Human Listeners and Virtual Assistants: Privacy and Labor Arbitrage in the Production of Smart Technologies

Paola Tubaro, Antonio A Casilli

10 Human Listeners and Virtual Assistants: Privacy and Labor Arbitrage in the Production of Smart Technologies

Paola Tubaro and Antonio A. Casilli

In spring 2019, public outcry followed media revelations that major producers of allegedly automated voice-activated devices recruit human operators to listen to, transcribe, and label users' conversations. These high-profile news stories originated in journalistic and scholarly investigations, and in the voluntary disclosures of whistleblowers. The scandal started with a *Bloomberg* report on April 10, 2019, disclosing that thousands of Amazon workers listen to the private conversations of users of the smart speaker Echo. Workers include both employees and subcontractors of Amazon, located in the US and in India. A company spokesperson claimed that these privacy violations concerned "an extremely small number" of the 100 million–strong Echo user base (Day, Turner, and Drozdiak 2019).

A few months later, the Flemish Radio and Television Broadcasting Organization (VRT) revealed that third-party outside contractors were eavesdropping on audio recordings of users of Google Home smart speakers and smartphone apps. The practice was not characterized as an exception; Google subcontractors were said to "systematically" listen after logging into a secure platform. Some of the conversations they accessed contained sensitive information, such as private arguments, confidential business conversations, and the whereabouts of children (Verheyden et al. 2019). Even more intimate details were said to reach Apple. According to a whistleblower who used to work for a subcontracting company, these included "confidential medical information, drug deals, and recordings of couples having sex." This data was collected via the company's voice assistant, Siri, which is included in most of its devices from iPhones to HomePod smart speakers and Apple Watches. As Amazon had done earlier, an Apple spokesperson claimed that this concerned "a very small random subset, less than one percent of daily Siri activations" (Hern 2019).

Following these revelations, which brought to the fore whistleblowers from within tech companies, other news stories followed at rapid pace. In August 2019, Microsoft and Facebook were exposed in the international press for subcontracting providers in

the Philippines, Bulgaria, and Mexico to listen to personal conversations on Skype (Cox 2019) and for recording and transcribing Messenger voice chats (Frier 2019). But even following this steady drip of news reports, tech companies have yet to demonstrate that they have built the necessary safeguards into their devices and software. Since 2019, very few of these companies have paused their eavesdropping programs or enhanced their privacy controls. The most common corporate responses have been limited to tweaking privacy policies and/or internalizing these processes, entrusting them to on-site employees who are submitted to more stringent nondisclosure agreements (NDAs) (Carr et al. 2019).

News reports of these egregious cases of privacy erosion display a remarkable common feature: they traditionally adopt the viewpoint of consumers worried about intrusions in their daily lives and intimate spheres, and about potential ensuing security issues. There is a tension between these recent privacy concerns and companies' need to collect the data necessary to the operation of the artificial intelligence (AI) systems that run smart speakers, smartphones, and mobile applications. Workers who remotely listen to consumers' conversations, called "data associates," "raters," or "reviewers," provide indispensable quality controls to improve automated processes such as voice-activated virtual-assistant software, machine translation, and speech-recognition systems. While the terms of use of most devices do inform users that personal data is used to train and test AI systems, they do not always explicitly state that human teams listen to recordings. Even when they do, they minimize the impact of their intervention, as, for example, Amazon for which it is "an industry-standard practice where humans review an extremely small sample of requests to help Alexa understand the correct interpretation of a request and provide the appropriate response in the future" (Amazon n.d.). Importantly, by characterizing these activities as AI training, to improve the tools, companies reinforce the belief—which, as we will see, is rather a misconception—that privacy violation is limited to a transitory period and that it will stop as soon as full-fledged machine learning kicks in. Moreover, the attempt to reassure consumers that this process "includes multiple safeguards to protect customer privacy" (Amazon n.d.) rarely provides further details.

In this chapter, we show that the job descriptions of these third-party workers include much more than just transcribing and annotating conversations to help automated speech algorithms "self-learn." These workers often verify the results of the software's calculations, fix errors, and compare automated transcriptions to human-made ones. In some instances, workers' tasks even consist of directly executing speech commands that the AI systems are unable to interpret, thus "impersonating" virtual assistants by completing the very same tasks they are supposed to perform automatically. These multitudes of contributors work in the shadows because their very existence—as

humans who listen to other humans—is embarrassing for the companies that sell allegedly automated voice-activated solutions. Indeed, admitting their role in developing smart technologies would be at odds with the common marketing claim that these technologies can become deeply integrated into our lives, precisely because they are activated by simple voice command (see, for example, Martinez and Cameron n.d.). Advertising campaigns have created the expectation that smart technologies effortlessly interpret and act upon these commands. If the recent scandals plainly exposed the unrealistic nature of these claims, they have not stopped the recruitment of human "remote listeners." It turns out, as we will discuss in more detail, that their function remains crucial and cannot be automated. Ironically, voice assistants need humans but obfuscate their contribution.

