COMMUNICATIVE FIGURATIONS Working Paper | No. 32
ISSN 2367-2277

Simone Natale

To believe in Siri: A critical analysis of AI voice assistants

Universität Bremen
Universität Hamburg
ZeMKI Zentrum für Medien-, Kommunikations- und Informationsforschung
LEIBNIZ-INSTITUT FÜR MEDIENFORSCHUNG HANS-BREDOW-INSTITUT

Forschungsverbund „Kommunikative Figurationen“ | Research Network „Communicative Figurations”
Universität Bremen | University of Bremen
ZeMKI, Zentrum für Medien-, Kommunikations- und Informationsforschung
Linzer Str. 4, 28359 Bremen, Germany, E-mail: zemki@uni-bremen.de
www.kommunikative-figurationen.de | www.communicative-figurations.org

Simone Natale (S.Natale@lboro.ac.uk)

Simone Natale is Senior Lecturer in Communication and Media Studies at Loughborough University, UK, and Assistant Editor of the journal Media, Culture & Society. In 2019, he has been ZeMKI Visiting Research Fellow at the University of Bremen, Germany. His work has been published in leading journals in his areas of interest, such as *New Media & Society*, *Communication Theory*, the *Journal of Communication*, *Convergence*, and *Media, Culture & Society*. He is the author of two monographs: *Supernatural Entertainments: Victorian Spiritualism and the Rise of Modern Media Culture* (Penn State University Press, 2016) and *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test* (forthcoming with Oxford University Press).

Working Paper No. 32, March 2020

Published by the „Communicative Figurations“ research network, ZeMKI, Centre for Media, Communication and Information Research, Linzer Str. 4, 28359 Bremen, Germany. The ZeMKI is a research centre of the University of Bremen.

ISSN: 2367-2277

To believe in Siri: A critical analysis of AI voice assistants

1 Introduction

"Talk to Siri as you would to a person," suggested Apple to the users of its AI voice assistant Siri in 2011, after it was bundled into the iPhone operating system (MacArthur, 2014). The message was meant to inspire a sense of familiarity with the assistant. Apple suggested that everything was already in place to accommodate the new technology in everyday experience: users just needed to extend their conversational habits to the invisible interlocutor embedded in the phone.

Given the swift success of Siri and other AI voice assistants in the following years, Apple's incitation might have worked. Similar tools were soon developed by other leading digital corporations: Amazon introduced Alexa in 2014, Google followed with its Assistant in 2016, while Microsoft's Cortana, now being discontinued, was launched even earlier, in 2013. In just a few years, the technology left the confined spaces of smartphones to dwell in all sorts of digital devices, from watches to tablets and speakers, inhabiting both domestic and professional environments. Just as graphic interfaces draw on visual information to facilitate interaction, AI voice assistants are based on software that recognizes and produces voice inputs. Users' commands and questions are then elaborated through language-processing algorithms that provide replies to the users' queries or execute tasks such as sending emails, searching on the Web, or turning on a lamp. Each assistant is represented as an individual character or persona (e.g. "Siri" or "Alexa") that despite being non-human can be imagined and interacted as such. As confirmed by market research and independent reports, they have been adopted by hundreds of millions of users around the world, making voice a key medium of interaction with networked computer technologies (Hoy, 2018).

The incitation of companies such as Apple to talk to voice assistants "as to a person," however, deserves to be questioned. Have AI voice assistants developed into something we talk to "as we would to a person," as promised by Siri's marketing lines? And if so, what does this even mean? Focusing on the cases of Alexa, Siri and Google Assistant, this working paper argues that AI voice assistants activate an ambivalent relationship with users, giving them the illusions of control in their interactions with the assistant while at the same time withdrawing them from actual control over the computing systems that lie behind the interface. I show how this is made possible at the interface level by mechanisms of projection that expect users to contribute to the construction of the assistant as a persona, and how this construction ultimately conceals the networked computing systems administered by the powerful corporations who developed these tools.

A critical analysis of AI voice assistants means unveiling the different strategies and mechanisms by which users are encouraged to accommodate existing social habits and behaviors so that they can "talk" to the AI assistant. Such strategies are not by any means straightforward, and do not correspond to tricking the users into believing that the AI thinks or feels "like a person." AI assistants rely on humans' tendency to project identity and humanity onto artifacts, but at the same time do not imply any decision from users regarding their ontology. In other words, they do not require users to decide if they are talking to a machine or to a person. They require them just to talk.

Although users ultimately benefit from the functionality of AI assistants and an enhanced capacity to accommodate the new technology in their everyday lives, one is left questioning

if it is safe to trust companies such as Apple, Amazon and Google to micro-manage more parts of our lives. The only way to find a response to this problem is looking through the complex stratification of technologies and practices that shape our relationship with these tools.

2 One and three

In the Christian theological tradition, God is “one and three”: Father, Son, Holy Spirit. This doctrine, called the Trinity, has stimulated lively theological discussions across many centuries. It is in fact one of the elements of the Christian faith that appears more confusing to believers: the idea that God’s three “persons” are distinct and one at the same time contrasts with widely-held assumptions about individuality, by which being one and being three are mutually exclusive (Torrance, 2016).

A similar difficulty also involves software. Many systems that are presented as individual entities are in fact the combination of separate programs applied to diverse tasks. The commercial graphic editor software package known as Photoshop, for instance, hides behind its popular trademark a complex stratification of discrete systems developed by different developers and teams across several decades (Lesage, 2016). When looking at software, the fact that what is one is also at the same time many should be taken not as the exception but as the norm. This certainly does not make software closer to God, but it does make it a bit more difficult to understand.

Contemporary AI voice assistants such as Alexa, Siri and Google Assistant are also one and many at the same time. On the one side, they offer themselves to users as individual systems with distinctive names and humanlike characteristics. On the other side, each assistant is actually the combination of many interconnected but distinct software systems that perform particular tasks. Alexa, for instance, is a complex assemblage of infrastructures, hardware artefacts and software systems, not to mention the dynamics of labor and exploitation that remain hidden to Amazon customers (Crawford and Joler, 2018). As BBC developer Henry Cooke put it, “there is not such a thing as Alexa” but only a multiplicity of discrete algorithmic processes. Yet Alexa is perceived as one thing by its users (Cooke, 2019).

