various everyday devices, such as smart speakers, smart TVs, smart watches, or the media interfaces of digitally connected cars.

In recent years, however, the high-flying economic and technological hopes initially pinned to voice assistant technologies have been critically reappraised. In 2022, an article in *Business Insider* asserted that billions of invested dollars

1 Apple (under CEO Steve Jobs) had bought the company of the same name, which had been founded in 2007 and developed the product in 2010 (Wikipedia 2024; see also Dürscheid 2023).

2 This is not the place to retell the media history of talking machines (see, e.g., Volmar 2019). For a detailed media theoretical and linguistic description of the technologies relevant here with a specific focus on smart speakers, see Hector (in preparation).

had been lost, and hoped-for profits had not been achieved – a "colossal failure of imagination", in the words of a former employee of Amazon (Kim 2022). According to the *Business Insider* article,

> most of those conversations were trivial, commands to play music or ask about the weather. That meant fewer opportunities to monetize. Amazon can't make money from Alexa telling you the weather – and playing music through the Echo gives Amazon only a small piece of the proceeds. (Kim 2022)

In addition, since the introduction of new language processing technologies such as text-generating ChatGPT, which was launched in 2022, the earlier voice assistant technology risks appearing unwieldy in comparison (Dürscheid 2023). Against this backdrop, Amazon and other companies are apparently trying to integrate modern generative AI into older voice assistance systems. According to press reports, an assistance system presented by Amazon in 2023 spoke

> in a far more natural and conversational voice than the friendly-but-robotic one that hundreds of millions have become accustomed to communicating with for weather updates, reminders, timers, and music requests. (Goldman 2024)

According to the reports, this 'new Alexa' engaged more naturally in conversations, delivered more natural voice output, and had a more pronounced personality (ibid.). However, it seems that the version demonstrated has not yet been convincingly implemented into the real performance of the systems (ibid.). Thus, the American magazine *Fortune* has claimed that Amazon and Apple – once pioneers in the development of talking machines – are now "desperately behind [their] Big Tech rivals Google, Microsoft, and Meta in the race to launch AI chatbots and agents, and floundering (in [their] efforts to catch up)" (Goldman 2024). One reason given for this is that the characteristic technological architecture of older voice assistants is required to retain certain characteristics in order to maintain existing features, but therefore is no longer up to date enough for the integration of recent AI. In addition, these circumstances make it difficult to collect or synthetically produce suitable linguistic training data for the further development of the voice assistants. Citing former employees, the article reports that Amazon has therefore repeatedly

deprioritized the further development of Alexa to focus on the development of generative AI for its cloud computing unit ("Amazon Web Services") – which could see the existing technology soon becoming a "digital relic" (ibid., see also Herbig 2024).

On the other hand, current usage studies show that the number of devices with voice user interfaces for different smart technologies is continually increasing. Amazon has not confirmed the reports about Alexa's economic failure, and it is evidently continuing to invest in such products (Amazon 2023). For example, further development of devices that combine voice user interfaces with camera, monitor, and touch interfaces seems to be ongoing. As Niklas Strüver (2023) points out, smart speakers are conceived as the central hub for the smart home – a field of consumer tech that is clearly continuing to gain ground. Thus, in the smart home, devices like smart speakers are what allow users to manage the entire orchestration of multiple interconnected smart home applications related to the kitchen, housekeeping, or security. While technology companies see internet-enabled devices in the home as a way to increase demand for many such products and associated services, critics point out that many of the devices are too expensive for most consumers and will take years to catch on.

Either way, there are ample reasons to examine language-processing machines and their future development from the perspective of interaction research and linguistics. Not only does human–machine dialogue offer a fruitful field for investigation, it also points to potential new approaches to research on human–human interactions, as Karola Pitsch (2015) has shown with the example of co-constructions: familiar conversational procedures are "broken open", making analytical access to more basic conversational phenomena possible. Furthermore, as Martin Porcheron, Joel Fischer, Stuart Reeves, and Sarah Sharples observed in 2018, social interaction among co-present participants changes when the use of machines is incorporated. The linguistic contributions in this volume address human–machine communication as well as human–human communication and can be read together as an overview of current research in this field.

However, many other academic disciplines also address the phenomenon of human–computer interaction (HCI), albeit from different research perspectives. In the social sciences, the focus tends not to be primarily on usability or usage modalities, the skills that people need to have in order to operate the devices, but above all on exploring how devices integrated into everyday life are changing the ways we live together, how new media and data practices are de-

veloping, and how privacy is being reinterpreted (e.g., Burgess et al. 2022; see Ochs, this volume). Within the social sciences, a field of research is emerging that builds on existing theoretical paradigms (including actor-network theory, diffusion research, science and technology studies, surveillance studies, and mediatization research), but which is also developing new innovative and complex methodological approaches.

2. Controversial Discourses, Household Publics, and Everyday Practices

Assessments of voice assistants in public discourse vary widely (see Habscheid, Hector, and Hrncal, this volume). On the one hand, they are advertised as an addition to a digitally-connected and thus smart lifestyle (Hennig and Hauptmann 2019). As assistance systems, they are said to have the potential to compensate for handicaps and facilitate a self-determined life for older people (Endter, Fischer, and Wörle 2023). On the other hand, they are also subject to critique, because the devices provide manufacturers with users' voice data from a particularly sensitive context, the private domestic sphere (Sadowski 2020; Turow 2021), largely as a result of "cooperation without consensus" (Waldecker, Hector, and Hoffmann 2023; for the concept see Star 1993). Although voice-controlled assistance systems are embedded in social interaction and everyday practice, those who want to make full use of their functional potential must adapt to technologized dialogue structures and platform logics. In doing so, they have to reveal a lot about themselves that is transmitted beyond the household as 'data' where it can be analyzed and exploited in ways and for purposes that are opaque to the user. Furthermore, the creation of social order in such contexts can be distorted by problematic biases (see, for example, Leblebici in this volume).