We argue that privacy issues related to the development of voice assistants are not isolated offenses. Rather, they are intrinsically related to the labor-intensive nature of today's models of automation, which are based on machine learning and fueled by human-produced data. The production systems underpinning these models cross borders, procuring workers in (mostly) low-wage countries to listen to the voice recordings of consumers in (primarily) Western Europe, North America, and other higher-income parts of the world. Indeed, digital technologies enable workers' outputs to be immediately transferred to any place, reducing the need for physical proximity. Companies' efforts to minimize labor costs are limited only by the need to match the language competencies of consumers and workers. The result is a complex geography that largely reproduces linguistic proximities inherited from the colonial past.

The conditions of labor and the remuneration of these remote listeners lie at the heart of the problem. Even less-intrusive computing techniques that mitigate privacy issues on the consumer side—primarily in the Global North—do not eliminate the need for humans in the loop in the Global South. In the remainder of this chapter, we describe the extent and relevance of this human contribution, and the harsh conditions under which it is obtained—including low salaries, precarity, and lack of social protection. We conclude that privacy issues are only one side of a more complex, multifaceted problem that cannot be solved without also addressing the working conditions and remuneration of the people who toil behind the production of purportedly automated voice assistants.

To support our argument, we use interviews that we conducted in 2018–2019 with start-ups specializing in automation and particularly voice technologies, as well as with a small number of whistleblowers who had formerly worked as remote listeners for international platforms and for subcontractors of major multinational technology companies. To reach out to them, we engaged in a broad effort to publicize our study

through social media, inviting participation widely; we also used snowballing and leveraged personal contacts with digital rights associations. Moreover, we rely on an extensive, two-year observation of platform websites, press releases, publications, and other public documentation. To a lesser extent, we also use some data from a questionnaire that we distributed to over 900 online workers on the platform then called Foule Factory (now Wirk.io) in France in 2018 (Casilli et al. 2019).

The Real Humans behind "Automated" Voice Assistants

Voice-activated assistants were among the first products of AI to be widely marketed worldwide. From general-purpose assistants such as Amazon's Alexa, Apple's Siri, and Microsoft's Cortana to more specialized products for connected objects such as voice-activated coffee machines (and other home- or car-automation devices), their use has become ubiquitous in the last few years. The current capacities of these smart solutions are the result of major advances in natural language processing and speech-recognition research. Answering even a mundane question about the weather at a certain location or the recipe for a home-cooked dish requires realizing a sequence of operations, each of which can be challenging for a machine. Before an actual answer can be provided, a question uttered by the user needs to be transcribed, analyzed for semantic and syntactic features, and adjudicated against a database of facts and preanswered questions. The computational solution to these problems relies on machine learning, a technique that "teaches" computers to find solutions from data without having to rely on explicitly programmed rules at each step of the process. Machine learning is the technique that has fueled the current AI boom. Given sufficiently large sets of examples, machine learning models can detect regularities and patterns in the data, which they then use as a basis to make predictions or decisions. Better-quality and larger datasets allow progressive refinement of results. Put differently, data is as necessary an input to AI as the algorithms (or computer programs) that handle it.

To illustrate the functioning of voice assistants more concretely, let us go back to the example of a question about the weather. First, an algorithm must be "trained" with sample data, which in this case will be audio recordings of people asking for the weather. Large numbers of such recordings are needed so that the algorithm can learn that they all mean the same thing despite differences in timbre, type and tone of voice, regional accents, and background noise level. But initial training is only one step toward the production of a functioning AI solution, as subsequently the algorithm needs to be tested and may undergo other phases of training to improve its performance. Some of these steps may include direct human interventions, whereby an operator takes over to

