

<http://doi.org/10.1590/15174522-114092>

Waiting for robots: the ever-elusive myth of automation and the global exploitation of digital labor

Antonio A. Casilli* 

Abstract

Discourses of robotic replacement and of the end of work have survived to the present day. But more and more voices now challenge the very idea that technological innovation is necessarily conducive to job loss. According to several studies, new high-tech jobs is accompanied by an even bigger low-tech job creation, and AI can be expected to be no exception. Based on new evidence about the role of human-annotated data in machine learning and algorithmic solutions, a new generation of scholars are now studying the germane phenomena of “heteromation”, “automation last mile” or, more simply, platform-based digital labor needed to generate, train, verify, and sometimes modify in real-time huge quantities of examples that machines are supposed to learn from. Digital labor designates datified and taskified human activities. The first type of platform occupation is *on-demand labor*. The second type of platform-based digital labor is *microwork*. Finally, the third type of digital labor is *social networked labor*.

Keywords: platform work, human work, on-demand labor, taskification, microwork.

* Polytechnic Institute of Paris, Paris, France.

À espera de robôs: o mito sempre evasivo da automação e a exploração global do trabalho digital

Resumo

Discursos sobre substituição robótica e o fim do trabalho persistem até hoje. No entanto, cada vez mais vozes contestam até mesmo a ideia de que inovação tecnológica leve necessariamente à extinção de empregos. Segundo diversos estudos, novos empregos de alta tecnologia são acompanhados pela geração ainda maior de empregos de baixa tecnologia, e pode-se assumir que a IA não seja exceção. Com base em novas evidências sobre o papel de dados comentados por humanos no aprendizado de máquinas e em soluções algorítmicas, uma nova geração de pesquisadores está agora estudando os fenômenos relacionados da “heteromação”, “automação do último quilômetro” ou, simplesmente, trabalho digital baseado em plataforma, necessário para gerar, treinar, verificar e, às vezes, modificar em tempo real grandes quantidades de exemplos com os quais as máquinas devem aprender. Trabalho digital significa atividades humanas *datificadas* e *taskificadas*. O primeiro tipo de ocupação em plataforma é o trabalho sob demanda. O segundo tipo de trabalho digital baseado em plataforma é o *microtrabalho*. Finalmente, o terceiro tipo é o trabalho baseado em redes sociais.

Palavras-chave: trabalho de plataforma, trabalho humano, trabalho sob demanda, taskificação, microtrabalho.

The monolith of full automation

The myth of full automation has been the proverbial Kubrickian monolith around which clusters of innovators and social commentators have been huddling for centuries. Tribes of investors have anticipated pervasive labor-saving technologies as a chance to further reduce the wage share while unevenly appropriating the output of productivity gains, as policymakers have been oscillating between the effort to mitigate their expected negative consequences and the fear that regulation may curb economic growth.

The pandemic crisis of Covid-19 has been a potent energizer for common narratives about the use of such technologies to pursue production in a world of socially distanced workers. Pandemic-induced recessions are

triggering employment losses in occupations easier to automate, leading to “jobless recoveries” (Hershbein; Kahn, 2018; Jaimovich; Siu, 2020). Rather than a linear process, automation results from strategic decisions that are easier to make in times of economic crisis. MIT researchers David Autor and Elisabeth Reynolds point out that the concomitant effect of working from home, urban de-densification, and employment concentration in large firms will stop the rapid employment growth observed in global societies until 2019 and bring about a phenomenon that they characterize as “automation forcing” (Autor; Reynolds, 2020). Companies seize the opportunity of labor market where demand surpass supply to impose innovations that allegedly replace workers.