It is in smart home environments that assistance systems as central interfaces come into their own, while at the same time opening up the household to the outside world far more than ever before. Whereas classic smart speakers' capacity for surveillance was limited to the perceptual mode of hearing (on "eavesdroppers" in physical or electronically mediated presence, see Goffman 1981, 132), smart homes incorporate camera-, monitor- and sensor-based systems and networks including various stationary and mobile devices and infrastructures, which can massively expand the scope for data collection. Under certain circumstances, this is accompanied by a further dissolution of the

boundaries of privacy, which on the one hand (e.g., in the case of surveillance of household members) may be perceived as abuse, but on the other hand (e.g., from a security perspective) may seem desirable.

As with all media, it is an open question as to how users will continue to adapt to new forms of media and how they make media adapt to the circumstances of their everyday lives. Individuals and households follow public discourse and interpret it in the light of their own household's public sphere, their concrete living conditions and interests. There is an ongoing debate within the humanities and social sciences, and especially within the domestication research paradigm (Hartmann 2023; Hector et al. 2023) on the adoption of digital media in household use settings. In principle, domestication research is based upon an analogy drawn between the process whereby media are appropriated and the process whereby cohabitation with farm animals or pets is established in the course of civilization. Domestication research, as summarized by Waldecker and Hector (2023, 5) "paints media as something that comes into the everyday life of users as foreign and wild, as something that has to be tamed and brought to relate to domestic routines". The metaphor of "taming" emphasizes the somewhat unpredictable and sometimes even threatening aspect of media technologies. This contrasts with the private household that often symbolizes a sense of security. With reference to Giddens (1984), Waldecker and Hector point out that this "ontological security" fundamentally establishes trust, supposedly guarantees the stability of one's own identity, the continuity of life and of the immediate environment. (Media) technologies that become entangled with this ontological security challenge it and can disrupt it: They become involved in everyday rituals, and even if ontological security is initially called into question by new media technologies (see Silverstone et al. 1992, 17), they (often) lose their threatening character as they are successively woven into everyday life, i.e., they become domesticated (see Bausinger 1984, 349–350).

In the process, everyday routines take new forms, and new practices emerge. First of all, new practices are required to get the novel devices and services to work at all. Further practices serve again and again to overcome the systems' technical unwieldiness and resistance. At the same time, the new usage practices become more or less deeply embedded in everyday life (see Waldecker and Hector 2023): They may (re-)shape, for example, the structuring of time between 'work' and 'leisure', ways of dealing with privacy, or the design of rooms and furnishings in the home. When users live together with other people in households, they must negotiate among themselves who uses which media, when, and how. In such contexts, economic decisions are also

discussed in connection with political and ultimately moral issues, such as whether to subscribe to streaming services, and if so, from which provider(s). Deliberation of such questions involves not only members of the household with its own power dynamics, but also voices from beyond the home; advice may be asked of friends, or sought in online forums or among reviews in which "online warm experts" reflect in accessible language on the possible uses of consumer technology as well as their limitations (Neville 2021; see also Waldecker and Hoffmann 2023).

3. Media Appropriation as a Linguistically Mediated Practice

Changing everyday practices as a consequence of media use is also observable at the linguistic level of everyday practice, all the more so when the technology concerned has a linguistic surface. This is the case, for example, for television, which is one of the classic mass media that has attracted particular interest in domestication and appropriation research. Unlike smart speakers, television does not require verbal input from users, neither at the level of content nor at the level of operation. Television broadcasts unidirectional communication, yet users have been shown to participate nonetheless. For television, "parainteraction" is characteristic, as Ayaß (1993) – drawing on Horton and Strauss (1957) as well as on Horton and Wohl (1986) – has shown: In unidirectional communication, forms of direct address and staged connection to everyday practices are used by on-screen performers to create an impression of interaction with those watching. Such utterances counterfactually imply that bidirectional interaction 'through' the screen could be possible (see also Böckmann et al. 2019, 145), and under certain circumstances, viewers pick up on this with forms of "parasocial" pseudointeraction in front of the screen (Ayaß 1993, 36).

The fact that in many cases the use of media is anchored in linguistic and interactional practice has been emphasized especially strongly by linguistic studies. These have addressed, among other topics, speaking while watching television together (Holly, Püschel, and Bergmann 2001) and intermission talk in theater (Gerwinski, Habscheid, and Linz 2018). It has been shown that viewers use the semiotic material their TV brings into the home as a resource for mutual "orientation" with respect to public issues (Holly 2001, 11–13). The studies also revealed that the appropriation of media – technologies as well as content – is affected not least by the possibilities of linguistic interac-

tion during and after reception. Examples include the format of "response cries" (Goffman 1981) and other forms of "terse speaking" (Baldauf 2002) in television-accompanied speech or reenactments and other reconstructive genres related to theater dialogues in intermission talk (Schlinkmann 2021). Accordingly, to study the appropriation of smart speaker technologies, it is necessary to ask how the linguistic conditions of their use enable and limit appropriation.