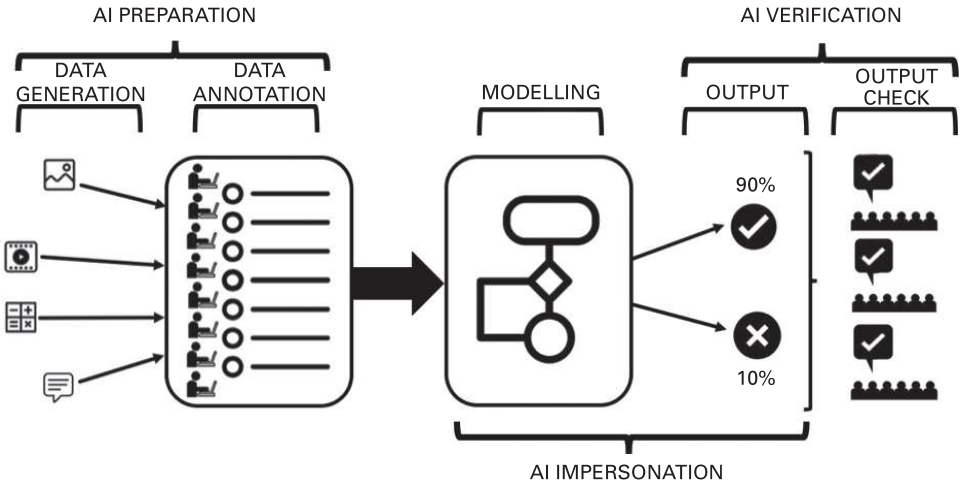

Figure 10.1
Platform-mediated human labor in three phases of the development of machine learning–based voice assistants. Note: Data is generated, then annotated, filtered, and enriched (AI preparation). It is subsequently used to train a machine learning model, which produces outputs verified by workers (AI verification). In some cases, workers replace automated processes in real time to complete training datasets or to correct mistaken outputs (AI impersonation).
Source: Tubaro, Casilli, and Coville (2020).

perform the tasks that the AI system is unable to execute due to lack of data, defective parameters, or flaws in the code. Later updates may be needed if, for example, new words emerge to describe known phenomena (like the annually updated names of tropical storms and hurricanes) or the very phenomena change (due to unusual weather activity or climate change, for example). In the following subsections, we show that these solutions require, at the same time, large amounts of personal data and large amounts of human work to make this data suitable for use in machine learning. It is from this mix of factors that the recent scandals originate, and as we will see, they present only the tip of the iceberg. Platform-mediated labor is required for each phase of the production of an AI system: preparation, verification, and impersonation (figure 10.1).

AI Preparation

We refer to the tasks needed for data generation (voice recordings) and for data enrichment or, more technically, data annotation (transcriptions, classifications) as "AI preparation" (Tubaro, Casilli, and Coville 2020). Often referred to as "AI training" in the industry, this is an early phase that occurs at input level, starting before a voice

assistant is marketed. At this stage, paid workers include not only listeners who take other people's recorded voices as inputs, but also speakers—workers who get remunerated to record their voices before listeners can start their operations.

In the preparation phase, companies use specialized digital platforms to recruit on-demand workers to generate the recordings needed to train algorithms before they are brought to market. These platforms (Casilli and Posada 2019) allow client companies to fragment large data projects into small tasks and allocate them to masses of anonymous providers, each of whom executes a tiny part of the whole. Tasks may consist of reading aloud and audio recording a few short sentences. One in five respondents to our French questionnaire reported that this was the latest task they had done. Some mentioned variants such as recording five different ways to ask the same question and reading aloud (and recording) a story for children. These tasks receive a fixed, usually low price. For example, on Atexto, a specialized platform that offers voice-related tasks exclusively, the recording of 10 sentences in Spanish was paid $0.50 in winter 2020 and was expected to take only a couple of minutes. Rates are similar (though slightly more variable depending on customers) on Clickworker, a larger platform with a significant presence throughout Europe and North America, which says on its blog that "voice recording is one of the most popular job types on our platform" (Bayhan 2019). On Upwork, an international freelancing platform, remunerations are higher (between $10 and $20) but for larger batches of 140–300 sentences taking up to 75 minutes.

These platforms also offer other tasks that serve to enrich the "raw" data obtained through recordings. Some of our questionnaire respondents said that they did tasks that entailed checking the quality of the audio recordings made by other workers (curiously enough, one of them reported that she once recognized her own voice). Other common enrichment tasks are transcriptions and classifications of audio datasets. Human-made transcriptions provide examples of how to associate sounds to words—a necessary step before giving them meaning. Classifications serve to add metadata—for example on the emotional state of the speaker—with the ultimate purpose of fine-tuning the voice assistant's understanding and nudging it toward more suitable responses. Depending on the platform, pay rates range from $0.085 to $1 per audio minute (an expression that denotes the length of the audio file to be transcribed, not the time spent working on the transcription).