These conclusions are shared by Oxford economist Michael Osborne who publicly stated that Covid-19 would finally fulfill the prophecy contained in a controversial working paper that he and his colleague Michael Frey penned back in 2013 (Williams, 2020). The study, entitled “The future of employment”, foresaw that machine learning and mobile robotics would destroy 47 percent of US jobs by 2030 (Frey; Osborne, 2013). By referencing this piece of academic literature, we intend to point out the murky links between the prophecy of full automation and the rhetoric of human replacement by machines. Frey and Osborne’s high-impact study was translated into myriad news stories that bought into the troublesome narrative of “the great replacement” – and inevitably echoed the theories of far-right and white supremacist milieus: “Robots could replace humans in a quarter of US jobs by 2030” (Carson, 2019), “Will robots take over your job?” (Haldane, 2015), “After the robot revolution, what will be left for our children to do?” (Seager, 2016).

Although the study’s methodology was already called into question back when the working paper was first published, and the results challenged by other estimates of a much more modest displacement effect on employment,¹ Osborne’s viewpoint on the inevitability of automation reactualizes an

¹ See specifically Arntz *et al.* (2016), a study commissioned by OECD, which estimates that, on average, just 9% of the jobs in OECD countries were at high risk of being automated.

even older perspective – the theory of the “end of work”.² Since the 1677 presentation by the French inventor Jean-Baptiste de Gennes of the first weaving machine that did not require workers, it has represented a utopian (or dystopian) prospect that consistently failed to materialize. In the first half of the 19th century, economists and thinkers of the first industrial revolution addressed the role of technologies in displacing labor. Nevertheless, their approach was less steeped in absolute certainty and more interested in detecting the factors that would bring about automation. In his *Lectures on the elements of commerce* (1801), Thomas Mortimer pointed out that while there is a category of machine conceived to reduce the drudgery of human work, another one was designed to “almost totally exclude the labour of the human race” (Mortimer, 1801, p. 72). Moral principles and sound policymaking, the author insists, counter the generalization of the latter. David Ricardo devoted the chapter 31 of the 1821 edition of his *Principles of political economy and taxation* to the role of machinery. In it, he insisted that the capitalist’s “temptation to employ machinery” stems not from a deterministic historical trend but from a purely pragmatic and microeconomic tendency to reduce labor costs. Automation is only one of the options available to business owners, who weigh its costs against those of “foreign trade” (which included offshoring of operational processes) and even the “labour of horses” (Ricardo, 1821). In his *Philosophy of manufacturers* (1835), Andrew Ure adds that by simplifying and standardizing complex tasks, automation does not aim to “supersede human labour altogether” but to make specialized workers compete against “ordinary labourers”, women, and children (Ure, 1835, p. 23) in order to bring down the cost of their labor.

Despite these calls for caution, the twin discourses of robotic replacement and of the end of work have survived to the present day. But more and more voices now challenge the very idea that technological innovation is necessarily conducive to job loss. According to several studies, new high-tech jobs is accompanied by an even bigger low-tech job creation, and

² See for instance Rifkin (1999).

AI can be expected to be no exception (Goos *et al.*, 2015). Economist David Autor (mentioned before) argued in his 2015 article “Why are there still so many jobs?” that productivity gains fostered by automation can sometimes have unexpected results. Take for example automatic cash dispensers implemented in banks all around the world since the late 1980s. Where one would expect a dramatic drop in the number of employees in the banking sector, the opposite happens: it increased over the following decades (Autor, 2015). Actually, ATMs made opening new branches less costly, thus encouraging banks to hire more tellers. Admittedly, nowadays a bank teller’s job description does not only involve counting and handing cash. Their competencies include counseling their clients on financial matters and selling banking products and services. But although the skills required for an automated occupation can change dramatically, the idea that automation necessarily destroys jobs is disproved by this and other comparable cases.

What is digital labor?

