Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces

Simone Natale
Loughborough University, UK; University of Turin, Italy

Henry Cooke
BBC Research & Development, UK

Abstract

Voice assistants such as Siri, Alexa, and Google Assistant have recently been the subject of lively debates in regard to issues such as artificial intelligence, surveillance, gender stereotypes, and privacy. Less attention, however, has been given to the fact that voice assistants are also web interfaces that might impact on how the web is accessed, understood and employed by users. This article aims to advance work in this context by identifying a range of issues that should spark additional reflections and discussions within communication and media studies and related fields. In particular, the article focuses on three key issues that have to do with long-standing discussions about the social and political impact of the internet: the role of web platforms in shaping information access, the relationship between production and consumption online, and the role of affect in informing engagement with web resources. Considering these issues in regard to voice assistants not only helps contextualize these technologies within existing debates in communication and media studies, but also highlights that voice assistants pose novel questions to internet research, challenging assumptions of what the web looks like as speech becomes one of the key ways to access resources and information online.

Corresponding author:
Simone Natale, University of Turin, via sant'ottavio 20, 10124 Torino, Italy.
Email: simone.natale@unito.it

The emergence of voice assistants such as Amazon's Alexa and Apple's Siri has recently sparked much debate. While techno-enthusiasts greet these technologies as a new epoch for human-computer interaction and for their potential benefits in regard to issues such as accessibility (Vtyurina et al., 2019), critics highlight that they produce or reproduce stereotypes of gender and race (Schiller and McMahon, 2019) and raise pressing questions regarding surveillance capitalism and privacy (Woods, 2018). Until now, however, there has been little critical examination of the quality of voice assistants as an interface to the internet. Yet, Amazon's Alexa, Apple's Siri, and Google Assistant – to mention some of the most popular of such tools (Hoy, 2018) – are online-based technologies providing new pathways to access the Internet through the mediation of huge corporations and their cloud services. These technologies, therefore, bear profound implications on how the Web and other resources are perceived, understood and employed by users.

Considering that the Internet, as widely acknowledged, plays a key role in access to information and the formation of the public sphere (Papacharissi, 2002), it is crucial to interrogate the dynamics of mediation enacted by voice assistants and how these impact on social, cultural, and political structures. If compared with the kind of access provided, for instance, by a standard Web browser, voice assistants retain certain aspects while departing markedly from them in other regards. We aim therefore to address the following question: how does the Internet look through the prism of voice assistants? While this piece alone cannot answer this question in all its complexities, our goal is to highlight the importance of this problem and identify a range of issues that should spark additional reflections and discussions within communication and media studies and related fields. Such an endeavor builds on like-minded explorations advanced in different disciplinary contexts, such as computer science (Wilks, 2019: 119–133), but has until now attracted relatively little engagement within communication and media scholarship, despite growing interest of the field on the intersections between communication and AI (Gunkel, 2020; Guzman, 2017, Hepp, 2020).

After briefly discussing theories about interfaces and their implications, we focus on three key issues that we believe central to examine voice assistants in their quality of Web interfaces: information access, consumption, and affect. Each of these issues have to do with long-standing discussions about the role of digital media and their impact on the state of the public sphere (Dahlberg, 2001; Downey, 2014): in particular, the impact of web interfaces and platforms, including search engines and social media, on users' access to information; the role of web users as consumers and/or producers of contents and resources; and the affective flows that web platforms are programmed to facilitate in order to enhance users' connectivity and interactions. Our goal, in this regard, is twofold: to contextualize the examination of the case of voice assistants in existing debates within media and communication studies, but also to highlight how voice assistants pose novel questions that need to be urgently addressed, as these tools become more widespread and established.

Voice assistants as interfaces

Marshall McLuhan (1964) famously argued that electric light is the only medium without a message, unless it is used to spell out some verbal ad or name. Its "message" depends on how it was used, for example to illuminate an apartment or project images on

the cinematic screen or on TV. To some extent, the Internet in the digital age shares the same characteristic. Broadly defined as a network of computer networks, the Internet has endless potential but strictly undetermined form. Interfaces including email messaging systems, search engines, browsers, mobile media apps, and the Web itself have played and continue to play a key role in determining what the Internet means to users, and as a consequence, its effects and impact on politics and on society at large (Brügger, 2016). The question of how the Internet impacts on the public sphere, therefore, can only be addressed in relationship with the different hardware and software technologies that provide access to the internet and establish the conditions for its use in different contexts and at different moments of time.

Voice assistants represent one of the most recent instances of such hardware and software systems. From this point of view, they are Web interfaces. At a technical level, interfaces are broadly defined as points of interaction between any combination of hardware and software components. The word "interface," however, is more broadly employed to describe devices or programs that allow users to interact with computer resources (Hookway, 2014). While graphic interfaces draw on visual information to facilitate interaction, voice assistants are based on software processing voice inputs (Nass and Brave, 2005). Language-processing algorithms elaborate users' commands and questions, converting these inputs into executable commands. To function properly, voice assistants such as Alexa, Siri, and Google Assistant need a constant connection to the internet, where they retrieve information and access services and resources. Given that in order to process and comprehend speech these assistants make extensive use of cloud processing, where user interactions are processed in large data centers operated by the platform owners, an internet connection is required for all but their most minimal functionality (Hoy, 2018). With a connection, these systems perform functions such as searching the Web to respond to queries, providing information and news, playing music and other media, and managing communications including emails, phone calls, and messaging apps (Ammari et al., 2019).