Despite the media revelations that we presented at the beginning of this chapter, the large amount of personal data that is collected, stored, and handled at this stage has attracted relatively little attention so far. Voice is one of several features unique to the physical and physiological identity of a person and is therefore considered personal data—protected in the European Union under the General Data Protection Regulation

(GDPR). In this sense, even the recording of a paid worker reading aloud a standard, client-provided sentence raises potential privacy issues. In a not-too-distant future, potential advances in voice recognition may make individual reidentification increasingly easy. While workers are asked for their consent when they accept these tasks, compliance with the GDPR requires further provisions such as the right to opt out later on—provisions that require coordination between platforms and final clients and are thus difficult to implement.

AI Verification

We alluded earlier to the common claim that AI training may be transitory and disappear as the technology matures. Indeed, one may speculate that future progress might reduce the data needs of machine learning. The potential development of "unsupervised" algorithms that are able to find patterns in data without any need for enrichment (such as classifications) may lower the demand for third-party labor for preparation tasks. But even under this scenario, there will still be a need for human work at output level, to review and, if necessary, to correct the results of the algorithm. This is what Tubaro, Casilli, and Coville (2020) call "AI verification." Once a voice assistant has been developed and brought to market, its functioning needs to be continuously checked for accuracy: Does it provide the correct answer when people ask a question about (to stick to the same example) the weather? If not, what may have gone wrong? What needs to be added or changed? This is precisely the work of the human remote listeners highlighted in the 2019 press stories and explains why there is a need not only for standardized examples created in the production phase but also for real-world voices of actual consumers in everyday settings—despite the ensuing privacy violation issues.

In the course of our fieldwork, we realized that verification tasks for voice assistants are not generally offered on generalist platforms, where anyone may register, because of the sensitive information they involve. They are more commonly outsourced through what Casilli et al. (2019) call "deep labor" platforms, such as Appen and Pactera Edge. These platforms often require an entry qualification test for new workers and usually pay by the hour rather than by piecework (see also chapter 8 in this volume). One of our interviewees, committing to work for a fixed number of hours every week, received $10 per hour—although after a few months, the number of assignments dropped significantly, reducing the hours and, consequently, the earnings. These platforms are usually tightly managed by a network of subcontractors, with a division of labor in which each of them deals with one aspect (contracting, day-to-day operations, payments, technical infrastructure, etc.). The complexity of the system does not facilitate tight control of privacy protection; this particular interviewee was never asked to sign

any confidentiality agreement regarding handling personal data (whether it included consumers' own voices or personal details about their lives).

When the voice assistant producer is a major multinational company and the amount of data to be processed is consistently large, verification tasks may be outsourced to salaried workers recruited by a subcontractor as part of a business process outsourcing (BPO) value chain. These are repetitive jobs often paid at minimum wage, but with the safeguards and protections associated with salaried employment, these workers at least enjoy more predictable flows of assignments and income. Workers usually must sign strong NDAs, and, according to our interviewees, they are not even allowed to share work-related information with their office mates. As discussed later, these jobs do not offer workers adequate protection against the psychosocial risks induced by the nature of their work. Both the problematic (and sometimes distressing) nature of the real-life content they listen to and the obligation to keep the information confidential constitute a burden that these human remote listeners are often left alone to handle.

AI Impersonation

According to Tubaro, Casilli, and Coville (2020), "AI impersonation" occurs when human paid labor does not support the processes of data production or algorithmic quality assurance but replaces them when they fail. Nontransparent use of low-paid humans instead of algorithms may seem deceptive to consumers and the general public. However, "faking" AI is sometimes done for legitimate reasons, such as to understand the production process of some functionality before actually designing an algorithm to automate it. In the field of AI, this "Wizard of Oz" technique dates back to the second half of the twentieth century. Initially designed for natural language processing, it consists of setting up an experiment where a hidden human (the "wizard") simulates the behavior of a smart computer application to test how users react to a system that they believe to be autonomous (Kelley 1983). In this sense, AI impersonation can be close to AI preparation, and dissimulation of the presence of human listeners may be justified by corporate officials as a way of "faking it until you make it." According to the founder of an AI start-up whom we interviewed in 2018, "What people don't realize is that the vast majority of B2B start-ups we know of are human-based." This is not fraud or deception but rather "a gamble on the future. They have to create their own data . . . and then, on that basis, develop machine learning models *hoping that one day* the process will be automated" (Casilli et al. 2019, emphasis added).