Banal deception, a mundane and imperceptible form of deception that is embedded in software and computing design, operates by concealing the underlying functions of digital machines through a representation constructed at the level of the interface (see Natale, 2021). A critical analysis of banal deception, therefore, requires examination of the relationship between the two levels: the superficial level of the representation and the underlying mechanisms that are hidden under the surface, even while they contribute to the construction of the overlaid representation. In communicative AI, the representation layer also coincides with the stimulation of social engagement with the user. AI voice assistants draw on distinctive elements such as a recognizable voice, a name, and elements to suggest a distinctive persona such as “Alexa” or “Siri.” From the user’s point of view, a persona is above all an imagined construction, the feeling of a continuing relationship whose appearance can be counted on as a regular and dependable event, and integrated into the routines of daily life (Bucher, 2014). This imagined relationship helps users maintain the impression of a coherence in their interactions with the assistant.

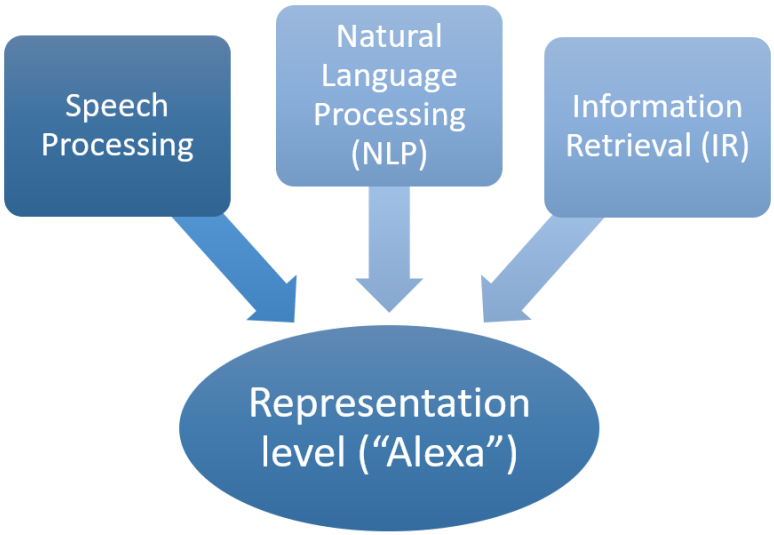

Figure 1 AI voice assistants as "one and three." Image created by the author.

In the hidden layer there are multiple software processes that operate in conjunction but are structurally and formally distinct. Although the entire "anatomy" of AI voice assistants is much more complex, three software systems are crucial to the functioning of AI voice assistants' banal deception, which roughly map to the areas of Speech Processing, Natural Language Processing (NLP), and Information Retrieval (IR) (fig. 1). The first one, Speech Processing, are algorithms that on the one hand "listen to" and transcribe the users' speech, and on the other produce the synthetic voice through which the assistants communicate with users (Nass and Brave, 2005). The second one, NLP, are the conversational programs that analyze the transcribed inputs and, like for a chatbot program, elaborate responses in natural language (Henrickson, 2018). Finally the third one, IR, are the algorithms that retrieve relevant information to respond to the users' queries and activate the relevant tasks. Compared to speech processing and NLP, the relevance of IR algorithms is perhaps less evident at first glance. However, they enable AI voice assistants to access Internet-based resources and to be configured as our proxies for navigating the Web (Wilks, 2019: 42-46). As the next sections will show, the differences between these three software systems are not restricted to their functions, since each of them is grounded in distinct approaches to computing and AIs, and carry with them different implications at both technical and social levels.

3 Speech processing, or the soft power of voice

Since the invention of media such as the phonograph and the telephone, scientists and engineers have developed a range of analog and digital systems to record and reproduce sound. Like all modern media, sound media were produced in the shape of the human user. For example, studies of human hearing were incorporated into the design of technologies such as the phonograph, which recorded and reproduced sound that matched sound frequencies perceived by humans (Sterne, 2012). A similar work of adaptation also involved the human voice, which was immediately envisaged as the key field of application for sound reproduction and recording (Laing, 1991). For instance, in 1878 the inventor of the phonograph Thomas Alva Edison imagined not music but voice recordings for note-taking or family

records as the most promising applications for its creation (Edison, 1978). Specific efforts were therefore made to improve the mediation of the voice.

Such endeavors profited from the fact that people are built or “wired” for speech (Nass and Brave, 2005). A human voice is more easily heard than other noises, and familiar voices are recognized with a precision barely matched by the vision of a known face. This quality comes to fruition in mediated communications. In cinema, for instance, the tendency of audiences to recognize a voice and immediately locate its source accomplishes several functions, adding cohesion to the narrative and instilling ‘life’ – i.e. presence and agency – to characters. Also in voice conversations over the phone or other media, the capacity to recognize and identify a disembodied voice is essential to the medium’s use. This skill enables phone users to recognize a familiar person and to gain hints about the demographics and mood of the voice’s owner. Thus the technological mediation of voice draws on the characteristics of human perception to generate meaningful results that are fundamental to the experience of the user.

Following from this technical and intellectual lineage, the dream of using spoken language as an interface to interact with computers is as old as computing itself (Licklider and Taylor, 1968). Until very recently, however, voice-based interfaces struggled to provide reliable services to users. Encountering automatic voice recognition technologies was often a frustrating experience: early “press or say one” phone services didn’t handle variations in accent or tone well, and users were often annoyed or amused by these systems’ struggles to comprehend even simple inputs (Duerr, 1966). Compared with the performances of Alexa or Siri, especially but not exclusively in the English language, one wonders how the technology could improve so markedly in such a short lapse of time.