As "mere" points of interaction, interfaces might be seen as secondary to the communication that occurs through them. Yet media scholars have often underlined that interfaces impact on the perception and the practical experience that people have of computing systems. Lori Emerson (2014) points out that the interface grants access but "also inevitably acts as a kind of magician's cape, continually revealing (mediatic layers, bits of information, etc.) through concealing and concealing as it reveals." Similarly, Wendy Hui Kyong Chun observes that interfaces 'offer us an imaginary relationship to our hardware: they do not represent transistors but rather desktops and recycling bins' (Chun, 2011: 66). Rather than being neutral, interfaces contribute powerfully to shape the experience of users. It is for this reason that voice assistants' quality as Web interfaces need to be taken very seriously.

Information access

The amount of information available through the Web is huge. It is estimated, as of May 2020, that there are more than 1.7 billion websites (data https://www.internetlivestats.com). Users make use of several kinds of interfaces to access such impressive mass of

information, such as browsers and search engines, which help them retrieve media and resources online. A potential way to see this process is that each of these interfaces reduces the breadth of information potentially available, so that the user's focus is limited to a more manageable range of information. The resources thereby located are supposed to respond to the queries and the needs of the user (Scott and Neil, 2009). Like all interfaces, however, Web interfaces are not neutral; they have significant biases that affect information access. For example, search engines rank their results based on algorithms that elaborate factors such as location, language and previous searches; moreover, they index only parts of the Web (Ballatore, 2015). Similarly, social networks impact on users' access to information through algorithms that rank different posts (Bozdag, 2013; Bucher, 2018).

Although each of these interfaces help users retrieve information they need, this also corresponds to a loss of control from the part of users. Since the rise of the Web, researchers have investigated to what extent different tools for Web navigation facilitate or hinder access to a plurality of information. Kjerstin Thorson and Chris Wells (2016) have proposed a framework for mapping information exposure of digitally situated individuals.

They argue that the conditions that define one individual's information access are defined by a proliferation of different contingencies. There is, in fact, not one but a plurality of content "flows" that are shaped and informed by factors including individual interests but also the infrastructures and platforms of digital communication: "unlike the mass media era, in which communication could be conceptualized as largely controlled by political elites and media actors, in the digital information environment processes of curation are also undertaken by actors such as friends and social contacts, computer algorithms, and individual media users themselves" (Thorson and Wells, 2016: 310). Interrogating the prerogatives of digital interfaces is essential to understand the directions and affordances of these content flows. This is why the question of voice assistants' impact on users' access to the Web resources is an urgent problem that needs to be tackled by communication and media scholars, developers and interface designers.

Our contention in this regard is that voice assistants give the impression of access to vast, nearly unlimited information that users customarily attributed to the Web, but dramatically reduce control over Web access to users, affecting their capacity to browse, explore and retrieve the plurality of information available through the Web. Comparing the search engine Google and the voice assistant Google Assistant helps clarify this point. If users search one keyword in the search engine, they are pointed to a plurality of sources. Although users rarely move beyond the first page of a search engine's results (Goldman, 2008), the interface still enables browsing into a large number of results. The same input directed to Google Assistant gives access to a much more limited amount of sources – in many cases, just one result provided through Wikipedia.

Due to the time that providing several potential answers by voice would need, compared with graphic and textual interfaces, the restriction of options is to be considered not just a design choice but an affordance of voice assistants as medium (Nass and Brave, 2005). Such restriction, however, is not transparent to the users of voice assistants. Despite the high degree of selectivity and the comparatively narrow range of information retrieved, these systems present themselves as neutral and entirely at the

service of the user. Google Assistant is again, in this regard, an appropriate example. As authors such as Siva Vaidhyanathan (2011) and John Durham Peters (2015) have compellingly shown, the Google corporation presents itself as an almost godlike entity, conflating itself with the omniscience attributed to the World Wide Web. Google Assistant, marketed as "Your own personal Google," takes up this representation by offering itself as an all-knowing entity that promises to have an answer for everything and to be "ready to help, wherever you are" (Google, 2019); one which you literally address as "Google" as if talking to the corporation in totality. Research has shown that the degree to which Google is trusted by users may affect their level of trust regarding results delivered through its search engine (Pan et al., 2007). Correspondingly, the level of trust assigned to the company may translate to the trust for the information provided by Google Assistant. Behind the corporate image and its benevolent omniscience, however, the control of the kind of information remains solidly in the hands of the company. This contrasts with the degree of relevance and the lack of information about its provenance (Dambanemuya and Diakopoulos, 2020).

By re-presenting information as the utterance of an assistant, moreover, its source may be obfuscated and labor erased (Hill, 2020). For instance, Wikipedia is used by voice assistants as a knowledge source to answer user queries. Wikipedia is run by Wikimedia, an independent non-profit organization, and its articles are written by volunteers (Loveland and Reagle, 2013). Not only does re-presenting information sourced from Wikipedia through the personality of an assistant erase the work done by Wikipedia's volunteer editors, it also appropriates that knowledge towards reinforcing the perception of that assistant (and its parent brand) as omniscient.

One further aspect is related to the question of voice assistants' agency in retrieving and using web resources. Like other kinds of computer software, digital interfaces are programmed to do things: they are not just abstractions that help users interact, they also operate changes in computing systems with corresponding effects in the real world. To mediate our interaction with the Web, voice assistants also incorporate software agents, that is autonomous components performing a variety of tasks related to web navigation, such as routing, searching, filtering, and caching. Every web browser or internet-connected device already makes use of software agents in one form or another (McKelvey, 2018). However, due to voice assistants' reliance on spoken words, voice assistants are expanding the breadth and depth of automatization to manage and enhance interactions with Web-based information on behalf of the user.