Sometimes, impersonation also verges on AI verification, when correction of algorithmic failures requires real-time human intervention. The development of Google Duplex, an early-stage AI-based voice assistant to make restaurant reservations by phone

on behalf of clients, illustrates both cases. Asked to comment on this in 2019, the company itself admitted that about 25 percent of calls placed through Duplex started with a human, and about 15 percent of those that began with an automated system had a human intervene at some point (Chen and Metz 2019). The human impersonator can be seen as preparing AI in the former case and verifying/correcting it in the latter.

Generally speaking, impersonation happens when humans outperform computers in terms of either efficacy or cost. The idea that prompted Amazon to launch its pioneering platform Mechanical Turk in 2006 was to integrate humans (amusingly termed "artificial artificial intelligence") directly into software programming. According to Irani (2015, 225), this platform was "born out of the failures of artificial intelligence to meet the needs of Internet companies." Ekbia, Nardi, and Šabanović (2015) further stretch this idea by suggesting that failures are actually constitutive aspects of AI production. When new technological solutions are introduced, they promise to automate some process and to be labor saving. But whenever technical limitations threaten delivery on this promise, "human beings are brought into the fold" (7). Cheap labor is one of the tools available to AI producers to compensate for the shortcomings of their technologies. This is an instance of the tension, illustrated by Sadowski (see chapter 13), between technology companies' promises and the masked inputs of human labor they nevertheless need.

Local and Remote AI-Oriented Labor

In virtually all fields of AI, cost efficiency factors in the decision to bring human workers into the loop. In principle, AI preparation, AI verification, and AI impersonation tasks are not geographically bound to any place and can be performed from anywhere, provided there is a computer (or even just a tablet or a smartphone) with broadband. When it comes to voice-activated solutions, linguistic factors largely determine the trade-off between recruitment of local and remote listeners. If a company wants to market a voice assistant to a new country, it needs to ensure its algorithm is trained in the languages spoken in that country. Data preparation and verification in particular often require perfect language fluency and sometimes a more fine-grained knowledge of regional dialects, idiomatic expressions, or domain-specific vocabulary. Therefore, these tasks will be primarily directed at native speakers, sometimes targeting specific countries: on platforms, it is not uncommon to see voice-recording tasks directed exclusively to, say, "Mexican Spanish speakers" or open only to residents of Mexico, while enrichment tasks are more often open to any native speakers of Spanish.

This poses additional constraints on speech recognition models compared to other fields of AI. Data annotation for computer vision algorithms is customarily outsourced

on behalf of European or North American companies to foreign workers wherever labor costs are low (Schmidt 2019; see also chapter 8 in this volume). Instead, we have encountered several instances of voice-related tasks performed within Europe—in Bulgaria, France, Ireland, and Spain. Linguistic needs are likely to keep at least some of these activities within the same countries where the final consumers are located. Nevertheless, recruiting local listeners does not necessarily mean that labor is negotiated locally. Deep labor platforms, for instance, put in place complex outsourcing networks that span national boundaries to enable cost optimization. Some of our interviewees identify as native speakers of French, live in France, and were paid to listen to recordings of French users. But they were contracted by international platforms on behalf of foreign clients, and their day-to-day managers, technical support officers, and other interlocutors were spread throughout Europe (figure 10.2).

Additionally, some degree of offshoring still takes place, at least among countries sharing a common language. This mirrors complex historical interdependencies, so that,

Figure 10.2
The circuitous supply chain of data produced by platform workers to train voice assistants. Note: The data produced by platform workers to train voice assistants establishes a circuitous supply chain spanning across several countries. In one case, AI verification is handled by a Chinese platform (1) that relies on a Japanese online service (2) and its Spanish subsidiary (3) to recruit workers in France (4). Workers are supervised by an Italian company (5) and use a technical infrastructure directly controlled by the final client, a producer of smart speakers located in the US (6).
Source: Authors' elaboration.

for example, a lot of French-language work goes to French-speaking African countries. It is for this reason that, while offshoring is possible only to a limited extent in the specific case of voice assistants, the companies targeted by the media revelations that opened this chapter rely heavily on data associates and raters located in low- and medium-income countries, especially in Southeast Asia. As attested in the literature about platform-based outsourcing, these flows of labor follow existing paths of economic, cultural, linguistic, and political dependency (Fuchs 2016; Casilli 2017; Graham, Hjorth, and Lehdonvirta 2017; Couldry and Mejias 2019). Major platforms and platform-based BPO vendors are located in countries or zones that offer particular legal and economic advantages, and they manage their international workforce from there. For tech companies, platformization represents a way to circumvent fiscal, privacy, and labor regulations, by arrogating to themselves a problematic double freedom of movement—both of labor and of data, at planetary scale. In the platform economy, the workforce is geographically dispersed and distributed along constantly reconfiguring supply chains. Digital intermediation services act as "techno-immigration systems" (Aytes 2012) that allow remote access to foreign workers without importing labor through immigration and even without offshoring via direct foreign investment.