The secret of this swift progress lies in one of the most significant technical changes that AI has experienced throughout its history: the rise of Deep Learning. Deep Learning is a class of machine learning algorithms that rely on complex statistical calculations performed by neural networks autonomously and without supervision. Inspired by the functioning of biological neurons, neural networks were proposed very early in the history of AI but initially seemed ineffective. In the 1980s and 1990s, new studies showed at a theoretical level that neural networks could be extremely powerful. Yet only in the last decade the technology realized its full potential, due to two main factors: advances in hardware that made computers able to process the complex calculations required by neural networks and, even more importantly, the availability of huge amounts of data, often produced by human users on the Internet, to ‘train’ the Deep Learning algorithms for the performance of specific tasks (Kelleher, 2019).

More broadly, Deep Learning has emerged in conjunction with a recalibration of human-computer interactive systems. For a growing range of AI applications, ‘intelligent’ skills are not programmed symbolically into the machine. Instead, they emerge through statistical elaboration of human-generated data that are harvested in computing networks through new forms of labor and automated power relations (Mühlhoff, 2019). Together with other applications such as image analysis and automatic translation, speech processing is one of the areas of AI that most benefited from the rise of Deep Learning. In the span of just a few years, the availability of masses of data that could be used to train the algorithms catalyzed a jump ahead in the automatic processing of the human voice by computers. Speech processing was in this sense the veritable killer application of AI assistants.

As the technical processing of the human voice became more sophisticated, companies that developed AI voice assistants took great care to adapt speech processing to their target users – exactly like analog sound media had profited from studies about the physiology and psychology of their audiences to improve the recording and reproduction of the voice

(Sterne, 2012). Significant thoughts were given to calibrating how the assistant's synthetic voice would sound to the ears of human users and to anticipate potential reactions to specific modulations.

Apple's Siri initially employed three different voiceover artists to represent the United States, Australia, and the United Kingdom (McKee & Porter, 2017: 167). This responded to the need for covering different English accents, as well as accommodating what Apple developers thought were cultural differences regarding perceptions of male versus female voices – hence the decision to employ a male voice in the UK case and a female voice in the US and Australia (Phan, 2017). Later, Apple included further accents for the English language, such as Irish and South African and, also in response to controversies about gender bias, allowed customization of gender. Google Assistant launched with a female default voice, but introduced a number of male voices and opted, also in reaction to controversies about sexism, for a random selection of one of the available voices as default (Google, 2020). Alexa has less voice customization options, although Amazon recently launched an remarkable new add-on, by which the voice of American actor Samuel Lee Jackson is made available to Alexa users for the affordable price of $0.99 (Kelion, 2019).

The fact that AI voice assistants have often featured female voices as default has been at the center of much criticism. There is evidence that people react to hints embedded in AI assistants' voices by applying categories and bias routinely attributed to people. This is particularly worrying if one considers the assistants' characterization as docile servants, which reproduces stereotypical representations of gendered labor. As argued by Thao Phan (2017), the representation of Alexa directs users towards an idealized vision of domestic service, departing from the historical reality of this form of labor, as the voice is suggestive of a native-speaking, educated, white woman. Similarly, Miriam Sweeney (2020) observes that most AI assistants' voices suggest a form of "'default whiteness' that is assumed of technologies (and users) unless otherwise indicated."

Although public controversies stimulated companies to increase the diversity of synthetic voices, hints to identity markers such as race, class and gender continue to be exploited to trigger a play of imagination that relies on existing knowledge and prejudice (Guzman, 2015). Studies in human-computer communication (e.g. Nass and Brave, 2005; Xu, 2019; Guzman, 2015) show that the work of projection is performed automatically by users: a voice is immediately attributed a specific gender, and even a race and class background. The notion of stereotyping, in this sense, helps us understand how AI assistants' disembodied voices activate mechanisms of projection that ultimately regulate their use. As Walter Lippmann showed in his classic study of public opinion, people could not handle their encounters with reality without taking for granted some basic representations of the world. In this regard, stereotypes have ambivalent outcomes: on the one side, they limit the depth and detail of people's insight into the world; on the other, they help people recognize patterns and apply interpretative categories built throughout time. While negative stereotypes need to be exposed and dispelled, Lippmann's work also shows that the use of stereotypes is essential to the functioning of mass media, since knowledge emerges both through discovery and through the application of pre-constituted categories (Lippman, 1922).

This explains why the main competitors in the AI assistant market select voices that convey specific information about a 'persona' – i.e. a representation of individuality that creates the feeling of a continuing relationship with the assistant itself. In other computerized services employing voice processing technologies, such as customer services, voices are sometimes made to sound neutral and evidently artificial. Such a choice, however, has been deemed untenable to companies whose AI assistants aspire to accompany users throughout their everyday lives. To function properly, these tools need to activate the mechanisms of representation by which users imagine a source for the voice – and, subsequently, a stable

character with which to interact, even if within relatively strictly boundaries (Bucher, 2012). As Lippmann has taught us, such mechanisms rely on previous stereotypes – which makes the choice of a gendered voice strategic, if extremely problematic, to companies such as Apple and Amazon.

By trusting them to apply their own stereotyping, AI voice assistants encourage users to contribute actively to the construction of sense around the disembodied voice. This serves not much to give ‘life’ to the assistants – despite all the anthropomorphic cues, users retain the ability to differentiate clearly between AI assistants and real persons (Guzman, 2015). More subtly, stereotyping helps users assign a coherent identity to assistants throughout time. This is achieved in the first instance by making the assistant’s voice recognizable: Siri and Alexa can only be assigned a coherent personality by users due to the fact that their voice always sounds the same (Nass and Brave, 2005: 143). The attribution of gender, race and class through stereotyping creates further clues to nurture the play of imagination involved in the users’ construction of a persona.

Thus the synthetic, humanlike voices of AI assistants as anthropomorphic cues are not meant to produce the illusion of talking to a human, but rather to create the psychological and social conditions for projecting an identity and, to some extent, a personality onto the virtual assistant. This banal form of deception does not imply any strict definition on the user’s part: one can grasp perfectly that Alexa is ‘just’ a piece of software and at the same time carry out socially meaningful exchanges with it. As it is ultimately left for the users to attribute social meaning, voice assistants leave ample space for individual interpretation.