AI scientist Yorick Wilks (2019) has speculated that this might eventually result in an increasing complexity required to access information on the Web, since websites and web-based resources would be organized in order to privilege interactions with machines and algorithms, rather than direct contact with human users. He points out that, as interfaces such as voice assistants enhance automatization of web access, "the web may become unusable for non-experts unless we have Companion-like agents to manage its complexity for us" (p. 120). After all, research has shown that most users are not particularly good at doing Web searches, and that even if they use search engines very often, they rarely implement filters and search strategies adopted by specialists (Singer et al., 2012). Interfaces that autonomously conduct searches and deliver results, such as AI assistants, can easily become appealing to many users. As confirmed by a recent study (Ammari et al., 2019), web

searches are already one of the three most popular operations requested by users of voice assistants, beside playing music and activating Internet of Things devices.

There are pitfalls associated with this level of mediation between the web and the user. As we have already discussed, voice assistants make choices on behalf of users, filtering potentially large search results down to a limited range of entries. This simplification removes agency from a user. Writing about Netflix's analogous recommendation system, Sarah Arnold (2016) states "[h]uman agency, here, is posited as an encumbrance, something best surrendered so that the user is not overwhelmed with uncertainty." She calls for the need to exercise critical scrutiny to how much of our agency we cede to algorithmic systems. Adam Greenfield (2017) warns us that the choices offered by voice assistants "arrive prefiltered through existing assumptions about what is normal, what is valuable, and what is appropriate." A software agent is not a neutral device; it has been created by designers and programmers who encode these assumptions into their products. Voice assistants' programming, moreover, benefit from the large masses of data that are gathered about users' previous behaviors, which allows designers and companies to anticipate common queries and to design user experiences accordingly (Fisher and Mehozay, 2019).

Finally in terms of information access, the rising relevance of voice assistants might create new forms of inequality or enhance existing inequalities. One first complication in this regard is language. Because the availability of large amounts of data on users is crucial to the development of effective voice processing systems that automatically transcribe voice inputs, it is more difficult to develop voice assistants for languages spoken by minority groups or in smaller communities. The functionality of voice assistants might thus create new forms of divides, privileging English speakers and other larger language groups. Moreover, these systems are usually less effective in transcribing queries when users speak a language that is not their mother tongue, which may inform access for migrant communities within specific linguistic contexts and reinforce existing bias (Moussalli and Cardoso, 2019). The second complication is related to disability. If on the one side voice-controlled interfaces may provide new opportunities, for instance, to people with visual impairments (Vtyurina et al., 2019), they might also work in the opposite direction for disabilities that concern hearing and speech. The availability of access from a plurality of interfaces beside voice will be crucial to minimize such risk.

The impact of voice assistants on information retrieval and access in the Web is not always considered negatively. Emily MacArthur, for instance, pointed out that assistants such as Alexa and Siri restore "a sense of authenticity" to web searches, as they shape them as a conversation rather than just an interaction with computers (MacArthur, 2014: 117). While more research may help ascertain the potentially problematic as well as the productive aspects of such conversations, what appears sure is that their impact on access to the Web cannot be taken lightly. Further increase in the use of interfaces based on speech might change dramatically how information is accessed, and consequently, how and which information is produced, shared, and prepared for use on the Web. Tools such as Google Assistant, Alexa, and Siri, in other words, may impact not only on perceptions about the boundaries between humans and machines (Sweeney, 2020), but perhaps even more importantly, on how the entryways as well as the barriers to information will be shaped in the future.

Consumption

The complex relationship between production and consumption has been a constant element of debate in the study of media industries and audiences. In the enthusiastic climate of the Web's early years and even more pronouncedly with the emergence of Web 2.0, some viewed digital media as naturally facilitating participation from the audience (Thorburn and Jenkins, 2003). In this context, online media were heralded as promoting participatory culture. The activity of bloggers, YouTubers, and the like seemed to justify the emergence of a new notion, prosumption, which points to the hypothesis that the traditional divide between production and consumption is under challenge in the digital age (Beer and Burrows, 2010). A wealth of research, however, demonstrates that the relationship of power between corporations and users is still much unbalanced in the age of the Web. José Van Dijck (2009), for instance, influentially showed that behind platforms such as YouTube lie a rigid power structure by which only a small number of users produce actual content, and even the participation of such users is embedded in vertical labor relations administered by the company owning that platform. More recent discussions have tended to abandon the rhetoric of user participation and empowerment, highlighting the impact of surveillance capitalism and the overwhelming power of giant digital corporations (Moore and Tambini, 2018; Nieborg and Helmond, 2019).

From the point of view of user interaction, voice assistants may further depart from the alleged logics of participatory online culture. Creating the impression of a dialog between the user and the interface, voice assistants shift the interactive element from the public to the private sphere, from discussions in public fora to private conversations with the software that is embodied in the assistant. Users are encouraged to perceive the platform as responding and adapting to their queries, while in fact their ability to produce content that is visible to the public (such as social media posts or comments) is limited by the very affordances of the interface. Voice assistants are still "prosumers," but in a different sense: what they "produce" is data about their interactions, which is monetized and employed by Amazon, Apple, or Google to improve the handling of conversations and therefore, the illusion that voice assistants are capable of meaningful social exchanges (Wünderlich and Paluch, 2018).

Voice assistants, in fact, are also surveillance systems programmed to collect and analyze data about users' queries (Woods, 2018). Thanks to this data, developers of voice assistants are able to anticipate common queries and assign the task of drafting appropriate answers to professional writers (Young, 2019). Since this process remains opaque to most users, the fact that voice assistants are able to address apparently random questions creates the impression that they are able to anticipate the users' queries – meeting expectations of what a "personal" assistant is for. Users are thereby encouraged to overestimate voice assistants' capacity of autonomous behavior. As Margaret Boden observes, voice assistants seem "to be sensitive not only to topical relevance, but to personal relevance as well," striking users as "superficially impressive" (Boden, 2016: 65).