In sum, each phase of the production of AI leverages a planetary labor arbitrage by coordinating local workers (both employees and local contractors of tech companies) and remote third-party workers. Overall, the production and development of voice assistants displays a complex geography, intertwining national and international dimensions, and organized along linguistic lines. This two-level labor arbitrage goes hand in hand with an equally complex privacy arbitrage, whereby clients and platforms seek to simultaneously lower labor costs and channel data to countries where privacy and data protection laws provide uneven levels of safeguard.

Consumers and Workers: A Dual Set of Privacy Risks

Due to the very nature of the technology underpinning them, voice assistants entail risks for all the humans involved—both consumers and workers. The presence of the latter engenders risks for the former: when producers of smart devices conceal the role of flesh-and-blood workers behind automation, consumers underestimate the likelihood and magnitude of potential privacy violations; the broader implications of consenting to terms of use are unclear. Consumers may not know that the assistant records not only bland questions such as "What is the weather like today?" but also their requests for, say, a certain type of porn videos or a prescription drug needed for a condition they are otherwise unwilling to disclose.

Our interviews with former remote listeners point to further risks largely obfuscated from view. Consumers are unlikely to realize that individual reidentification is difficult but not impossible. In principle, recordings are anonymized and fragmented into small bits of a few seconds before reaching human remote listeners, but they may occasionally include mentions of names, addresses, or even social security numbers. Consumers also dictate text to their voice assistants—personal and professional messages and documents that may include sensitive or identifying information. Additionally, remote listeners told us that they might come across several pieces of the same conversation in the same series of transcriptions, enough to draw up a basic profile of the user or their mood at the moment. An even more widespread problem is that a voice assistant may switch on inadvertently and record conversations that were not intended for it, including, for example, interactions with children. This occurs because the algorithms for detection of "wake words" (that is, the special words or phrases that are meant to activate voice assistants, like "Siri" and "OK Google") are still largely imperfect (Coucke et al. 2019).

Another set of risks emerges when we shift our attention from AI consumers to the invisibilized workers that make AI possible. Workers do not only analyze and transcribe consumers' data; they also produce data themselves, especially in the AI preparation phase. While the short recording of a standard sentence is generally harmless (for example, pronouncing the name "Alexa"), multiple recordings of information about the worker (such as their location, skills, preferences, etc.) may be matched with other data or metadata (type of device used, time to perform a task, platform logs, etc.) to produce identifying information.

Casilli et al. (2019) also identify four main categories of psychosocial risk for all these workers—whether they be primarily speakers or listeners. The first risk is losing sight of the objective of their work: especially those who perform tasks that involve recording their voice for data generation are often ignorant of the fact that this activity serves to produce AI-based virtual assistants. The companies that post these tasks on platforms do not always explain their purposes, and our fieldwork indicates that even some experienced, assiduous workers miss this point. The remote listeners who transcribe and check conversations for verification purposes are more likely to know—or at least guess—the purpose of their work. However, as one interviewee told us, especially on platforms, workers are sometimes left alone to handle the personal information of users. This case occurred before GDPR legislation entered into force in Europe; nevertheless, it suggests that platforms shift the responsibility to protect the personal data they handle onto workers. The objectives of the work become fuzzy—is it about improving the technology or enabling massive intrusions into other people's lives?

A second psychosocial risk is loss of control over the quality of work. Especially on platforms, tasks may be rejected, but workers are not systematically given feedback or allowed to redo their assignments. Contact with support services is often difficult. Under these conditions, it becomes hard to define quality standards, and any rejection is perceived as an abuse. A third risk involves suffering and conflict at work. On platforms, demand for labor is highly volatile, and workers are exposed to broad competition. Even with BPO vendors that secure more predictable flows of tasks, job contracts are often short-lived. Hence, workers constantly feel that there is insufficient demand and that they have to fight against others for tasks. The fourth risk is isolation. Especially on platforms, work from home and anonymity may prevent communication and sharing with others. Some platforms do offer online forums to let workers get in touch with each other, but others do not, and a few actively limit communication in order to avoid any claims or protests (Casilli et al. 2019). In some cases, workers do not even manage to share their experience and concerns with their closest social group of family, partner, and friends. Many of our questionnaire respondents said they find it difficult to explain what they do on platforms, how, and why. Isolation may prevent the emergence of a common reflection on working conditions and on possible improvements. The individual has no control over his or her working environment.