Unlike traditional television, internet technologies are two-way media: To put it bluntly, they not only bring the world into the household, but also the household into the world, with the latter in the form of specifically collected, aggregated, and processed data. The use of this data impacts on everyday life in ways that are noticed but cannot be traced, for example, in the form of personalized advertising or sensor-based environments in the smart home that adapt to usage habits. Thus, in the case of digital household technologies, not only are digital media domesticated in the home, households are also "externalized" (Brause and Blank 2020), or, in Hepp's words "deeply mediatized" (Hepp 2020).

The world that comes into the home with smart speakers is also linguistic on the surface – to a certain extent, it resembles the spoken language of interpersonal interaction. However, linguistic exchange with the machines differs not only in that dialogue involves non-human conversation partners, but that the technical language-processing systems upon which the latter depend have a limited ability to cooperate (Suchman 2007). The linguistic contributions to this volume discuss the range of forms such conversation can take: focusing on the human–machine dialogues, the social interaction they take place within, and the everyday practices that are realized – or not – as a result.

The sociotechnical relationships under discussion also raise fundamental questions for social theory. From a conversational linguistics perspective, the ANT approach, whereby all participating entities are conceptualized as equally significant actants (Latour 2005) seems unsatisfactory to us. For example, language-processing machines like Alexa are participants in practice, but not participants in social interaction as it is understood by conversation analysis (Habscheid 2023; Hector, in preparation; Habscheid, Hector, and Hrncal, this volume). From an ethnomethodological perspective, it can be shown that users orient towards machinic conversation partners with attitudes that, depending on the situation, sometimes reflect a more anthropomorphizing and at other times a more instrumentalized approach to the technology. Accordingly, Antonia Krummheuer (2010) characterizes the sociotechnical dialogue with an embodied conversational agent (ECA) as a "hybrid" or "ambiguous" exchange:

The exchange between human and machine shows similarities to interpersonal interaction, which is simulated to a certain extent (see also Hennig and Hauptmann, 2019), but also differences that require users to adapt to the limited communication capabilities of the machines (see also Lotze, this volume). Agency of the voice assistants is an object of negotiation both in everyday practice (Habscheid, Hector, and Hrncal 2023) and at the level of public discourse (Lind and Dickel 2024).

4. Smart Speaker Use and the Social Consequences for Everyday Reality

The use of digital technology is just as much a part of everyday life as the use of many other devices and communication with people who are physically present (Keppler 2018, 73). With the integration of a smart speaker into one's private household, this is extended by a technical artifact that is designed to function as a kind of interaction partner. Based on studies of social robotics, Michaela Pfadenhauer and Tobias Lehmann (2021) propose that a smart speaker can also be regarded as an "artificial companion" in everyday life. Smart speakers are expected to execute various commands as reliably as possible, search for and provide information, manage operation of networked devices, and offer services. Although their dialogue capabilities are still limited (Habscheid 2023) and communication is prone to disruption and often inconclusive or unpredictable (Pins et al. 2020; see also Lutz and Newlands, this volume), it can be assumed that this will improve significantly in the future, not least through the implementation of artificial intelligence. As an everyday companion, the smart speaker is certainly part of household communication: as an omnipresent third party. This participation at the locus of everyday life not only creates a social and emotional relationship with the device or with devices, but will also change how we communicate socially in everyday life. In the words of Hepp (2015), the communicative figuration of households, i.e., the communicative arrangement and role behavior of their members, is currently undergoing transformation. It is therefore of sociological interest to explore the extent to which the artificial companions can be regarded as "vehicles to cultural worlds of experience" (Pfadenhauer and Lehmann 2021) and prompt new fundamental questions of sociality (see also Hepp et al. 2022).

Furthermore, sociological investigation into sociotechnical practices and their consequences for the protection of privacy is called for. Through the

appropriation and use of smart technology, users reveal personal data about themselves (such as their taste in music, their shopping behavior, their account data, their everyday routines, their address book) and allow their home environment to be (acoustically) recorded. Huge volumes of data are transmitted to tech companies, stored, and evidently used as training data or for other purposes. Users are not always aware of this and it is largely beyond their control, although within the EU at least the Digital Service Act is intended to ensure greater transparency (see the conversation with Nikolai Horn in this volume). On this matter, it is important to examine users' own attitudes and explanations for how they deal with data protection and privacy. The narrative 'nothing to hide, nothing to fear' is expressed by many users as a pragmatic approach to data protection settings and issues for a variety of reasons (see Waldecker, Martin, and Hoffmann, this volume). Existing studies of ways of dealing with and justifying decisions concerning the data protection settings of digital applications have tended to neglect to consider those indirectly affected, such as visitors to households in which such devices are installed and used as a matter of course (e.g., Hoffmann 2023). Discourses on media and critiques of corporate data practices not only shape public debate, but are also negotiated in the private sphere (see Vermeulen and Mols, this volume). It remains to be seen how these smart technologies and media practices will 'conventionalize' in the future and how social scientists will study the ongoing developments.