Given these empirical and theoretical shortcomings, why is the deterministic belief in the job displacement effect of automation still so prevalent among experts and laypersons alike? The main reason is the disproportionate emphasis put by present-day research on the usage side of technological innovation at the expense of the production side. Especially when it comes to studying AI-systems and intelligent solutions, analyses tend to focus on the deployment phase of business-process automation, while largely failing to take into account the material conditions and the living labor necessary to produce artificial intelligence. This tendency, combined with lack of transparency due to industrial IP, creates a serious knowledge gap that recent approaches are trying to bridge. Based on new evidence about the role of human-annotated data in machine learning and algorithmic solutions, a new generation of scholars are now studying the germane phenomena of “heteromation” (Ekbia; Nardi, 2017), “automation last mile”

(Gray; Suri, 2019) or, more simply, platform-based digital labor needed to generate, train, verify, and sometimes modify in real-time huge quantities of examples that machines are supposed to learn from (Tubaro *et al.*, 2020).

Digital labor designates datified and taskified human activities. By datification we mean that within platform ecosystems every human productive gesture can be turned into a data point. Datification is both a mode of extraction of value in the form of data and a mode of governance of labor by cycles of algorithmic management (Rosenblat; Stark, 2016). Taskification is in turn the process of atomizing and fragmenting complex processes by reducing them to standardized tasks, easier to coordinate to produce scalable and modular technological arrangements.

One way of systematizing the types of digital labor on contemporary platforms is to look at geographic embeddedness and at levels of remuneration.

The first type of platform occupation is *on-demand labor*, which takes place on services such as Uber, Lyft, Airbnb, Taskrabbit, Glovo and Deliveroo that provide logistic, transportation, and care services. Workers' activity is digital insofar as it is app-based, but it can be also described as geographically sticky (Ojanpera *et al.*, 2017), i.e., as connected to an urban area, to a city or a neighborhood. On-demand workers have been unevenly impacted by Covid-19. Of course, drivers and those working on hospitality platforms such as Airbnb have seen a sharp decrease of revenues during periods of lockdown, but this has been compensated by the increased demand in logistic, delivery, and business-to-business services. Moreover, some platform workers have acquired strong visibility during the pandemic crisis, and this has boosted political awareness and workers' organization among riders and delivery operators (ILO, 2021).

The second type of platform-based digital labor is *microwork*, which consists of small repetitive tasks outsourced to crowds of independent providers. Although regarded as low-skilled – and thus poorly paid – this activity is important to automate a number of functions, from accounting, to sales, to manufacturing. Less visible and less recognized than their bike-riding and cab-driving counterparts, microworkers can be found in several

countries. They represent a sizable portion of the over 160 million estimated online workers all around the world (Kässi *et al.*, 2021). In some cases, microwork is performed by users on specialized platforms, such as Amazon Mechanical Turk, where workers can look for HITs (Human Intelligence Tasks) and contribute to what Jeff Bezos famously presented as “artificial artificial intelligence” (Bezos, 2006). Each task is usually paid a few cents and lasts, in some cases, less than a minute. The average hourly rate on Amazon Mechanical Turk was estimated at less than 2 dollars (Hara *et al.*, 2017). Given the volatility of the demand of microwork and the power of requesters to arbitrarily deny remuneration to workers, this type of activity does not represent a stable income source. Nevertheless, it is increasingly popular in several countries in the global South, where it can constitute the primary revenue of lower-income households.

Since Covid-19 kicked in, microwork platforms have become even more popular. Some of them sport over ten million workers, and an increase of over 30% of contributor activity since March 2020 (Ambalina, 2020; Rochweger, 2020). They attract users by presenting themselves as marketplaces for labor with low barriers to entry and no qualifications required, where people can earn money just by performing simple actions involving images, videos, and text. The purpose of these tasks is often not well-defined, for reasons that include both trade secrets and the concern that workers might bias the process if they knew in advance the expected outcome.

Finally, the third type of digital labor is *social networked labor*. It designates the activities that are indispensable to run social media and generic user-based platforms, such as Google, Facebook, Instagram, YouTube. Despite the emphasis that these companies have put on the specialized work of computer scientists, engineers, and code developers, social networked labor does not relate to these elite workers, but rather to unrecognized contributors – either precarious clickworkers or unpaid generators of content and data (Casilli, 2020).