Conclusion

The collection and disclosure of the personal data of very large numbers of people are needed for the development and commercialization of AI-driven assistants. Human voice, which as discussed constitutes personal data, must be gathered from the platform workers who execute data generation tasks and from those who use the product. Their recorded voices are necessarily disclosed to other people, whom we have referred to here as listeners: the workers who perform data enrichment tasks and those who check whether the AI has correctly understood users' utterances.

This is not to say that there are no efforts to design privacy-preserving algorithms. For example, recent research uses "federated learning" as a decentralized procedure that enables the training of a central model on the local data of many users, without the need to ever upload this data to a central server (Leroy et al. 2019). However, this alleviates the problem on the side of consumers without reducing that of the workers. What is at stake here is the privacy of both groups. In this respect, research in computer science is working toward solutions to design tasks that keep platform workers' identities more protected (see Duguépéroux and Allard 2020), but risks still remain high as of mid-2021.

Contrary to optimistic claims commonly heard in the industry, it appears that the need for human labor to support AI voice assistants is unlikely to be temporary. Even when a new voice-activated technology is sufficiently mature and no longer needs human impersonators, the need for workers to prepare datasets and, more importantly, to perform quality checks on outputs remains high. It can even be expected to grow as more and more applications of voice technologies emerge, from home automation to industrial production and even health services (where, for example, nurses can simultaneously provide care to patients and dictate notes about their condition).

Privacy risks for both consumers and workers and psychosocial risks for workers go together and stem from the very nature of production in these systems. The way voice-based AI is produced today, requiring massive amounts of data, generates the need for ever-larger masses of recordings at all stages of the production process, with potentially disclosive effects. The models of industrial organization that best serve these technologies, based on subcontracting and platformization, place workers in a position of disadvantage and expose them to the consequences of precarity and low pay. Every country in which a voice-activated device is sold needs an internal supply of workers for audio data production and verification—a need that is going to grow, rather than diminish, as these technologies spread worldwide. At the same time, whenever linguistic factors allow recruitment of workers in lower-income countries, the production chains extend across borders to (for example) Madagascar from France. In these cases, competition among workers from multiple sites brings remunerations further down, and deterritorialization disconnects workers from the very purposes of their activities. As data crosses national boundaries, working conditions deteriorate and risks of leakages increase. Thus, any solution must be dual—protecting workers to protect consumers.

References

Amazon. n.d. "Alexa and Alexa Device FAQs." Accessed September 10, 2021. https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230.

Aytes, Ayhan. 2012. "Return of the Crowds: Mechanical Turk and Neoliberal States of Exception." In *Digital Labor: The Internet as Playground and Factory*, edited by Trebor Scholtz, 79–97. New York: Routledge.

Bayhan, Rabia. 2019. "Get Paid to Read Out Loud—Audio Recording Jobs." Clickworker, October 16. https://www.clickworker.com/2019/10/16/get-paid-to-read-out-loud-audio-recording-jobs/.

Carr, Austin, Matt Day, Sarah Frier, and Mark Gurman. 2019. "Silicon Valley Is Listening to Your Most Intimate Moments." *Bloomberg Businessweek*, December 11.

Casilli, Antonio A. 2017. "Digital Labor Studies Go Global: Toward a Digital Decolonial Turn." *International Journal of Communication* 11: 3934–3954. https://ijoc.org/index.php/ijoc/article/viewFile/6349/2149.

Casilli, Antonio A., and Julián Posada. 2019. "The Platformization of Labor and Society." In *Society and the Internet: How Networks of Information and Communication Are Changing Our Lives*, 2nd. ed., edited by Mark Graham and William H. Dutton, 293–306. Oxford: Oxford University Press.

Casilli, Antonio A., Paola Tubaro, Clément Le Ludec, Marion Coville, Maxime Besenval, Touhfat Mouhtare, and Elinor Wahal. 2019. *Le Micro-travail en France. Derrière l'automatisation, de nouvelles précarités au travail?* [Micro-work in France. Behind Automation, New Forms of Precarious Labor?]. Final Report of the DiPLab (Digital Platform Labor) Project. http://diplab.eu/wp-content/uploads/2019/05/Le-Micro-Travail-En-France_DiPLab-2019.pdf.