5. On the Contributions in this Volume

This volume presents a wide spectrum of recent research on voice-operated systems and services, including analyses focusing on their (linguistically mediated) use and appropriation, on users' appraisals of them, and on the question of the exploitative utilization of the data they transmit. Perspectives from conversation analysis and media linguistics, media sociology, media studies, surveillance studies, the critique of political economy and related aspects of consumer research, domestication research, pragmatist and praxeological sociology as well as critical theory are brought together to shed light on the practical entanglement of users, devices, algorithms, data, and corporate interests. By encompassing these diverse approaches, this volume sets out to analyze the phenomenon of IPAs at multiple levels: from that of interaction, to everyday practices in households, to the level of users' perceptions and evaluations, and

not least in relation to global processes of data processing and exploitation. Our aim is to provide a comprehensive view of the transformation and persistence of everyday practices under platformized conditions and usage practices mediated by novel interfaces.

The majority of the contributions to this volume have evolved from presentations given at the conference "Voice Assistants in Private Homes. Media, Data, and Language in Interaction and Discourse", which took place on May 8 and 9, 2023, at the University of Siegen, Germany, organized by the research project "Un/desired observation in interaction: Intelligent Personal Assistants", which from 2020–2023 empirically investigated media practices with voice assistants as a key technology in the field of data-intensive digital media, taking a dual approach combining media sociology (Waldecker, Martin, and Hoffmann, this volume) and applied linguistics (Habscheid, Hector, and Hrncal, this volume). The project was part of the Collaborative Research Center "Media of Cooperation", which brings together numerous sub-projects investigating diverse phenomena but all taking as their point of departure a praxeological media theory paradigm that conceptualizes practice as the "mutual making of common goals, means and processes" and, in this context, media as "cooperatively created conditions of cooperation" or, in short, as "media of cooperation" (Schüttpelz 2017, 24). The "means" that can be cooperatively produced as "media" can – but do not have to – be of a linguistic nature (see Goodwin 2018; Habscheid, Hector, and Hrncal, this volume). In accordance with the interdisciplinary agenda of the Collaborative Research Center, as editors of this volume we seek to examine the complex phenomenon of data-intensive, AI-based assistance systems by addressing its multiple layers. The aim is to shed light on the intricate interrelationships between use and users, language, devices, algorithms, data, organizations, and economic exploitation.

The volume is structured in four parts. The first section – **Voice Assistants in Private Homes. Conceptual Considerations –** focuses on the theoretical foundations of key areas of IPA research and showcases various methodological approaches and findings of empirical studies. **Carsten Ochs** begins by examining the affective reactions of people who wonder why the users of smart speakers seem so unconcerned about their privacy. He traces the emergence of the modern practice of privacy protection, which was established in the 20th century, and now, since the advent of smart technologies in private homes, is being renegotiated. Ochs attempts to show what actually happens to the data collected and processed by smart speaker infrastructures that reach

into private homes, and concludes that the term "surveillance capitalism" is an apt one under the circumstances. Taking a Marxist approach and drawing on feminist theory, **Markus Kienscherf**'s contribution investigates the role of voice assistants in the reproduction of labor and capital. The appropriation of user-generated voice data by smart speakers is positioned within a more general history of the role of surveillance in the (re)production of capitalist social relations. The author shows that surveillance is central to the appropriation of surplus value in the spheres of production, the social reproduction of labor power, and the management of circulation and consumption. He then looks at the business models of tech companies and argues that the appropriation of user-generated data transmitted via smart speakers represents an extension of capitalist surveillance into the sphere of social reproduction.

The chapter by **Caja Thimm**, **Phillip Engelhardt**, and **Julia Schmitz** deals with anthropomorphism and communication accommodation to voice assistants. The focus is on how assistance systems with VUIs (voice user interfaces) are used and affectively engaged with in multi-person households, based upon a case study with households including physically impaired people with special support needs. The authors observe that these users' assumptions, attitudes, and expectations were not stable but varied according to contextual factors. As a theoretical basis for the research, "Communication Accommodation Theory" (CAT) is developed and adapted for the study of HCI constellations, focusing on strategies of anthropomorphization, which are shown to partially – and perhaps increasingly – influence the ways people interact with machines as well as to shape the discourse, interface design, and self-image of users. Last but not least, the authors reflect on the different insights into usage gained by their methodological combination of interviews and media diaries. The last contribution in this first section by **Netaya Lotze** traces the development of a complex sociolinguistic model that can bring findings concerning the anthropomorphization of HCI technologies together with evidence of cognitive and linguistic adaptation to the (more or less) limited communicative capacity of machines. After a comprehensive research overview, Lotze presents the results of her own studies conducted since 2000, which she summarizes and interprets in the light of the model (and vice versa). The model integrates various approaches from the philosophy of language, computer science, cognitive science, and linguistics, and is structured to take into account 'external factors', 'system variables', and 'user variables', while incorporating a user typology as well as enabling diachronic analysis.