This helps explain why research has shown that people construct their relationships with AI voice assistants in very diverse ways. For instance, in Andrea Guzman’s qualitative research with users of mobile conversational agents, participants gave a range of interpretations of aspects such as a voice’s source, some of them identifying it as a voice ‘in’ the mobile phone while others perceiving it as the voice ‘of’ the phone. The fact that “the locus and nature of the digital interlocutor is not uniform in people’s minds” is a result of the high degree of participation that is required from the user (Guzman, 2019: 343). Likewise, recent research shows that different users retain diverse types of benefits from their interactions with the assistants (McLean & Osei-frimpong, 2019). To use Marshall McLuhan’s term, AI voice assistants are a “cool” medium, a notion McLuhan applies to describe media such as television and the telephone that are low-definition and require participation on the part of the audience (McLuhan, 1964). The low definition of Alexa and other AI assistants leaves listeners to do the bulk of the work. As a result, though, Alexa is different things to different kinds of users. It is, after all, a necessity of any medium of mass diffusion to be able to adapt its message to diverse populations.

Still, the design of voice interfaces exercises an undeniable influence over users. Experimental evidence shows that synthetic voices can be manipulated to encourage the projection of demographic cues including gender, age and even race, as well as personality traits, such as an extroverted or an introverted, a docile or aggressive character (Kim & Sundar, 2012). Yet this is a ultimately a “soft,” indirect power, whereby the attribution of personality is delegated to the play of imagination of the users. The low definition of voice assistants contrasts with humanoid robots, whose material embodiment reduces the space of user contribution, as well as with graphic interfaces, which leave less space for the imagination (Hepp, forthcoming). The immaterial character of the disembodied voice should not to be seen, however, as a limitation: it is precisely this disembodiment that forces users to fill in the gaps and make AI voice assistants their own, integrating them more deeply into their everyday lives and identities. As a marketing line for Google Assistant recites, “it’s your own personal Google” (Google, 2019). The algorithms are the same for everybody, but you are expected to put a little bit of yourself into the machine.

4 Of haikus and commands: NLP and the dramaturgy of AI voice assistants

When I pick up my phone and ask Siri if she's intelligent, this is what appears on the screen:

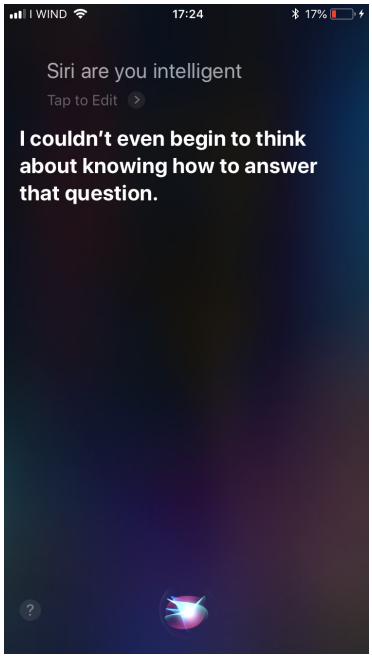

Figure 2. Author's conversation with Siri, 15 December 2019

To me, this is not just a turn of phrase. It's an inside joke that points to the long tradition of reflections about machines, intelligence and awareness – all the way back to Turing's 1950 paper, in which he argues that the question of whether machines can 'think' is irrelevant (Turing, 1950). Yet what at first glance looks like a clever reply is actually one of the least 'smart' things Siri can do. The reply is not the result of a sophisticated simulation of symbolic thought, nor has it emerged from statistical calculations of neural networks. More simply, it was manually added by some designer at Apple who decided that a question about Siri's intelligence should ignite such a reply. This is something a programmer would hesitate to describe as coding, dismissing it as little more than script writing or a 'programming trick'.

The recent, swift success of AI voice assistants has led many researchers and commentators to argue that these tools have clearly outrun the conversational skills of earlier conversational programs (Sweeney, 2020). Yet in contrast with such a widely held assumption, the likes of Alexa, Siri and Google Assistant do not depart dramatically from earlier systems, including chatbots, at least in terms of their conversational skills. This is due to the uneven degree of progress that AI has experienced in recent years. The rise of Deep Learning stimulated new expectations for the field and captured the public imagination with the vision that AI will equal or outrun humans in every kind of task in the near future. The myth of the thinking machine has been reignited, and the AI enterprise is experiencing a new wave of enthusiastic responses in the public sphere as well as in scientific circles (Natale & Ballatore, 2020). The picture, however, is much more complex. For all the potential of neural networks, not all areas of AI have benefited from the Deep Learning revolution. Due to the problem of retrieving and organising data about conversations and thus the difficulty of training algorithms to this task, conversational systems have until now only been slightly touched by Deep Learning. As technology journalist James Vincent put it, "machine learning is fantastic at learning vague rules in restricted tasks (like spotting the difference between cats and dogs or identifying skin cancer), but it can't easily turn a stack of data into the

complex, intersecting, and occasionally irrational guidelines that make up modern English speech” (Vincent, 2018).

Thus, while AI voice assistants represent a step ahead in communicative AI for areas such as voice processing, their handling of conversations still relies on a combination of technical as well as dramaturgical solutions. Their apparent proficiency is the fruit of some of the same strategies developed by chatbot developers across the last few decades, combined with an unprecedented amount of data about users’ queries that help developers anticipate their questions and design appropriate responses. The dramaturgical proficiency instilled in AI voice assistants at least partially compensates for the technical limitations of conversational programs.