Historically, inquiries into this form of digital labor tended to concentrate on the latter, limiting themselves to content production described as “free

labor” (Terranova, 2000). But pro-bono contribution of amateur content providers is far from being its most relevant instance. The professionalization and monetization of influencers, live-streamers, video creators, community managers have progressively made social networked labor less “free” than initially thought. In addition, users’ contributions are mainly valuable to platforms for the data and metadata they produce, rather than for the content of their posts, images or videos. From this point of view, each user, even those who do not envision to make a career out of their content, produces value for these platforms by simply liking, sharing, following and opting-in to geolocation.

Finally, the very presence of unpaid “producers” (Bruns, 2009) on social media rests on the existence of other back-office socially networked workers, such as content moderators, whose activity is essential to the operation of these platforms. Moderation consists in filtering and flagging problematic content in accordance with ever-evolving guidelines and corporate policies established by commercial platforms. This occupation often cuts across the same logics of labor taskification and outsourcing highlighted before. Underpaid moderators perform psychologically taxing and often challenging duties, and have been adversely impacted by Covid-19. Given the violent, graphic, and sensitive content they manage, moderators cannot simply work from home. Social platforms have been rushing to announce that they would safeguard the health and wellbeing of their moderation teams during lockdowns, while at the same time prioritizing them for the swiftest return to office work (Hatmaker, 2020).

In order to highlight the implications of platforms for automation, we need to delve into the three categories of digital labor.

On-demand labor

On-demand labor is probably best epitomized by its most prominent platform, Uber. Despite the ebbs and flows of the company’s performance, its business model is a *bona fide* cultural success, which gave birth to the word

“uberization”, which now stands from disrupting a market by introducing algorithmic processes.

The technological and organizational innovations introduced by Uber are also paradigmatic of a specific style of labor governance. Since the second half of the 2010s, numerous court cases have rejected the classification of Uber drivers as self-employed partners. In 2017 the European court of justice ruled that Uber should be regulated as a transport services company. Yet, Uber displays several features of a technology business whose drivers are, first and foremost, data providers. Despite the changes in the service over time, almost 2/3 of drivers’ workday is actually spent on producing data on the company’s app while driving around or waiting for new gigs (Cassano, 2016).

The data produced by Uber drivers are in some cases personal information, such as their names, likeliness, address and payment details, that are entered in their profile. Subsequently, Uber drivers produce data whenever they enroll in specific time slots. Drivers send and reply to messages and calls, both from the platforms and from customers, thus generating more information. Location data are very important too, as drivers use several apps in order to negotiate the best route. They sometimes have to use several apps on two or more devices, to compensate for possible gaps in the information provided by Uber’s GPS, thus turning their everyday activity into a hybrid between urban transportation and multi-device data management.

Their reputation score is probably the single most important data point drivers have to take into account. The evaluations and reviews left by riders are used to compute it, based on standard criteria such as hygiene and punctuality, or more elusive indicators such as “chattiness” and driving style. This very subjective, and in some cases arbitrary way of assessing workers’ performance can have adverse consequences. If their score falls below a given threshold, drivers face deactivation of their profile and exclusion from the platform.

The data produced by drivers operationalize Uber’s algorithmic matching models, that put in place economic incentives for drivers based

on real-time estimates of customers' demand for the service. The patented surge pricing algorithm that is used by Uber to match supply and demand has been described by Uber's R&D researchers. Economist Jonathan Hall provides a practical example: on March 21, 2015, at the end of a concert at the Madison Square Garden in New York, several hundred people opened the mobile app at the same time (Hall *et al.*, 2015). Based on that huge surge in the number of online users originating from a specific location, Uber's algorithm was able to calculate a certain positive coefficient by which the price of each ride would be multiplied. For consumers, that translated into a higher price for their ride. But for drivers, that meant an opportunity to earn 40 or 50 dollars for a ride that would usually pay only 10. The Uber app was optimized to include this information. The city maps it features is a mosaic of yellow, orange, and red zones. When a zone become suddenly red, that means better earning opportunities for drivers, who are incentivized to converge to that point. Of course, not all consumers are ready to pay a higher price. They either wait for the price to return to normal or move to another area where the coefficient is 1. In both cases they continue using the app, thus producing data themselves.