Section 2, **Linguistic Exchange with Voice Assistants as a Practical Problem** presents studies from the field of linguistics examining the practical use of and critical discourse about language assistants. The chapter by **Florence Oloff** provides an empirically underpinned perspective on the usability and learnability of voice assistants as everyday technologies. Oloff examines specific instances of older users' first encounters, during adult education courses, with hitherto unknown voice-operated applications. She shows how, in non-profit, professionally guided practical training sessions, participants explore the potential benefits and problems of multimodal interfaces – the first stage of appropriation. Furthermore, a mismatch between the actual needs of the learners, the spatial, temporal, and medial limitations of the settings, and the teaching methods used by instructors to deal with these factors in an improvisational way becomes clearly evident. Oloff makes some suggestions on how to improve teaching and learning in these contexts. The contribution by **Didem Leblebici** provides insights into the experiences of Turkish-speaking users of non-Turkish-speaking voice assistants in Germany. The author expands upon a media linguistics interest in voice user interfaces by drawing on theoretical understandings of multilingualism from sociolinguistics and critical discourse analysis. The chapter, which is based on the linguistic analysis of ethnographic interview data, advances a critical discussion of the ways that language-processing technologies reinforce the standardization of language. Detailed examples are drawn upon to illustrate and analyze different phenomena of stylized language use in interaction with IPAs. A contribution by **Stephan Habscheid**, **Tim Hector,** and **Christine Hrncal** concludes this second section. The authors present an overview of the results to date from the linguistic strand of the project "Un/desired observation in interaction: Intelligent Personal Assistants". In the theoretical part of the chapter, the conceptual foundations of the "Media of Cooperation" Collaborative Research Center are further elaborated from a linguistic praxeological perspective, discussing approaches taken in interaction research on the one hand and linguistic media research on the other, as well as the domestication approach in media and communication research. This is followed by analyses of empirical findings from the research project, which underscore how instrumental linguistic practices are in embedding smart speakers into domestic routines, and illustrate how newly acquired technology reshapes social practices and communication within households.

The third thematic section brings together contributions that deal with the issues of **Privacy and Data Protection as Practical Problems**. Concerns relating to the extraction of personal data and its subsequent use have been raised

for almost as long as digital media technologies and applications have been available, with the disclosure of data from the domestic sphere often attracting particularly critical attention. One instance when contextual privacy is called into question is when devices behave unexpectedly. Glitches occur, which can appear as technical anomalies and expose critical privacy vulnerabilities. Taking a glitch studies approach, **Christoph Lutz** and **Gemma Newslands** focus on users' experiences of malfunctions, which can also have wider societal implications and raise questions about surveillance, data security, and the ethical responsibilities of technology companies. Although glitch studies is an interdisciplinary field that tends to use qualitative methods, Lutz and Newslands draw on quantitative data to identify the four most common glitches experienced by Amazon Echo users and how they categorize the consequences of those glitches in relation to levels of trust and concerns about privacy. The findings highlight a critical aspect of smart speaker technology: the delicate balance between their perceived benefits and the fears of potential negative consequences of using them. Such considerations and fears also play a major role when people decide whether (or not) to purchase voice-operated devices in the first place. In their chapter, **Jasper Vermeulen** and **Anouk Mols** present a multi-methods study that investigated the privacy perceptions of users and non-users of smart speakers. Based upon data from in-depth interviews and focus groups, they elaborate on Dutch users' and non-users' assessments of risks and benefits. They found that users generally appreciated affordances such as controllability, support, conversation, linkability, and recordability, while some indicated they would prefer greater transparency regarding corporations' use of data. Non-users associated recordability and locatability with privacy risks that were seen as significant enough to not use such technology at all. In addition to rational considerations, the study also pointed to the role of emotions in shaping adoption considerations and decisions. **David Waldecker**, **Alexander Martin**, and **Dagmar Hoffmann** also look at users' attitudes towards data protection issues in connection with the use of smart speakers and, in particular, how they deal with them. In doing so, they draw on various studies that show the extent to which users of digital media technologies develop a kind of "online apathy", "data protection cynicism", or even "digital resignation". Based on qualitative interviews with smart speaker users in Germany, the authors report how users cultivate certain attitudes towards the devices and the discourse surrounding them and how they explain their usage routines and pragmatic considerations. In addition to the findings of the studies cited, the authors' analysis of their own interviews reveals an attitude that Andreas Pettenkofer has termed "prag-

matic fatalism". Users who adopt such a stance more or less accept the data practices of companies and at the same time declare them to be irrelevant to their everyday lives. To conclude this section, **Nikolai Horn**, data protection expert and currently political advisor to iRights.Lab, discusses in a conversation with Dagmar Hoffmann and David Waldecker the legal and political aspects of protecting voice-based data. The new possibilities offered by AI and natural language processing are also addressed. Questions are raised about the extent to which voice recordings can be used to draw conclusions about identity characteristics of users and how voice recordings could be misused. The interview also explores the question of how users can be made more aware of data protection issues and how EU regulations such as the GDPR can ensure greater transparency in data use and give users more control of their own data.

The final (fourth) section – **Technical Infrastructures as a Practical Problem** – brings together a contribution from the field of social informatics and one from the sociology of technology to focus more explicitly on the IT processes and infrastructures that enable smart speaker technology but are not always transparent for users. Over a period of three years, **Dominik Pins**, **Fatemeh Alizahdeh**, **Alexander Boden**, **Sebastian Zilles,** and **Gunnar Stevens** used the living lab approach to investigate users' uncertainties with regard to the data collected as a consequence of their use of smart speakers in everyday life. Based on findings from interviews, field research, and participatory design workshops with 35 households, the authors developed a tool called "CheckMyVA" that supports users in accessing and visualizing their own VA data. The observations and findings presented in the chapter offer suggestions for tools and design strategies that could foster data literacy and enable users to reflect on their long-term interactions with VAs, ultimately "demystifying" the technology. The final chapter, by **Niklas Strüver,** takes a look behind the scenes to explore the practices involved in the ongoing development of automatic language processing. Amazon was once a pioneer in this field, but the launch of new large language models (LLMs) has posed major challenges for the company. Strüver conducted expert and narrative interviews with participants from university research teams who competed in the most recent Alexa Prize Competitions (APCs) to advance Alexa technology. These interviewees are able to offer fascinating insights into development practices, especially concerning the integration of LLMs into existing technology. Examining how the participants in these competitions deal with the conditions set by Amazon and the resources it makes available to competitors, Strüver outlines

some of the path dependencies, risks, benefits, and structuring aspects that participants encountered in their attempts to innovate Alexa.