In fact, many of voice assistants' interactions that appear socially meaningful to users are scripted. In contrast, the conversational proficiency of these tools remains relatively limited (Natale, 2020). Differently from artificial companions and social robots, which are created with the explicit purpose of constructing the illusion of sociality (Caudwell

and Lacey, 2019), Siri, Alexa, and Google Assistant are designed around a transactional, command-response model of conversation. Therefore, they mainly treat conversational inputs as prompts to execute appropriate tasks. Thus, if voice assistants are an example of a computer interface that takes up the language of humans, they are also an interface that stimulates humans to take up the "language" of computers, that is programming language. An expert user of voice assistants will learn the commands that are most effective in order to have voice assistants operate as they wishes – similar to how a programmer will memorize the common commands of a programming language, or how an expert "power user" learns to effectively use Google search syntax.

While software developers have control on the operations performed by the system, the agency of the users of voice assistants can be best described as the ability to choose among a pre-defined range of interactions that the companies already anticipated for their systems. Developer documentation for Google Assistant describes possible use cases for the platform as "booking a flight," "finding out when their favorite sports team plays next," or "receiving coaching during yoga" (Google, 2020). Amazon's equivalent documentation for Alexa lists "Look up information from a web service," "order a car from Uber, order a pizza from Domino's Pizza," or "Turn lights on and off" (Amazon, 2020a). These are short interactions with a functional goal, which cast the user in a consumer role and as a mainly passive receiver of information. Amazon's Alexa Design Guide (Amazon, 2020b) refers to the user as a "customer" throughout: a clear indication of their intended role.

Also software developers working on these platforms are limited to defining expected commands ("intents") and returning text (to be read by a speech synthesizer) or short audio clips in response; no other interactions are possible within the software frameworks supplied by the platform owners. Design research carried out by one of the authors explored the limitations of this interaction model. Musicians and sound artists were invited to a workshop in which they worked with technologists from BBC R&D to explore the possibilities of Alexa as a creative tool. Participants found it very hard to create interactions outside of the command/response model, and only one developer managed to create software which allowed a user any meaningful creative agency (Cowlishaw, 2018).

In the future, the technologies that power the most advanced voice assistants could become available to independent companies, non-profit organizations, public institutions, and even individual programmers, opening up for systems that do not reproduce the bias and the interests of big corporations.[1] This potential development, however, has already been incorporated into the business model of a company such as Amazon, which provides a range of developers' functions available for customization by private companies and programmers through tools such as the Alexa Skills Kit (ASK) or Alexa Skill Blueprints. These include personalization options such as adding voice control to connected products, building conversational programs, customize speech feature such as intonation, timing, and pronunciation (Gunkel, 2020: 149), and creating apps – "Skills" in Alexa's terminology – which ultimately, however, remain under the auspices and the control of the Amazon company (Lee, 2018). It is unlikely, considering the reliance on huge datasets of user information to train AI technologies in areas such as speech recognition and synthesis and natural language understanding, that cutting-edge systems will

be developed outside of the main actors of digital capitalism. The key technologies required for the development of these systems require scales of operation that are not achievable by smaller organizations or individual developers. Cutting-edge voice recognition, natural language understanding and speech synthesis are all machine learning problems that rely on vast datasets of spoken and written language for their training. Some of the best systems that currently exist to capture that data are voice assistants themselves; in a neat self-supporting turn, the systems gather the data needed to create better iterations of themselves. Crawford and Joler (2018) have remarked that users of Alexa are "a resource, as their voice commands are collected, analyzed and retained for the purposes of building an ever-larger corpus of human voices and instructions." They argue that users of voice assistants are a hybrid of consumer, resource, and laborer, unwittingly working to improve the platforms that they use.

Ultimately, voice assistants such as Siri, Alexa, and Google Assistant can only exist embedded within a wider system of material and algorithmic structures that guarantee market dominance to companies such as Amazon, Apple, and Google (Hill, 2020). They are gateways to the cloud-based resources administered by these companies. Voice assistants in this sense not only constitute an additional departure from participatory culture, which adds to the unbalanced power relationship between corporations and users. As interfaces giving access to the internet, they further erode the distinction between the web and the proprietary cloud services that are controlled by a few multi-billion company actors.

As Judith Donath (2019) has recently suggested for the case of social robots, voice assistants in the future might be exploited for advertising purposes, since the ability to market that an intelligent conversation agent would master would put them "in an entirely new category" (p. 21). Donath perceptively notes that advertising has been the key source of income for numerous "new" media of the past, from the penny press to broadcasting, from web search engines to social media. Arguably, voice assistants already retain a promotion function in directing customers to Amazon, Apple, and Google products and services. Explorative studies in marketing have already highlighted voice assistants' potential in building engagement and loyalty (Moriuchi, 2019; Vernuccio et al., 2020). The feelings of empathy that anthropomorphized software (as voice assistants can be considered, due to their use of humanlike voices, names, gender, and other characteristics, see Araujo, 2018) may stimulate in users will create novel and pressing ethical questions. The boundaries between persuasion and deception will have to be explored and redrawn (Natale, 2021; Schuetzler et al., 2019), similarly to how the problem of certain uses of social media such as Facebook in regard with political propaganda and disinformation are presently at the center of lively controversies (Chadwick and Vaccari, 2019).

Affect

As social media and social networks developed into one of the key platforms for users to access information, entertainment and sociality online, connectivity rose to the double status of an idealized outcome for Web 2.0 and a business model for companies such as Facebook (Fisher, 2018). Since the main source of income for these platforms is advertising, the degree to which they are able to encourage participation and constant

connectivity from the part of users impacts directly on their commercial potential. As shown by Karppi (2018), social media companies therefore see disengagement as an existential threat. This has led them to incorporate a range of features aimed at stimulating "affective bonds" (Karppi, 2018) that encourage users to interact extensively with social media and commercial web resources. The "like" feature, for instance, ensures that users receive a form of emotional reward that motivates them to continue using the platform, thereby improving the platform's appeal to advertisers. Similarly, other Web interfaces such as browsers and search engines are also designed in ways that facilitate affective reactions, as research has shown that a web user's initial emotional response have carried-on effects on their subsequent interactions (Deng and Poole, 2010).