Data generated both by drivers and by riders is an important revenue source for the company that monetizes this information by allowing access to municipalities, brands, and advertisers. Another use for the same data, is mainly automation. Geolocation, transaction, and personal data are used to train Uber autonomous vehicles. Despite the recent transfer of their Advanced Technologies unit to Amazon-backed startup Aurora, Uber continues to use data to develop "driverless" cars. Since several decades, autonomous vehicles have been imagined as sci-fi boudoirs on wheels, where passengers can read or chat, while the actual driving is fully automated. Reality is of course different: car manufacturers are no closer to the goal of level 5 automation. On a recurrent basis, manufacturers have admitted that the dream of fully self-driving cars does not match engineering reality, and the presence of persons actively involved in driving is still necessary in autonomous vehicles (Hull, 2021). Nevertheless, Uber does not necessarily

describe these persons as “drivers”, but rather as operators, whose role is to intervene in case of car malfunctioning (Cellan-Jones, 2020). In fact, companies like Uber make everything to render the human labor of their drivers invisible. Some even go as far as disguising them as car seats, allegedly to research how other drivers react to them (Etherington, 2017). By de-emphasizing the role of human drivers in “driverless” cars, platforms have become morally responsible for an increasing number of accident deaths that have occurred over the last decade. In March 2018, an Arizona woman became the first person killed in a road accident involving a Uber autonomous vehicle. Videos of the accidents have been circulated by the police. In the first sequence, taken by the car camera pointing towards the road, we see the victim just one second before the fatal impact. In the second sequence, taken by the camera pointing towards the interior of the car, the “operator” is sitting at the driver’s seat, wearing a seatbelt. Despite reports to the contrary, this person is neither sleeping nor watching a TV show on a tablet, but actually communicating with the Uber platform (KJZZ Phoenix, 2018). If only the role of “operators” as actual and active drivers had been stressed, if this person had received a better training stating their crucial role in situations like this, this accident could have been avoided.

Apart from single human workers that are still needed inside an autonomous car, there is also a multitude of other invisible workers that operate around it. This new category of workers allows us to transition to the second type of digital labor – microwork. Uber self-driving cars are basically mobile communicating devices that exchange data with the platform and with other connected vehicles. But they also capture huge amounts of information from the environment surrounding them. If a pedestrian crosses the street, the car’s lidars and cameras are supposed to record that and then process it to act consequently. In order for the car to autonomously determine the best course of action, recorded images need to be annotated, filtered, documented. Raw data need to be transformed into enriched quality data that Uber cars’ computer vision systems can use.

In the automotive industry, several platforms provide this service by putting companies in touch with workers operating from remote locations. They tag images recorded by the vehicles and identify objects such as other cars, traffic lights, and lanes via a process called semantic image segmentation (Tubaro; Casilli, 2019). The computer vision systems that operate the car are based on machine learning solutions that need billions of examples and data to “learn” how to recognize objects and pedestrians. The persons who metaphorically “teach” computer vision systems are described by former Uber commercial vehicle unit director Anthony Levandowski as “human robots” (Levandowski, 2013). They are basically the humans who perform the work that automated systems are supposed to do: recognize objects, process information, drive...

Microwork

The reason why data annotation is so inextricably associated with the production of artificial intelligence systems is that human-level AI remains unattainable by current methods. The dream of Artificial General Intelligence that drove 20th century pioneers of the field, has been supplanted by “weak” or “narrow AI”, limited and ineffective without humans-in-the-loop. This type of AI relies on machine learning (deep learning, neural networks, generative adversarial networks, convolutional networks etc.), which in turn is predicated on the availability of quality data enriched by myriad workers.