It can be summarized that research in the field of smart technologies will certainly continue to be necessary, and that lines of inquiry are always shaped by disciplinary conventions, hence interdisciplinary exchange should continue to be promoted in the future.

Acknowledgements

As stated above, most of the contributions to this volume were developed from presentations held at the conference "Voice Assistants in Private Homes. Media, Data and Language in Interaction and Discourse" organized by the research project "Un/desired observation in interaction: Intelligent Personal Assistants". We express our gratitude to all those who participated in the conference and to all the authors who have contributed to this book. We would also like to thank our student assistants Chris Dangelmaier, Sarah Diehl, Aileen Halbe, Johanna Klein, Alexander Martin, Franziska Niersberger-Guéye, and Leonie Tittel, who helped to make the conference such a success.

The research project was part of the Collaborative Research Center "Media of Cooperation" at the University of Siegen, Germany, in its second funding phase (2020–2023). In its first funding phase (2016–2019), the project was led by Wolfgang Ludwig-Mayerhofer and investigated the strategic practices by which young people using social media tried to attract 'requested' attention and avoid 'undesired' observation. After Wolfgang Ludwig-Mayerhofer's retirement, Stephan Habscheid and Dagmar Hoffmann took over as principal investigators in the second funding phase, continuing to follow the general idea of un/desired observation but applying it to smart speakers with VUIs. This volume is a key publication that resulted from the research conducted in those four years. In its third funding phase (2024–2027), the project is once again shifting its focus, this time to smart home environments, in order to explore further transformations accompanying the recent massive expansion of interfaces as well of registrable data in smart homes through sensor-based mechanisms.

The project team in the second funding phase of the Collaborative Research Center (2020–2023) was led by principal investigators Stephan Habscheid and Dagmar Hoffmann, with Tim Hector, David Waldecker, Christine Hrncal, and Kathrin Englert as (post-)doctoral researchers. As the editors of this volume,

we would like to thank all of our colleagues for their long-term collaboration and their contribution to the success of this book. We want to thank Pip Hare for her extremely dedicated comprehensive copy editing as well as Sarah Diehl, Franziska Niersberger-Guéye, and Christopher Wegner for their extensive editorial work.

This research was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) – Project ID 262513311 – SFB 1187 Media of Cooperation.

References

Ayaß, Ruth.1993. "Auf der Suche nach dem verlorenen Zuschauer." In *Medienrezeption als Aneignung. Methoden und Perspektiven qualitativer Medienforschung*, edited by Werner Holly, and Ulrich Püschel, 27–41. Opladen: Westdeutscher Verlag.

Baldauf, Heike. 2002. *Knappes Sprechen*. Tübingen: Niemeyer.

Bausinger, Hermann. 1984. "Media, technology and daily life." *Media, Culture and Society* 6: 343–351. https://doi.org/10.1177/016344378400600403.

Böckmann, Barbara, Dorothee Meer, Michelle Mohn, Anastasia-Patricia Och, Ilaria Paltrinieri, Alina Renelt, Christine Ramdorf, Daniel Rettinghausen, Katharina Staubach, and Martin Tenz. 2019. "Multimodale Produktbewertungen in Influencer-Videos auf YouTube: Zur parainteraktionalen Konstruktion von Warenwelten." *Zeitschrift für Angewandte Linguistik* 70: 139–171. https://doi.org/10.1515/zfal-2019-2003.

Borbach, Christoph. 2018. "Sprachspiele|Stimmsynthesen: Zur Nachrichtentechnischen Genese Des Auditiven Pendants Von ELIZA." In *Hello, I'm Eliza: Fünfzig Jahre Gespräche Mit Computern*, edited by Marianna Baranovska and Stefan Höltgen, 177–98. Bochum: projektverlag.

Brause, Saba Rebecca, and Grant Blank: "Externalized domestication: smart speaker assistants, networks and domestication theory." *Information, Communication & Society* 23(5): 751–763. https://doi.org/10.1080/1369118X.2020.1713845.

Burgess, Jean, Kath Albury, Anthony McCosker, and Rowan Wilken. 2022. *Everyday data cultures*. Cambridge: Polity Press.

Dürscheid, Christa. 2023. "Mit Menschen und Chatbots. Digitale Kommunikation früher und heute. Conference Paper: LingColl 2022 in Tampere." *Re-*

search Gate. Accessed August 10, 2024. https://www.researchgate.net/publication/369506979_Mit_Menschen_und_ Chatbots.

Endter, Cordula, Florian Fischer, and Tobias Wörle. 2023. "Doing Home by Using Digital Assistive Technologies. On the Role of Meaning and Materiality in the use of Digital Assistive Technologies at Older People's Domesticity." *Digital Culture & Society* 9 (1): 179–200. https://doi.org/10.14361/dcs-2023-0109.