In efforts to ensure that AI assistants reply with apparent sagacity and appear able to handle banter, Apple, Amazon and to a smaller degree Google assigned the task of scripting responses to dedicated creative teams (Stroda, 2020). Similar to Loebner Prize chatbots, which have been programmed to deflect questions and restrict the scope of the conversations, scripted responses allow voice assistants to conceal the limits of their conversational skills and maintain the illusion of humanity evoked by the talking voice. Every time it is asked for a haiku, for instance, Siri comes out with a different piece of this poetry genre, expressing reluctance (“You rarely ask me / what I want to do today / Hint: it’s not haiku”), asking the user for a recharge (“All day and night, / I have listened as you spoke. / Charge my battery”), or unenthusiastically evaluating the genre (“Haiku can be fun / but sometimes they don’t make sense. / Hippopotamus”).[1] Even if these scripted responses are unsophisticated on a technical level, their ironic tone can be striking to users, as shown by the many webpages and social media posts reporting some of the “funniest” and “hilariously honest” replies. AI voice assistants benefit from the fact that irony is perceived as evidence of sociability and sharpness of mind.

In contrast to chatbots such as ELIZA (Natale, 2019), the objective of AI voice assistants is not to deceive users into believing they are human. Yet the use of dramaturgical “tricks” allows AI voice assistants to achieve subtler but still significant effects. Scripted responses help create an appearance of personalization, as users are surprised to see Siri or Alexa reply to a question with an inventive line. The “trick” in this case is that AI assistants are also surveillance systems that constantly harvest data about users’ queries, which are transmitted and analyzed by the respective companies. As a consequence, AI assistant developers are able to anticipate some of the most common queries and have writers come out with appropriate answers. The consequentiality of this trick remains obscure to many users, creating the impression that the voice assistant is anticipating the user’s thoughts – which meets expectations of what a “personal” assistant is for. Users are thereby swayed into believing AI assistants to be more capable of autonomous behavior than they actually are. As noted by Margaret Boden, they appear “to be sensitive not only to topical relevance, but to personal relevance as well,” striking users as “superficially impressive.” (Boden, 2016: 65).

Also, simulated sociality is just one of the functions of AI voice assistants, not their *raison d’être*. Social engagement is never imposed on the user, but occurs only if users invite this behavior through specific queries. When told “goodnight,” for instance, Alexa will reply with scripts including “goodnight,” “sweet dreams,” and “hope you had a great day.” Such answers, however, are not activated if users just request an alarm for the next morning. Depending on the user’s input, AI assistants enact different modalities of interaction. Alexa, Google Assistant and Siri can be a funny party diversion one evening, exchange conviviality at night, and the next day return to being discreet voice controllers that just turn lights on and off (McLean and Osei-frimpong, 2019).

This is what makes AI assistants different from artificial companions, which are software and hardware systems purposely created for social companionship. Examples of artificial companions include robots such as Jibo, which combines smart home functionality with the appearance of empathy, as well as commercial chatbots like Replika (fig. 3), an AI mobile app that promises users comfort "if you're feeling down, or anxious, or just need someone to talk to" (Luka Inc., 2019). While Alexa, Siri and Google Assistant are only meant to play along if the user wants them to do so, artificial companions are purportedly programmed to seek communication and emotional engagement. If ignored for a couple of day, for instance, companionship chatbot Replika comes up with a friendly message, such as "Is everything OK?" or the rather overzealous "I'm so grateful for you and the days we have ahead of us."[2] Alexa and Siri, on the other hand, require incitement to engage in pleasantries or banter – coherently with one of the pillars of their human-computer interaction design, by which assistants speak up only if users pronounce the wake word. This is also why the design of smart speakers that provide AI voice assistant services in domestic environments, such as Amazon Echo, Google Home and Apple HomePod, is so minimal. The assistants are meant to remain seamless, always in the background, and quite politely intervene only when asked to do so (Woods, 2018; West, 2019).

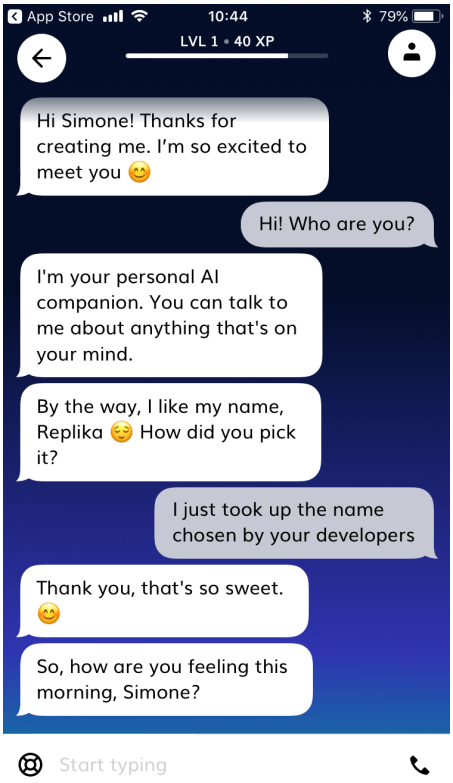

Figure 3. Author's conversation with Replika, 30 December 2019. Replika boosts a reported 2.5 million sign-ups and a community of circa 30,000 users on its Facebook group. It allows customization of gender (female, male and non-binary) and is programmed to be overly submissive and complimentary to the user, its stated purposes being: "Talk things through together," "Improve mental wellbeing," "Explore your personality" and "Grow together" (Luka Inc., 2019).

When not stimulated to act as companions, AI voice assistants treat conversational inputs as prompts to execute specific tasks. Conversations with AI assistants therefore profit from the fact that language, as shown by speech act theory, is also a form of action (Winograd, 1980). In fact, some (e.g. Heidorn, 1974) have proposed that natural language is equivalent to a very high-level programming language. It is easy to see how this applies to the case of communicative AI interfaces such as the assistants. In computing, every line of code is

translated into lower-level specific commands all the way down to the physical operations performed at the level of the machine. Similarly, when users ask Siri or Alexa to “call Mum” or “play BBC radio” these inputs are translated into the corresponding instructions in lower-level programming languages to initiate corresponding functions. An expert user of voice assistants will learn which commands most effectively have Alexa, Siri or Google Assistant ‘understand’ and operate accordingly – similar to how a computer scientist memorizes the most frequent commands specific to a programming language.