Some digital platforms specialize in providing these services. The aforementioned Amazon Mechanical Turk was introduced by Jeff Bezos in the mid-2000s. Its name refers to the legendary robot chess player invented in the 18th century by Austrian nobleman Von Kempelen. The original Turk was presented as the first artificial intelligence, capable of simulating the cognitive processes of human chess players and thus outperform them. The ruse was that inside the automaton was hiding an actual person who would move the pieces. There are several hypotheses as to their actual identity: a disabled veteran, a small person, a skinny boy... In any case, the operator

was not a grandmaster, but a person whose chess abilities were average at best. Amazon Mechanical Turk is based on a similar idea. Instead of placing data scientists and computer geniuses in charge of training and verifying its automatic solutions, it entrusts this responsibility to laypersons – hundreds of thousands of them.

Imagine, for example, a startup wanting to “teach” smart speakers to play songs and playlists based on titles suggested by users. Not only this startup would need a database containing billions of examples of human voices pronouncing song titles (in different languages, with different accents and speech patterns...), but would also require that the data be arranged so to differentiate music genres, discriminate songs with the same title, associate similar songs etc. Similarly, companies who want to automate business processes turn to Amazon Mechanical Turk or to other microwork platforms to hire persons who will perform these data annotation tasks. They include, for instance, recognizing languages or types of conversations, transcribing text, providing keywords.

Generally, neither platforms nor client companies formally hire the workers. They are a disposable workforce, with no actual employment contract and paid by the piece. Amazon considers them as “consumers-workers” who do not receive a salary but a “reward”, and whose only link to the platform is having agreed to its Terms of Service. When communicating with investors and clients, these platforms often describe workers who accept such precarious conditions and low pay as young, educated, originating from country of the global North, and interested in a freelance activity. This rhetoric mirrors on-demand platforms characterization of their drivers and delivery peoples as flexible and independent “partners”. The underlying reality is more complicated. The majority of these workers are located in low-income countries in South-East Asia, Latin America, and Africa, where average wages are lower and informal economy is significantly more prevalent. Compared to pre-internet offshoring, outsourcing these tasks to a remote crowd on a platform is less demanding for companies, which do not need to create factories or to open up subsidiaries abroad.

Amazon is not the only company that leverages microwork to train artificial intelligence solutions. Microsoft, for instance, has been using it for years to calibrate and optimize its products. The results of its search engine Bing are supervised by microworkers recruited on the proprietary platform UHRS (Universal Human Reference System). Like on Amazon Mechanical Turk, workers accept to perform microtasks paid a few cents, but they do not work for other companies – their only client is the Redmond’s giant. Microsoft is an interesting case because it does not limit itself to recruit paid microworkers. Average internet users can be put to work to train artificial intelligence products too, like in the infamous case of the chatter bot Tay experiment (Wolf *et al.*, 2017). The conversational agent was supposed to simulate the personality and online behavior of a teenage girl. Instead of training it on a dataset of examples of conversations previous to its launch, Microsoft decided to train it live on Twitter. A few hours in, a coordinated campaign from the members of an online forum succeeded in “teaching” Tay all sorts of profanity, homophobic and racist slur, and problematic political opinions. Microsoft immediately discontinued the experiment. Its failure contained a valuable lesson on the importance of quality training for AI solutions. It also was evidence to the tendency of Big Tech platforms to rely on unpaid consumer work to produce them.

In the same way, Alphabet recruits paid microworkers on its proprietary platform Rater Hub to refine and improve the results of Google Search engine. Rater Hub was introduced around 2007 and allows microworkers to verify the quality of the results of sampled search engine queries. Whenever a user looks up for a term, there is a chance that a “rater” is presented with a task that consists in judging if the algorithm has correctly interpreted the query. If the result is relevant, it is marked with a green button, otherwise it is flagged in red. In other cases, results are flagged because they point to banned content, fake news, or contain censored search terms. Like Amazon Mechanical Turk, Rater Hub does not offer proper jobs, but rather fragmented gigs rewarded a few cents each time workers help improve Google’s results.