The introduction of the voice assistant adds new layers of complexity to the managing of affect in Web interfaces. Constructing appearances of personality that stimulate specific responses from users (Guzman, 2015), voice assistants create forms of affect that are grounded not much or not only in interactions with other users as in interactions between the user and the machine. This enhances the bonds that stimulate users to engage with digital platforms (Bucher, 2018) through the building of an imagined relationship with the interface itself. This does not mean, to be clear, that users are led to believe that the assistant is a real person: for all the anthropomorphising features, studies have confirmed that users generally remain aware of the distinction between voice assistants and "real" people (Guzman, 2019). The characterization features embedded in this kind of software, however, can still have very meaningful consequences. Nass and Brave (2005) have shown that the personality that a machine adopts while presenting information to a user affects how the user feels about that information, and research has also shown that the style of interaction (e.g. social vs task oriented) informs the effectiveness of the retrieval of resources and functions from the Web (Chattaraman et al., 2019). Consequently, designers have proposed practical solutions to instil distinctive personality into software and hardware objects as a way to stimulate particular forms of engagements in users (Marenko and Van Allen, 2016) and have pointed to the impact of anthropomorphic features in orienting users' perception and affect (Araujo, 2018; Caudwell and Lacey, 2019).

The question then deserves to be asked, to what extent the flows of affect constructed by the specific characterization of each voice assistant influence how we feel about the information we access through it? In order to answer this question, a deeper understanding is needed of how voice assistants mobilize mechanisms of representation and stereotyping in their interactions and relationships with users (Natale, 2021). From a technical viewpoint, there isn't anything like one monolithic "Alexa" or "Siri." Rather than being individual entities, they are the integration of a wide range of different systems and algorithms. "Alexa" and "Siri" do exist, however, at a semiotic level. Similarly to how graphic interfaces employ metaphors such as the desktop or the bin to help users acclimatize with the system (Emerson, 2014), voice assistants employ several semiotic means to facilitate and direct users' interactions.

The first of these means is the name given to the software. This replicates a strategy employed also for other types of software, which are sometimes assigned names that suggest humanity and also gender (Zdenek, 2003). As Taina Bucher has shown in regard with the case of social media bots, such feature contributes to constructing the illusion of a "persona," that is a continuing relationship with an individual entity (Bucher, 2014).

Such a token of individuality conceals the fact that behind what we call "Alexa" or "Siri" lie a plurality of discrete software systems, practices, and platforms (Natale, 2020).

The second means of representation that is mobilized by voice assistants is the use of a humanlike, gendered voice. Although the possibility was open for developers of voice assistants to implement a machine-like and/or a gender neutral voice, all main actors have opted for humanlike and gendered voices – quite often, a female one, a circumstance that awakens much concern regarding the replication and proliferation of gender stereotypes through these platforms (Sweeney, 2020; Woods, 2018; see also Zdenek, 1999). Asked about the reasons for this choice, Deborah Harrison – one of the personality designers of Microsoft's voice assistant Cortana (a project that was recently discontinued after running between 2014 and 2019) – explained that in her team's intention "the female voice was just about specificity. In the early stages of trying to wrap our mind around the concept of what it is to communicate with a computer, these moments of specificity help give people something to acclimate to" (Young, 2019: 117). Harrison's answer suggests that a key rationale for characterizing the assistant's voice through gender is the necessity to capitalize from the existing familiarity of users with human voices. As authors such as Lippmann (1922) and Bourdieu (1977) have convincingly shown, people need to rely on existing representations, stereotypes and habits in order to accommodate new information and experiences. The play of stereotyping encouraged by a gendered voice may thus conduct, in the companies' intention, to a better integration of voice assistants within users' existing experiences and interpretive frameworks; in Deborah Harrison's own words, it may "give people something to acclimate to" (Young, 2019: 117).

Stereotypical representations mobilized by the assistants' voice do not concern only gender, but also other elements of characterization. As argued by Thao Phan (2017) and Heather Woods (2018), for instance, Alexa directs users towards a representation of a character that has not only a clear gender but even hints at class and race, as the voice is suggestive of a native-speaking, educated, white woman. Humphry and Chesher (2020) point out that this is meant "to reduce anxieties about their potential to exceed their roles as loyal helpers and cross the boundary into the monstrous," (p. 2) that is the negatively charged connotation of robotic and alien voices that has been firmly established by science fiction and popular culture's representation of hostile robots. Such an approach, moreover, assuages concerns and resistance from consumers "to the surveillant potential of these technologies, while keying into more positive associations for robots and the automated home of the future" (Humphry and Chesher, 2020: 3).

The third means of representation mobilized by voice assistants is the simulation of a personality. While much of the characterization is left to the imagination and interpretation of users (Guzman, 2015), each assistant is instilled to some degree of its own specific character. Interestingly, there are significant points of connection between the personality staged by the main voice assistants and the business model of the companies producing them. Amazon represents Alexa as a docile servant (Woods, 2018). This helps conceal Amazon's structures of labor, whereas the exploitation of the workforce needs to remain invisible to users who access Amazon services and products online (Hill, 2020). In comparison to Alexa, Siri's digital persona makes more use of irony and storytelling (Thorne, 2020). This fits with Apple's corporate image, through which the company has

aimed to match their target customers' self-representation based on creativity and unique-ness (Magaudda, 2015). Finally, Google decided to give their assistant less evident tokens of personality (Greenfield, 2017). This choice reflects the wider communication strategy of a company that traditionally downplays individual characters to present Google as an immanent entity (see Natale et al., 2019).