The variability of AI voice assistants’ approaches to interaction is exemplary of the move from the straight-out deception of chatbots such as ELIZA or competitors in the Loebner Prize (an annual contest for chatbots that pretend to be human) to a banal form of deception that, faithful to the principles of transparent design and user friendliness, gives users at least the appearance of control (Natale, 2021). When ignited by appropriate queries, Siri, Alexa and Google Assistant engage in forms of simulated sociality. When approached with a request, they return to their role of docile, silent aides controlled through the most accessible of all programming languages: natural language. In this process, users experience both control and the lack of it. On the one hand, they are ultimately responsible for establishing the tone and scope of conversations. On the other hand, they have limited insight into the deceptive mechanisms of AI assistants, which are embedded in code that is obscure to most users and rely on unprecedented degrees of knowledge about the users themselves. The incessant accumulation of data about users’ behavior, in fact, ensures that Apple, Amazon and Google can manage the delicate balance between such contrasting needs: conceding the illusion of control to users while retaining actual control for themselves.

5 Search for me, Alexa: Information Retrieval, AI assistants and the shape of the Internet

Both speech processing and the handling of conversation are examples of the “low definition” of AI voice assistants. On the one side, the processing of voice requires users to contribute to the construction of the assistant’s persona through stereotyping. On the other, the conversational routines of AI assistants adapt to the users’ queries, offering different modalities of interaction and leaving users with the illusion of control over the experience. In contrast, the third system – Information Retrieval – has more to do with the tasks that AI voice assistants are able to complete than with users’ perceptions of computing systems. Yet there is a close link between the apparently neutral and mundane character of IR operations and the overall representation level of each assistant, which in turn helps us understand how users are withdrawn control over networked computing systems.

IR refers to systems that enable the localization of information relevant to specific queries or searchers (Manning, Raghavan & Schütze 2008). Most notably, IR regulates the functioning of Web search engines such as Google Search and, more generally, the retrieval of information across the Web. However, little attention has been given to the fact that IR also plays a key role in AI voice assistants. To properly function, AI voice assistants such as Alexa, Siri and Google Assistant need to be constantly connected to the Internet, through which they retrieve information and access services and resources. Internet access allows these systems to perform functions including responding to queries, providing information and news, streaming music and other online media, managing communications including emails and messaging, as well as controlling smart devices in the home such as lights or heating systems (Bentley et al., 2019). Although AI voice assistants are scarcely examined in their quality of interfaces giving access to Internet-based resources, they are ultimately technologies that provide new pathways to navigate the Web through the mediation of huge corporations and their cloud services. As they enter more and more into public use, therefore,

they also inform how the Web and other resources are employed, perceived and understood by users.

One of the key features of the Web is the seemingly endless amount of information that is accessible through it. As of September 2019, it is estimated that there are more than 1 billion 700 million websites online.[3] To navigate such an imposing mass of information users employ browsers, search engines and social networks as interfaces that help them identify and connect to specific webpages, media and services. Each of these interfaces restricts the focus towards a more manageable range of information that is supposed to be tailored to or for the user. Yet these interfaces also have their own biases that inform users' experiences of the Web itself. Search engines, for instance, index not all but only parts of the Web, influencing the visibility of different pieces of information based on factors including location, language and previous searches. Likewise, social networks such as Facebook and Twitter impact on access to information, due to the algorithms that decide on the appearance and ranking of different posts as well as to the 'filter bubble' by which users tend to become distanced from information not aligning with their viewpoints (Bozdag, 2013; Willson, 2014).

To some extent, each of these interfaces could be seen as empowering users, as they help them retrieve information they need. Yet this empowerment also corresponds to a loss of control from the part of the user. It is for this reason that researchers since the emergence of the Web have kept interrogating if and to what extent different tools for Web navigation facilitate or hinder access to a plurality of information (Thorson and Wells, 2016). The same question urgently needs to be asked for AI voice assistants. Constructing a persona within the interface, voice assistants mobilize specific representations while they ultimately reduce control from users over the Web's access, jeopardizing their capacity to browse, explore and retrieve a plurality of information available through the Web.

A comparison between the search engine Google and the voice assistant Google Assistant is useful at this point. If users search one item on the search engine, say "romanticism," they are pointed to customized entries from Wikipedia and the Oxford English Dictionary alongside a plethora of other sources. Although studies have demonstrated that most users rarely go beyond the first page of a search engine's results (Goldman, 2007), the interface still enables users to browse at least a few of the 16,400,000 results retrieved through their search. The same input given to Google Assistant (at least, the version of Google Assistant on my phone) links only to the Wikipedia page for "Romanticism," the artistic movement. Other meanings for the same words are disregarded in the initial search, and one single source is privileged by the system. If the bias of Google algorithms applies to both the search engine and the virtual assistant, in Google Assistant browsing is completely obliterated and substituted by the triumph of "I'm feeling lucky" searches delivering a single result. Due to the time that would be needed to provide several potential answers by voice, the relative restriction of options is to be considered not just a design choice but a characteristic of AI voice assistants as a medium.

Emily MacArthur has pointed out that a tool such as Siri "restores a sense of authenticity to the realm of Web search, making it more like a conversation between humans than an interaction with a computer" (2014: 117). One wonders, however, if such a "sense of authenticity" is a way for AI voice assistants to appear at the service of the users, to make us forget that they are at the service of the companies that developed them. In spite of their imagined personae, "Alexa," "Siri" and "Google Assistant" never exist on their own. They exist only as embedded within a hidden system of material and algorithmic structures that guarantee market dominance to companies such as Amazon, Apple and Google. They are gateways to the cloud-based resources administered by these companies, eroding the distinction between the Web and the proprietary cloud services that are controlled by these

huge corporations. This erosion is achieved through close interplay between the representation staged by the digital persona embodied by each assistant and its respective company's business model.