Chen, Brian X., and Cade Metz. 2019. "Google's Duplex Uses A.I. to Mimic Humans (Sometimes)." *New York Times*, May 22.

Coucke, Alice, Mohammed Chlieh, Thibault Gisselbrecht, David Leroy, Mathieu Poumeyrol, and Thibaut Lavril. 2019. "Efficient Keyword Spotting Using Dilated Convolutions and Gating." *ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech, & Signal Processing*, 6351–6355. https://doi.org/10.1109/ICASSP.2019.8683474.

Couldry, Nick, and Ulises Ali Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.

Cox, Joseph. 2019. "Revealed: Microsoft Contractors Are Listening to Some Skype Calls." *Vice*, August 7. https://www.vice.com/en/article/xweqbq/microsoft-contractors-listen-to-skype-calls.

Day, Matt, Giles Turner, and Natalia Drozdiak. 2019. "Amazon Workers Are Listening to What You Tell Alexa." *Bloomberg*, April 10.

Duguépéroux, Joris, and Tristan Allard. 2020. "From Task Tuning to Task Assignment in Privacy-Preserving Crowdsourcing Platforms." In *Transactions on Large-Scale Data- and Knowledge-Centered Systems XLIV*, edited by A. Hameurlain, A. M. Tjoa, P. Lamarre, and K. Zeitouni, 67–107. Berlin/Heidelberg: Springer.

Ekbia, Hamid R., Bonnie Nardi, and Selma Šabanović. 2015. "On the Margins of the Machine: Heteromation and Robotics." *iConference 2015 Proceedings*. https://www.ideals.illinois.edu/handle/2142/73678.

Frier, Sarah. 2019. "Facebook Paid Contractors to Transcribe Users' Audio Chats." *Bloomberg*, August 13.

Fuchs, Christian. 2016. "Digital Labor and Imperialism." *Monthly Review* 67 (8): 14–24. https://monthlyreview.org/2016/01/01/digital-labor-and-imperialism/.

Graham, Mark, Isis Hjorth, and Vili Lehdonvirta. 2017. "Digital Labour and Development: Impacts of Global Digital Labour Platforms and the Gig Economy on Worker Livelihoods." *Transfer: European Review of Labour and Research* 23 (2): 135–162. https://doi.org/10.1177/1024258916687250.

Hern, Alex. 2019. "Apple Contractors 'Regularly Hear Confidential Details' on Siri Recordings." *The Guardian*, July 26. https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings.

Irani, Lilly. 2015. "Difference and Dependence among Digital Workers: The Case of Amazon Mechanical Turk." *South Atlantic Quarterly* 114 (1): 225–234.

Kelley, J. F. 1983. "An Empirical Methodology for Writing User-Friendly Natural Language Computer Applications." In *CHI '83: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 193–196. New York: Association for Computing Machinery.

Leroy, David, Alice Coucke, Thibaut Lavril, Thibault Gisselbrecht, and Joseph Dureau. 2019. "Federated Learning for Keyword Spotting." *ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech, & Signal Processing*, 6341–6345.

Martinez, Michael, and Lori Cameron. n.d. "Voice Technology: As Google Duplex Wows and Scares, a Post-screen World Emerges with Questions That the Smart Speakers Cannot Answer." IEEE Computer Society. Accessed September 10, 2021. https://www.computer.org/publications/tech-news/trends/voice-assistants-technology-smart-speakers.

Schmidt, Florian A. 2019. "Crowdsourced Production of AI Training Data: How Human Workers Teach Self-Driving Cars How to See." Working Paper Forschungsförderung no. 155, August. Düsseldorf: Hans-Böckler-Stiftung. https://www.boeckler.de/pdf/p_fofoe_WP_155_2019.pdf.

Tubaro, Paola, Antonio A. Casilli, and Marion Coville. 2020. "The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence." *Big Data & Society* 7(1). https://doi.org/10.1177/2053951720919776.

Verheyden, Tim, Denny Baert, Lente Van Hee, and Ruben Van Den Heuvel. 2019. "Hey Google, Are You Listening?" *VRT NWS*, July 10. https://www.vrt.be/vrtnws/en/2019/07/10/google-employees-are-eavesdropping-even-in-flemish-living-rooms/.