The extent to which the personalization of voice assistants might inform how infor-mation is accessed and interpreted by users needs to be the subject of research not only from the perspective of computer science and interaction design, but also from a social sciences viewpoint. One crucial question is the extent to which trust in the information provided can be enhanced by manipulating the representation and characterization of the voice assistants. This issue may link to existing discussions about social media's alleged "elective affinity" (Gerbaudo, 2018) with phenomena such as disinformation, fake news, and populism. If voice assistants' role as Web interfaces will further expand, in fact, the question might arise if voice assistants can be conducive of similar or different chal-lenges as those observed for social media. Another important issue is related to the prob-lem of persuasion. As discussed briefly in the previous section, voice assistants might potentially be used to influence users to consume specific products or to vote for a spe-cific party or political figure. In this context, elements of characterization that produce affect and activate mechanisms of empathy might enhance the persuasiveness of these tools. As suggested by initial explorations of this issue, persuasion might be achieved by voice assistants through means that are usually confined to interpersonal communica-tion, including flirting (Tyutelova et al., 2020). The application of these mechanisms in sensitive areas such as political communication or in the spread of misinformation can-not be excluded. In fact, chatbots employing natural language processing software – similar to that which is employed in voice assistants – to generate the illusion of social engagement, are already being used for political propaganda: in the 2019 Israeli election, for instance, a chatbot impersonating Prime Minister Benjamin Netanyahu was used in its official campaign to target supporters (Ben-David, 2019).

Conclusion

Although voice assistants have until now scarcely been examined in their quality of web interfaces, these and other technologies that use speech as a way to access information and resources on the Web are likely to become more and more important in the near future. Considering the swift success of voice assistants, AI researcher Yorick Wilks (2019) recently suggested that "it may become impossible to conceive of the web with-out some kind of a human face that renders it personal" (p. 121). Although it is always difficult to predict future technological trends, it is a fact that the retrieval of information from the web is already one of the key functions of voice assistants. It is therefore urgent that media and communication scholars, as well as computer scientists, designers and policy-makers, consider more programmatically their potential impact in this regard. By advancing an initial agenda in this direction, this article has shown that some key ongo-ing debates about online media – such as the role of web platforms in shaping informa-tion access, the relationship between production and consumption online, and the role of

affect in informing engagement with web resources – may help in assessing and understanding the implications of voice assistants as web interfaces.

Our analysis has focused on voice assistants developed by Western and more specifically US-based corporations, such as Amazon, Apple, and Google. While Siri, Alexa, and Google Assistant are made available in a large number of different languages and have achieved significant global reach (Hoy, 2018), competing products have also been developed in different parts of the world (Xiao and Kim, 2018). More research, in this sense, is needed that considers the implications of regional, national, cultural, and linguistic specificities informing how these tools provide Web access to different communities of users.

If the potential challenges and problems have been the subject of particular attention in this article, voice assistants also bear significant potential as web interfaces. A more optimistic way to look at the issue may be to consider that the growing role of web agencies embedded in voice assistants could result in more accurate systems to assess information, distinguishing between reliable and unreliable information, and counteracting the dissemination of fake news. Indeed, the role of voice assistants as interfaces to the Internet is not written in any of their technical features, especially considering that a technology and its impact can never be separated from the social, political, economic and institutional structures that surround them (Williams, 1974). Rather than rejecting or stigmatizing these tools, media scholars, legislators, and designers face the challenge of identifying and developing the most appropriate technical solutions, ethical principles, and governance tools so that change can be anticipated and directed toward fruitful and sustainable ways.

This article calls for media and Internet researchers to take up this challenge, conducting further research and engaging with policymakers and developers in order to counteract potential risks and complexities. The article originated from a collaboration between academia and the industry: one of the authors is a media studies scholar while the other is a software developer in the media sector with a background in digital storytelling. Considering how social media have shifted in the lapse of just a few years from being widely regarded as harbingers of new opportunities for citizen journalism (Wall, 2015) to being criticized by many as leading to the populism, disinformation, and fake news (Chadwick and Vaccari, 2019), it is especially important that the impact of voice assistants in accessing information on the web is investigated in a timely way. This will require the diverse competences and concerted efforts of researchers, developers and designers who are ready to question existing assumptions of what is the Web and how it may look like as speech, beside graphic and textual interfaces, becomes one of the key ways to access the resources and information available through it.

Funding

ORCID iD

Simone Natale (iD) https://orcid.org/0000-0003-1962-2398

Note

1.   A few open source projects of this kind already exist, including Rhasspy, Mycroft, and Mozilla DeepSpeech.

References

Amazon (2020a) Build skills with the Alexa Skills Kit. Available at: https://developer.amazon. com/en-US/docs/alexa/ask-overviews/build-skills-with-the-alexa-skills-kit.html   (accessed 16 November 2020).

Amazon (2020b) Alexa Design Guide. Available at: https://developer.amazon.com/en-US/docs/ alexa/alexa-design/get-started.html (accessed 16 November 2020).

Ammari T, Kaye J, Tsai JY, et al. (2019) Music, search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26(3): 1–27.

Araujo T (2018) Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior* 85: 183–189.

Arnold S (2016) Netflix and the myth of choice/participation/autonomy. In: McDonald K and Smith-Rowsey D (eds) *The Netflix effect: Technology and entertainment in the 21st century*. New York: Bloomsbury, pp. 49–62.

Ballatore A (2015) Google chemtrails: A methodology to analyze topic representation in search engine results. *First Monday* 20(7). Avalaible at: https://firstmonday.org/ojs/index.php/fm/ article/view/5597/4652 (accessed 15 July 2020).

Beer D and Burrows R (2010) Consumption, prosumption and participatory web cultures: An introduction. *Journal of Consumer Culture* 10(1): 3–12.