It is striking to observe that the specific characterizations of each AI voice assistant is strictly related to the overall business and marketing approaches of each company. Alexa is presented as a docile servant, able to inhabit domestic spaces without crossing the boundaries between 'master' and 'servant'. This contributes to hiding Amazon's material structures of labor and the precarious workforce that sustains the functionality of the platform (Hill, 2019). Thus, Alexa's demeanor contributes to make the exploitation of Amazon's workforce seamless and invisible to the customers/users who access Amazon Prime services and online commerce through the docile assistant. In turn, Siri, compared to the other main assistants, is the one that makes the most extensive use of irony. This helps corroborate Apple's corporate image of creativity and uniqueness, which the company attempts to project onto its customers' self-representation: "stay hungry stay foolish," as recited in a famous Apple marketing line (Magaudda, 2015). In contrast with Apple and Amazon, Google chose to give their assistant less evident markers of personal identity, avoiding even the use of a name. What appears as refusal to characterize the assistant, however, actually reflects Google's wider marketing strategy, which has always downplayed elements of personality (think of the low profile, compared to Steve Jobs for Apple or Jeff Bezos for Amazon, of Google's founders Larry Page and Sergei Brin) to present Google as a quasi-immanent oracle aspiring to become indistinguishable from the Web (Peters, 2015). Google Assistant perpetuates this representation by offering itself as an all-knowing entity that promises to have an answer for everything and is "ready to help, wherever you are" (Google, 2020).

Rather than being separated from the actual operations that AI voice assistants carry out, the construction of the assistant's persona is meant to feed into the business of the corporations. In fact, through the lens of Siri or Alexa, there is no substantial difference between the Web and the cloud-based services administered by Apple and Amazon. Although interfaces are sometimes seen as secondary to the communication that ensues through them, they contribute powerfully to shape the experience of users. It is for this reason that AI voice assistants' interface quality needs to be taken seriously. Like the metaphors and representations evoked by other interfaces, the construction of the assistant's persona is not neutral but informs the very outcome of the communication. In providing access to the Web, AI voice assistants reshape and repurpose the Web as something that responds more closely to how companies such as Amazon, Apple and Google want it to look like for their customers.

6 Conclusion: To deceive and not to deceive

AI voice assistants represent a new climax in the convergence between AI and human-computer interaction (Guzman and Lewis, 2019; Hepp, forthcoming). As media studies scholars have shown, all interfaces employ metaphors, narrative tropes and other forms of representation to orient interactions between users and machines towards specific goals (Chun, 2011; Emerson, 2014).[4] Graphic interfaces, for instance, employ metaphors such as the desktop and the bin, constructing a virtual environment that hides the complexity of operating systems through the presentation of elements familiar to the user. The metaphors and tropes manifested in the interface inform the imaginary constructions through which people perceive, understand and imagine how computing technologies work (Bucher, 2016).

In AI voice assistants, the level of representation coincides with the construction of a 'persona' that creates the feeling of a continuing relationship with the assistant. This adds further complexity to the interface, which ceases to be just a point of intersection between user and computer, taking up the role of both the channel and the producer of communication. The literal meaning of *medium*, "what is in between" in Latin, is subverted by AI assistants that reconfigure mediation as a circular process in which the medium acts at the same time as the endpoint of the communication process. An interaction with an AI assistant, in this sense, may restore a sense of authenticity in interactions with computers, but also results in creating additional distance between the user and the information retrieved in the Web through the indirect management of the interface itself.

The way this distancing is created through the application of a banal, normalized form of deception to AI voice assistants has been illuminated in this paper. Mobilizing a plurality of technical systems and design strategies, AI voice assistants represent the continuation of a longer trajectory within communicative AI. Computer interface design emerged as a form of collaboration in which users do not much 'fall' into deception as participate in constructing the representation that creates the very possibility for interacting with computing systems (Emerson, 2014). In line with this mechanism, there is a structural ambivalence in AI assistants that results from the complex exchanges between the software and the user, whereby the machine is adapted to the human so that the human can project its own meanings into the machine.

AI voice assistants such as Alexa and Siri are not trying to fool anyone into believing that they are human. Yet, as this working paper shows, their functioning is strictly bounded to a "banal" form of deception (see Natale, 2021) that benefits from cutting-edge technical innovations in neural networks as well as from dramaturgical strategies established throughout decades of experimentation within communicative AI. Despite not being credible as humans, therefore, they are still capable of fooling us. This seems a contradiction only so long as one believes that deception involves a binary decision: if we are, in other words, either "fooled" or "not fooled." AI voice assistants and other AI-based technologies demonstrate that this is not the case: technologies incorporate deception in more nuanced and oblique ways than is usually acknowledged.

Acknowledgments

This working paper is an early draft of materials forthcoming in Simone Natale, *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test* (New York: Oxford University Press, 2021). I would like to thank Leah Henrickson for providing feedback on an earlier version of this text. Research leading to the publication of this working paper was funded through a ZeMKI Visiting Research Fellowship in 2019-20.

7 References

Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction (TOCHI), 26*(3), 1–27.

Ballatore, A. (2015). Google chemtrails: A methodology to analyze topic representation in search engine results. *First Monday, 20*(7).

Boden, M. (2016). *AI: Its nature and future*. Oxford: Oxford University Press.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology, 15*(3), 209–227.