Social networked labor

Still one can make the case that regular Google users are, to an extent, data annotators as well. Whenever they type a new query in the search engine, they are producing examples of terms and expressions for the algorithm to learn from. This, too, is part of the machine learning that we have already mentioned. This approach is based on axioms that are as much epistemic as they are political: firstly, that a machine *learns from its users*, secondly, that *users must produce* an increasing number of examples to create new rules that machines do not abide by at the outset (Hogan, 2014).

Indeed, using a search engine can hardly qualify as conventional “work”, both from a legal and from an economic point of view. Yet, Google constantly invites its users to produce data for its platforms and services, and in many cases, it has been pretty transparent that the production of data constitutes work. ReCaptcha is probably the most prominent case of user contributions being harvested to produce value in the form of distributed and unrecognized digital labor. For users, ReCaptcha is a page or a pop-up that appears whenever a website prompts them to show that “they are not bots”. In the early 2010s, users were urged to transcribe distorted words. These words were too difficult to read for a text-recognition software, but easily interpretable for an individual with basic literacy skills. While transcribing these words, users were actually training optical character recognition (OCR) algorithms to read scanned texts hosted on the Google Books platform. Books were largely digitized by hand, which sometimes skewed or distorted scanned words. By transcribing these words, initial reCaptcha users were actually annotating the scanned document by adding a tag corresponding to the illegible terms.

As early as 2008, the inventor of the system, Luis Von Ahn, had published an article in the *Journal Science* defining its users as a “workforce” and maintaining that their activity was “work [...] contributing to the digitization of human knowledge” (von Ahn *et al.*, 2008). Meanwhile some platforms like the Russian Kalotibablo were already offering comparable Captcha solving services performed by crowds of microworkers as early as 2010. On these grounds, in 2015 a few hundred citizens of Massachusetts decided to start

a (ultimately unsuccessful) class-action against Google to be reclassified as actual employees (*Rojas-Lozano v. Google Inc.*, 2015). Around that time Google decided to discontinue the text-based reCaptcha and to come up with its new image-based one, where users detect traffic lights and vehicles to improve computer vision systems at Alphabet's autonomous car division—Waymo.

What is the difference between micropaid data annotation work and user-provided unpaid data annotation work? In both cases, platform-mediated human agents are producing data and value. In the case of ReCaptcha, it is common sense to say that users' activity cannot be construed as work because they receive a service in exchange for their contribution. But sometimes the service is not even an actual product available on the market. Translation tools, for instance, can be eternal prototypes, trial versions of actual software. They attempt to translate based on masses of examples suggested or rated by users. Facebook provides automatic translations of posts and messages. Each transaction attempt comes with a button prompting users to come up with a better translation. This user-generated edit is included in a training database for an algorithm, which the users unknowingly improve. This type of activity, too, blurs the lines between work and consumption (Huws, 2003; Dujarier, 2008).

Facebook is a platform and, as such, it captures the value produced by the community of its users. More precisely, Facebook is a "lean platform" that does not own the content it distributes,³ an economic entity halfway between a content farm and a marketplace for data. Unlike YouTube, Facebook does not monetize content. It does not sell the photo, video or post produced by its users, because the data component of each piece of content is way more valuable to the platform and to its advertisers that use them for targeted ads.

Here again, data monetization isn't the only important source of value. Data is also crucial for automation. Content moderation is an example of how AI is fueled by human data labor, both paid and unpaid. Commercial content moderation is the process that allows users to access a family-friendly Facebook, with no violent images, vile messages, or inappropriate

³ For a definition of lean platforms, see Srnicek (2017).

adult content (Roberts, 2019). Facebook has long claimed that it is entirely automated, but since a few years, studies have revealed that teams of microworkers, similar to those on Amazon Mechanical Turk or on Rater Hub, are training filtering algorithms. Beyond their nominal service of removing problematic content, moderators are micropaid to watch videos and tag them, to annotate pictures, to read and classify messages. This moderation and annotation work is performed by humans who are remunerated. In their case, what they do is clearly work. But sometimes content moderation on Facebook happens from below, when the platform involves in this activity average, unpaid users. For instance, whenever users come across an offensive message, an action button invite them to flag it. After that, more details are requested from the users, as to the type of offense, the nature of the content, the persons targeted by it. The flagged content can then be sent to a moderator, who ultimately adjudicates it against the platform guidelines. Both ends of this process involve human activities: at the beginning the unpaid user flags content, at the end the paid moderator decides whether to ban it or not. In the middle, the artificial intelligence is crushed by the volume of tasks necessary to uphold the illusion of full automation.