Englert, Kathrin, Dagmar Hoffmann, and David Waldecker. 2022. "Tut mir leid, ich verstehe nicht ganz – Smart Speaker als vermeintliche Gesprächspartner*innen." *merz | Zeitschrift für Medienpädagogik* (66) 2: 24–34.

Gerwinski, Jan, Stephan Habscheid, and Erika Linz. 2018. *Theater im Gespräch. Sprachliche Publikumspraktiken in der Theaterpause.* Berlin and Boston: de Gruyter.

Giddens, Anthony. 1984. *The Constitution of Society. Outline of a Theory of Structuration*. Cambridge: Polity Press.

Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Goldman, Sharon. 2024. "How Amazon blew Alexa's shot to dominate AI, according to more than a dozen employees who worked on it." *Fortune*, June 13, 2024 https://fortune.com/2024/06/12/amazon-insiders-why-new-alexa-llm-generative-ai-conversational-chatbot-missing-in-action/.

Goodwin, Charles. 2018. *Co-Operative Action*. New York: Cambridge University Press.

Habscheid, Stephan. 2023. "Socio-technical dialogue and linguistic interaction. Intelligent Personal Assistants (IPA) in the Private Home." *Sprache und Literatur* 51 (2): 167–196. https://doi.org/10.30965/25890859-05002020.

Habscheid, Stephan, Tim Hector, and Christine Hrncal. 2023. "Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology". *Social Interaction. Video-Based Studies of Human Sociality* 6 (1). https://doi.org/10.7146/si.v6i1.137378.

Hartmann, Maren, ed. 2023. *The Routledge Handbook of Media and Technology Domestication*. London and New York: Routledge.

Hector, Tim. In preparation. *Smart Speaker als Gesprächsbeteiligte? Sprachliche Praktiken im Prozess der Domestizierung mit Voice-User-Interfaces.* Dissertation, University of Siegen, Faculty of Arts and Humanities – Faculty I. Defense: April 12, 2024.

Hector, Tim, David Waldecker, Niklas Strüver, and Tanja Aal, eds. 2023. *Thematic issue: "Taming Digital Practices – On the Domestication of Data-Driven Technologies"*. *Digital Culture & Society* 9 (1).

Hennig, Martin and Kilian Hauptmann. 2019. "Alexa, optimier mich! KI-Fiktionen digitaler Assistenzsysteme in der Werbung." *Zeitschrift für Medienwissenschaft* 11 (2): 86–94. https://doi.org/10.25969/mediarep/12636.

Hepp, Andreas. 2020. *Deep mediatization*. Abingdon and New York: Routledge.

Hepp, Andreas. 2015. "Kommunikative Figurationen: Zur Beschreibung der Transformation mediatisierter Gesellschaften und Kulturen." In *Theorien des Medienwandels*, edited by Susanne Kinnebrock, Christian Schwarzenegger, and Thomas Birkner, 161–188. Köln: Herbert von Halem.

Hepp, Andreas, Wiebke Loosen, Stephan Dreyer, Juliane Jarke, Sigrid Kannengießer, Christian Katzenbach, Rainer Malaka, Michaela Pfadenhauer, Cornelius Puschmann, and Wolfgang Schulz. 2022. "Von der Mensch-Maschine-Interaktion zur kommunikativen KI. Automatisierung von Kommunikation als Gegenstand der Kommunikations- und Medienforschung." *Publizistik* 67 (4): 449–474. https://doi.org/10.1007/s11616-022-00758-4.

Herbig, Daniel. 2024. "Amazon: Verbesserte Alexa kommt laut Bericht nicht voran." *Heise online*, June 14, 2024. https://www.heise.de/news/Amazon-Verbesserte-Alexa-kommt-laut-Bericht-nicht-voran-9763110.html.

Hoffmann, Dagmar. 2023. Digitaler Alltag, Datenpraktiken und Subjektautonomie. In: *merz | Zeitschrift für Medienpädagogik*, 67 (1): 51–59.

Holly, Werner. 2001. "Der sprechende Zuschauer." In *Der sprechende Zuschauer. Wie wir uns Fernsehen kommunikativ aneignen*, edited by Werner Holly, Ulrich Püschel, and Jörg Bergmann, 12–24. Wiesbaden: Westdeutscher Verlag.

Holly, Werner, Ulrich Püschel, and Jörg Bergmann, eds. 2001. *Der sprechende Zuschauer. Wie wir uns Fernsehen kommunikativ aneignen*. Wiesbaden: Westdeutscher Verlag.

Horton, Donald, and Anselm Strauss. 1957. "Interaction in Audience Participation Shows." *American Journal of Sociology* 62 (6): 579–587.

Horton, Donald, and R. Richard Wohl. 1986. "Mass communication and parasocial interaction: Observation on intimacy at a distance." In *Inter/Media. Interpersonal communication in a media world* (3rd edition), edited by Gary Gumpert, and Robert Cathcart, 185–206. New York, Oxford: Oxford University Press.

Huq, Oliver. 2011. "iPhone 4S: Auf die inneren Werte kommt es an." *Heise online*, October 4, 2011. https://www.heise.de/news/iPhone-4S-Auf-die-inneren-Werte-kommt-es-an-1354456.html.

Keppler, Angela. 2018. "Medien, Lebenswelt und Alltagshandeln." In *Mediensoziologie. Handbuch für Wissenschaft und Studium*, edited by Dagmar Hoffmann and Rainer Winter, 71–85. Baden-Baden: Nomos. https://doi.org/10.5771/9783845264196-70.