Ben-David A (2019) How we got Facebook to suspend Netanyahu's chatbot (and why it matters). Available at: https://www.anatbendavid.info/post/2019/10/10/how-we-got-facebook-to-suspend-netanyahu-s-chatbot-and-why-it-matters (accessed 9 June 2020).

Boden M (2016) *AI: Its Nature and Future*. Oxford: Oxford University Press.

Bourdieu P (1977) *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.

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

Brügger N (2016) Introduction: The Web's first 25 years. *New Media & Society* 18(7): 1059–1065.

Bucher T (2014) About a bot: Hoax, fake, performance art. *M/C Journal* 17(3). Available at: http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/814 (accessed 15 July 2020).

Bucher T (2018) *If. . . Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.

Caudwell C and Lacey C (2019) What do home robots want? The ambivalent power of cuteness in robotic relationships. *Convergence* 41(8): 1176–1191.

Chadwick A and Vaccari C (2019) *News Sharing on UK Social Media: Misinformation, Disinformation, and Correction*. Loughborough: Online Civic Culture Centre, Loughborough University.

Chattaraman V, Kwon WS, Gilbert JE, et al. (2019) Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior* 90: 315–330.

Chun WHK (2011) *Programmed Visions: Software and Memory*. Cambridge, MA: MIT Press.

Cowlishaw T (2018) IRFS Weeknotes #265. Available at: https://www.bbc.co.uk/rd/blog/2018-04-irfs-weeknotes-number-265 (accessed 15 November 2020).

Crawford K and Joler V (2018) Anatomy of an AI System. Available at: https://anatomyof.ai/ (accessed 20 September 2019).

Dahlberg L (2001) Computer-mediated communication and the public sphere: A critical analysis. *Journal of Computer-Mediated Communication* 7(1). Available at: https://academic.oup.com/jcmc/article/7/1/JCMC714/4584246 (accessed 13 December 2020).

Dambanemuya HK and Diakopoulos N (2020) "Alexa, what is going on with the impeachment?" Evaluating smart speakers for news quality. *Proceedings of the Computation and Journalism Symposium* 1: 1–4.

Deng L and Poole M (2010) Affect in web interfaces: A study of the impacts of web page visual complexity and order. *MIS Quarterly* 34(4): 711–730.

Donath J (2019) The robot dog fetches for whom? In: Papacharissi Z (ed.) *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*. New York, NY: Routledge, pp.10–24.

Downey J (2014) Flux and the public sphere. *Media, Culture & Society* 36(3): 367–379.

Emerson L (2014) *Reading Writing Interfaces: From the Digital to the Book Bound*. Minneapolis: University of Minnesota Press.

Fisher E (2018) When information wanted to be free: Discursive bifurcation of information and the origins of Web 2.0. *Information Society* 34(1): 40–48.

Fisher E and Mehozay Y (2019) How algorithms see their audience: Media epistemes and the changing conception of the individual. *Media, Culture & Society* 41(8): 1176–1191.

Gerbaudo P (2018) Social media and populism: An elective affinity? *Media, Culture & Society* 40(5): 745–753.

Goldman E (2008) Search engine bias and the demise of search engine utopianism. In: Spink A and Zimmer M (eds) *Web Search: Multidisciplinary Perspectives*. Berlin: Springer, pp.121–133.

Google (2019) Google Assistant. Available at: https://assistant.google.com/ (accessed 19 September 2019).

Google (2020) Conversational Actions. Available at: https://developers.google.com/assistant/conversational/overview (accessed 16 November 2020).

Greenfield A (2017) *Radical Technologies: The Design of Everyday Life*. New York: Verso.

Gunkel DJ (2020) *An Introduction to Communication and Artificial Intelligence*. Cambridge, UK: Polity.

Guzman AL (2015) *Imagining the voice in the machine: The ontology of digital social agents*. PhD Dissertation, University of Illinois, Chicago.

Guzman AL (2017) Making AI safe for humans: A conversation with Siri. In: Gehl RW and Bakardjieva M (eds) *Socialbots and Their Friends: Digital Media and the Automation of Sociality*. London: Routledge, pp.85–101.

Guzman AL (2019) Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior* 90: 343–350.

Hepp A (2020) Artificial companions, social bots and work bots: Communicative robots as research objects of media and communication studies. *Media, Culture & Society* 42(7–8): 1410–1426.

Hill DW (2020) The injuries of platform logistics. *Media, Culture & Society* 42(4): 521–536.

Hookway B (2014) *Interface*. Cambridge, MA: MIT Press.

Hoy MB (2018) Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly* 37(1): 81–88.

Humphry J and Chesher C (2020) Preparing for smart voice assistants: Cultural histories and media innovations. *New Media and Society*. Epub ahead of print 22 May 2020. DOI: 10.1177/1461444820923679.

Karppi T (2018) *Disconnect: Facebook's Affective Bonds*. Minneapolis: University of Minnesota Press.

Lee H (2018) *Voice User Interface Projects: Build Voice-Enabled Applications Using Dialogflow for Google Home and Alexa Skills Kit for Amazon Echo*. Birmingham: Packt Publishing Ltd.

Lippmann W (1922) *Public Opinion*. New York: Harcourt, Brace & Co.

López G, Quesada L and Guerrero LA (2018) Alexa vs. Siri vs. Cortana vs. Google Assistant: A comparison of speech-based natural user interfaces BT. In: Nunes IL (ed.) *Advances in Human Factors and Systems Interaction*. Cham: Springer, pp.241–250.

Loveland J and Reagle J (2013) Wikipedia and encyclopedic production. *New Media & Society* 15(8): 1294–1311.

MacArthur E (2014) The iPhone Erfahrung: Siri, the auditory unconscious, and Walter Benjamin's Aura. In: Weiss DM, Propen AD and Emmerson Reid C (eds) *Design, Mediation, and the Posthuman*. Lanham: Lexington Books, pp.113–127.