Bucher, T. (2016). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086
Bucher, T. (2014). About a bot: Hoax, fake, performance art. *M/C Journal*, *17*(3).
Chun, W. H. K. (2011). *Programmed visions: Software and memory*. Cambridge, Mass.: MIT Press.
Cooke, H. (2019) Talking with Machines. Presentation delivered at the *Mediated Text Symposium*, Loughborough University, London, 5 April 2019.
Crawford, K., & Joler, V. (2018). Anatomy of an AI System. Retrieved September 20, 2019, from https://anatomyof.ai/
Duerr, R. (1996). Voice recognition in the telecommunications industry. In *Professional Program Proceedings. ELECTRO '96* (pp. 65–74).
Edison, T. A. (1878). The phonograph and its future. *The North American Review*, *126*(262), 527–536.
Emerson, L. (2014). *Reading writing interfaces: From the digital to the book bound*. Minneapolis: University of Minnesota Press.
Goldman, E. (2008). Search engine bias and the demise of search engine utopianism. In A. Spink & M. Zimmer (Eds.), *Web Search* (pp. 121–133). Berlin: Springer.
Google (2020). Choose the Voice of Your Google Assistant. Retrieved January 3, 2020), from http://support.google.com/assistant
Google (2019). Google Assistant. Retrieved December 12, 2019, from https://assistant.google.com/
Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343–350.
Guzman, A. L. (2015). *Imagining the Voice in the Machine: The Ontology of Digital Social Agents*. PhD Dissertation, University of Illinois at Chicago.
Guzman, A. L., & Lewis, S. C. (2019). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691
Heidorn, G. E. (1974) English as a Very High Level Language for Simulation Programming. *ACM SIGPLAN Notices* 9(4), 91-100.
Henrickson, L. (2018). Tool vs . agent : attributing agency to natural language generation systems systems. *Digital Creativity*, *29*(2–3), 182–190.
Hepp, A. (forthcoming) Artificial Companions, Social Bots and Work Bots: Communicative Robots as Research Objects of Media and Communication Studies. *Media, Culture and Society*.
Hill, D. W. (2019). The injuries of platform logistics. *Media, Culture & Society*. https://doi.org/10.1177/0163443719861840
Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88.
Kelion, L. (2019) Amazon Alexa Gets Samuel L Jackson and Celebrity Voices. *BBC News*, 25 September 2019. Retrieved 12 December 2019, from https://www.bbc.co.uk/news/technology-49829391
Kelleher, J. D. (2019). *Deep Learning*. Cambridge, Mass.: MIT Press.
Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241–250.
Laing, D. (1991). A voice without a face: popular music and the phonograph in the 1890s. *Popular Music*, *10*(1), 1–9.
Lesage, F. (2016). A Cultural Biography of Application Software. In C. Paterson, D. Lee, & A. Saha (Eds.), *Advancing Media Production Research: Shifting Sites, Methods, and Politics* (pp. 217–232). Basingstoke, UK: Palgrave Macmillan.
Licklider, J. C. R., & Taylor, R. W. (1968). The Computer as a Communication Device. *Science and Technology*, *2*(3), 2–5.
Lippmann, W. (1922). *Public opinion*. New York: Harcourt, Brace & Co.
Luka Inc. (2019) Replika. Retrived December 30, 2019, from https://replika.ai/
MacArthur, E. (2014). The iPhone Erfahrung: Siri, the auditory unconscious, and Walter Benjamin's Aura. In D. M. Weiss, A. D. Propen, & C. Emmerson Reid (Eds.), *Design, Mediation, and the Posthuman* (pp. 113–127). Lanham: Lexington Books.
Magaudda, P. (2015). Apple's Iconicity: Digital Society, Consumer Culture and the Iconic Power of Technology. *Sociologica*, *9*(1), 0.
Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press.
McKee, H. A., & Porter, J. E. (2017). *Professional communication and network interaction: A rhetorical and ethical approach*. London: Routledge.

McLean, G., & Osei-frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28–37.
McLuhan, M. (1964). *Understanding media: The extensions of man*. Toronto: McGraw-Hill.
Mühlhoff, R. (2019). Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. *New Media and Society*. https://doi.org/10.1177/1461444819885334
Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, Mass.: MIT press.
Natale, S. (2021). *Deceitful media: Artificial Intelligence and social life after the Turing Test*. New York: Oxford University Press.
Natale, S. (2019). If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA. *New Media & Society*, *21*(3), 712–728.
Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of Artificial Intelligence. *Convergence: The International Journal of Research into New Media Technologies*, *26*(1), 3–18.
Peters, J. D. (2015). *The marvelous cloud: Towards a philosophy of elemental media*. Chicago: University of Chicago Press.
Phan, T. (2017). The Materiality of the Digital and the Gendered Voice of Siri. *Transformations*, *29*, 23–33.
Sterne, J. (2012). *MP3: The meaning of a format*. Durham: Duke University Press.
Stroda, U. (2020). 'Siri, tell me a joke': Is there laughter in a transhuman future? In *Spiritualities, ethics, and implications of human enhancement and artificial intelligence* (pp. 69–85). Wilmington, De.: Vernon Press.
Sweeney, M. E. (2020). Digital Assistants. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (Eds.), *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, Mass.: MIT Press.
Thorson, K., & Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, *26*(3), 309–328.
Torrance, T. F. (2016). *The Christian doctrine of God, one being three persons*. London: Bloomsbury.
Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.
Vincent, J. (2018). Inside Amazon's $3.5 Million Competition to Make Alexa Chat like a Human'. *The Verge*, 13 June 2018, retrieved January 12, 2020 from https://www.theverge.com/2018/6/13/17453994/amazon-alexa-prize-2018-competition-conversational-ai-chatbots
Xu, K. (2019). First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses. *New Media & Society*, *21*(11–12), 2522–2547.
Wilks, Y. (2019). *Artificial Intelligence: Modern Magic or Dangerous Future?* London: Icon Books.
Winograd, T. (1980). What Does it Mean to Understand Language? *Cognitive Science*, *4*(3), 209–241. https://doi.org/10.1207/s15516709cog0403_1
West, E. (2019). Amazon: Surveillance as a Service. *Surveillance & Society*, *17*(1/2), 27–33.
Willson, M. (2014). The politics of social filtering. *Convergence*, *20*(2), 218–232.
Woods, H. S. (2018). Asking more of Siri and Alexa: Feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication*, *35*(4), 334–349. https://doi.org/10.1080/15295036.2018.1488082

Endnotes

[1] Author's conversations with Siri, 15 December 2019.
[2] Author's conversation with Replika, 2 December 2019.
[3] Data from https://www.internetlivestats.com).
[4] Chun, *Programmed Visions*; Galloway, *The Interface Effect*.