Kim, Eugene. 2022. "Amazon is gutting its voice assistant, Alexa. Employees describe a division in crisis and huge losses on a wasted opportunity." *Business Insider*, November 19, 2022. https://www.businessinsider.com/amazon-alexa-job-layoffs-rise-and-fall-2022-11.

Krummheuer, Antonia. 2010. *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts.* Stuttgart: Lucius&Lucius.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory.* Oxford: Oxford University Press.

Lind, Miriam, and Sascha Dickel. 2024. "Speaking, but having no voice. Negotiating agency in advertisements for intelligent personal assistants." *Convergence* 30 (3): 1008–1024. https://doi.org/10.1177/13548565231192100.

Neville, Stephen J. 2021. "The domestication of privacy-invasive technology on YouTube: Unboxing the Amazon Echo with the online warm expert." *Convergence* 27 (5): 1288–1307. https://doi.org/10.1177/1354856520970729.

Pfadenhauer, Michaela, and Tobias Lehmann. 2022. "Affects after AI: Sociological perspectives on artificial companionship." In *The Routledge Social Science Handbook of AI*, edited by Anthony Elliott, 91–106. London: Routledge.

Pins, Dominik, Alexander Boden, Gunnar Stevens and Britta Essing. 2020. "'Miss understandable' – A study on how users appropriate voice assistants and deal with misunderstandings." In *Proceedings of Mensch und Computer 2020 (MUC20)*, 349–359. ACM, Magdeburg.

Pitsch, Karola. 2015. "Ko-Konstruktion in der Mensch-Roboter-Interaktion. Kontingenz, Erwartungen und Routinen in der Eröffnung." In *Ko-Konstruktionen in der Interaktion*, edited by Ulrich Krafft, Elisabeth Gülich, and Ulrich Dausendschön-Gay, 229–58. Bielefeld: transcript. https://doi.org/10.1515/9783839432952-013.

Porcheron, Martin, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. "Voice Interfaces in Everyday Life." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*, 1–12. https://doi.org/10.1145/3173574.3174214.

Sadowski, Jathan. 2020. *Too Smart. How Digital Capitalism is Extracting Data, Controlling Our Lives, and Taking Over the World.* Cambridge: MIT Press.

Schlinkmann, Eva (2021): *Kommunikative Rekonstruktion in der Theaterpause. Eine Gattungsanalyse von Pausengesprächen im Theater.* Stuttgart: Metzler.

Schüttpelz, Erhard. 2017. "Infrastructural Media and Public Media." *Media in Action* 1 (1): 13–61. http://dx.doi.org/10.25819/ubsi/7935.

Silverstone, Roger, Eric Hirsch, and David Morley. 1992. "Information and communication technologies and the moral economy of the household." In *Consuming Technologies. Media and Information in Domestic Spaces*, edited by Roger Silverstone and Eric Hirsch, 13–28. London: Routledge.

"Siri". Wikipedia article. Accessed August 10, 2024. https://en.wikipedia.org/wiki/Siri.

Star, Susan. 1993. "Cooperation Without Consensus in Scientific Problem Solving: Dynamics of closure in open Systems." In *CSCW. Cooperation or Conflict?*, edited by Dan Daiper, Closton Sanger, and Steve Easterbrock, 93–106. London: Springer. https://doi.org/10.1007/978-1-4471-1981-4_3.

Statista. 2021. "Statista Global Consumer Survey. ID 999790" Accessed August 21, 2024. https://de.statista.com/statistik/daten/studie/1271603/umfrage/anteil-der-haushalte-in-deutschland-mit-smart-speaker/.

Stresing, Laura. 2011. "Siri sagt: Warum die Apple-Sprachsoftware nicht nur Geeks fasziniert." *Tagesspiegel*, December 7, 2011. https://www.tagesspiegel.de/gesellschaft/medien-_-ki/warum-die-apple-sprachsoftware-nicht-nur-geeks-fasziniert-6708628.html.

Strüver, Niklas. 2023. "Frustration Free: How Alexa Orchestrates the Development of the Smart Home." *Digital Culture & Society* 9 (1): 99–123. https://doi.org/10.14361/dcs-2023-090106.

Suchman, Lucy. 2007. *Human-machine reconfigurations. Plans and situated actions* (2nd edition). Cambridge: Cambridge University Press.

Turow, Joseph. 2021. *The Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet.* New Haven and London: Yale University Press.

Volmar, Axel. 2019. "Productive Sounds." In *The Democratization of Artificial Intelligence*, edited by Andreas Sudmann, 55–76. Bielefeld: transcript. https://doi.org/10.1515/9783839447192-004.

Waldecker, David, and Tim Hector. 2023. "A Praxeological Approach Towards the Domestication of Connected Media Technologies. Introduction." *Digital Culture & Society* 9 (1): 5–22. https://doi.org/10.14361/dcs-2023-0102.

Waldecker, David, and Dagmar Hoffmann. 2023. "Inszenierung von kritischen Kompetenzen in Nischenöffentlichkeiten. Bewertungen von Smart Speakern auf YouTube." *kommunikation@gesellschaft* 23 (1). https://doi.org/10.15460/kommges.2022.23.1.1000.

Waldecker, David, Tim Hector, and Dagmar Hoffmann. 2024. "Intelligent Personal Assistants in practice. Situational agencies and the multiple forms of cooperation without consensus." *Convergence* 30 (3): 975–991.