Magaudda P (2015) Apple's iconicity: Digital society, consumer culture and the iconic power of technology. *Sociologica* 9(1): 1–15.

Marenko B and Van Allen P (2016) Animistic design: How to reimagine digital interaction between the human and the nonhuman. *Digital Creativity* 27(1): 52–70.

McLuhan M (1964) *Understanding Media: The Extensions of Man*. Toronto: McGraw-Hill.

McKelvey F (2018) *Internet Daemons: Digital Communications Possessed*. Minneapolis: University of Minnesota Press.

Moore M and Tambini D (2018) *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Oxford: Oxford University Press.

Moriuchi E (2019) Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing* 36(5): 489–501.

Moussalli S and Cardoso W (2019) Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*. Epub ahead of print 15 April 2019. DOI: 10.1080/09588221.2019.1595664.

Nass C and Brave S (2005) *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA: MIT Press.

Natale S (2020) To believe in Siri: A critical analysis of AI voice assistants. *Communicative Figurations Working Papers* 32: 1–17.

Natale S (2021) *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. New York: Oxford University Press.

Natale S, Bory P and Balbi G (2019) The rise of corporational determinism: Digital media corporations and narratives of media change. *Critical Studies in Media Communication* 36(4): 323–338.

Nieborg DB and Helmond A (2019) The political economy of Facebook's platformization in the mobile ecosystem: Facebook Messenger as a platform instance. *Media, Culture & Society* 41(2): 196–218.

Pan B, Hembrooke H, Joachims T, et al. (2007) In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication* 12(3): 801–823.

Papacharissi Z (2002) The virtual sphere: The internet as a public sphere. *New Media & Society* 4(1): 9–27.

Peters JD (2015) *The Marvelous Cloud: Towards a Philosophy of Elemental Media*. Chicago, IL: University of Chicago Press.

Phan T (2017) The materiality of the digital and the gendered voice of Siri. *Transformations* 29: 23–33.

Schiller A and McMahon J (2019) Alexa, alert me when the revolution comes: Gender, affect, and labor in the age of home-based artificial intelligence. *New Political Science* 41(2): 173–191.

Schuetzler RM, Grimes GM and Giboney JS (2019) The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior* 97: 250–259.

Scott B and Neil T (2009) *Designing Web Interfaces: Principles and Patterns for Rich Interactions*. Sebastopoli, CA: O'Reilly.

Singer G, Norbisrath U and Lewandowski D (2012) Ordinary search engine users carrying out complex search tasks. *Journal of Information Science* 39: 346–358.

Sweeney ME (2020) Digital assistants. In: Thylstrup NB, Agostinho D, Ring A, et al. (eds) *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, MA: MIT Press. Available at: https://ir.ua.edu/handle/123456789/6348 (accessed 7 November 2020).

Thorburn D and Jenkins H (2003) *Rethinking Media Change: The Aesthetics of Transition*. Cambridge: MIT Press.

Thorne S (2020) Hey Siri, tell me a story: Digital storytelling and AI authorship. *Convergence* 26(4): 808–823. DOI: 10.1177/1354856520913866.

Thorson K and Wells C (2016) Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory* 26(3): 309–328.

Tyutelova LG, Lisovitskaya VN, Shevchenko ES, et al. (2020) The voice-activated digital assistants in the brand-consumer communication. In: *2020 IEEE communication strategies in digital society seminar* (ComSDS), 8 April 2020, pp.137–140. New York: IEEE.

Vaidhyanathan S (2011) *The Googlization of Everything (and Why we Should Worry)*. Berkeley: University of California Press.

Van Dijck J (2009) Users like you? Theorizing agency in user-generated content. *Media, Culture & Society* 31(1): 41–58.

Vernuccio M, Patrizi M and Pastore A (2020) Brand anthropomorphism and brand voice: The role of the name-brand voice assistant. In: Martínez-López F and D'Alessandro S (eds) *Advances in Digital Marketing and eCommerce*. Cham, Switzerland: Springer, pp.31–39.

Vtyurina A, Fourney A, Morris MR, et al. (2019) Bridging screen readers and voice assistants for enhanced eyes-free web search. In: *The world wide web conference*, San Francisco, CA, May 2019, pp.3590–3594. New York, NY: Association for Computing Machinery.

Wall M (2015) Citizen journalism: A retrospective on what we know, an agenda for what we don't. *Digital Journalism* 3(6): 797–813.

Wilks Y (2019) *Artificial Intelligence: Modern Magic or Dangerous Future?* London: Icon Books.

Williams R (1974) *Television: Technology and Cultural Form*. London: Fontana.

Woods HS (2018) Asking more of Siri and Alexa: Feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication* 35(4): 334–349.

Wünderlich NV and Paluch S (2018) A nice and friendly chat with a bot: User perceptions of AI-based service agents. In: *ICIS 2017: Transforming Society with Digital Innovation*, pp.1–11. Available at: https://aisel.aisnet.org/icis2017/ServiceScience/Presentations/11/ (accessed 13 December 2020).

Xiao XT and Kim SI (2018) A study on the user experience of smart speaker in China: Focused on Tmall Genie and Mi AI Speaker. *Journal of Digital Convergence* 16(10): 409–414.

Young L (2019) 'I'm a cloud of infinitesimal data computation' when machines talk back: An interview with Deborah Harrison, one of the personality designers of Microsoft's Cortana AI. *Architectural Design* 89(1): 112–117.

Zdenek S (1999) Rising up from the MUD: Inscribing gender in software design. *Discourse & Society* 10(3): 379–409.

Zdenek S (2003) Artificial intelligence as a discursive practice: The case of embodied software agent systems. *AI & Society* 17(3–4): 340–363.