The perception of this persistent intermingling between artificial intelligence and the digital labor that is necessary to produce it varies considerably across the global North/South divide. Companies that sell AI solutions are mostly situated in high-income countries such as the US, Europe, South Korea, and Japan. Big players have also emerged in China and India.

In order to provide, in real-time, huge quantities of annotated data to train, verify, and manually operate these systems, platforms negotiate digital labor. Sometimes when this workforce is needed to provide on-demand services, it is recruited locally. Other times it is hired in low-income countries. Companies base these strategic choices on economic factors that are inevitably influenced by global dependencies and imbalances of wealth and power between North and South countries. Forms of international polarization

between regions where demand for digital labor is high (as they concentrate major producers of AI solutions) and regions that specialize in supplying this type of labor become increasingly visible. Existing empirical explorations have already underlined how South-Asian and Indian workers produce data for US and UK-based tech companies (Graham *et al.*, 2017; Berg *et al.*, 2018). More recent and ongoing research projects focus on Portuguese-, Spanish-, and French-speaking platforms both in Latin America and in Africa⁴.

So, after all is said and done, the message I want to leave you with is that while we are all waiting for robots – like Beckett’s characters waited for a transcendent Godot – we need to focus on the material conditions of digital labor. Automation is not going to happen without the contribution of underpaid, micropaid, and unpaid workers. This means that our systems will never be “fully” automated. They will always require some level of human intervention.

Nevertheless, something has already happened to the way we work. Endogenous trends towards platformization have combined with exogenous one like the Covid-19 pandemic, to force the taskification and datification of numerous facets of our jobs. Nowadays, formally employed staff members and precarious gig-workers alike are at risk of being, if not fully automated, at least fully platformized. Although robots are not necessarily bound to take them away, this should not sound as good news to you. For an increasing portion of the global workforce, the only occupations that will be available in the future are contingent and underpaid. In the absence of strong regulation, the prospect of accessing formal social security arrangements fades away from them. We should take this into account while designing policies and organizing social movements for the coming decades.

Antonio A. Casilli is professor of sociology at the Télécom Paris school of the Polytechnic Institute of Paris and researcher at the Interdisciplinary Institute on Innovation.

✉ antonio.casilli@telecom-paris.fr

⁴ See the projects TRIA (El trabajo de la inteligencia artificial) and HUSH (Human Supply Chain) that I am developing with my colleagues at the DiPLab (Digital Platform Labor) research group, available at: <http://diplab.eu/>. Cf. also Casilli *et al.* (2019), Moreschi *et al.* (2020) and Miceli *et al.* (2021).

References

1. AMBALINA, Limarc. Companies hiring remote workers during COVID-19. **Lionbridge**, 2 Apr. 2020. Available at: <https://lionbridge.ai/articles/10-companies-hiring-remote-workers-during-covid-19/>
2. ARNTZ, Melanie; GREGORY, Terry; ZIERAHN, Ulrich. **The risk of automation for jobs in OECD countries**: a comparative analysis. OECD Social, Employment and Migration Working Papers, no. 189, OECD Publishing, Paris, 2016. <http://dx.doi.org/10.1787/5jlz9h56dvq7-en>
3. AUTOR, David H. Why are there still so many jobs? The history and future of workplace automation. **Journal of Economic Perspectives**, v. 29, n. 3, p. 3-30, 2015.
4. AUTOR, David H.; REYNOLDS, Elisabeth B. The nature of work after the COVID crisis: too few low-wage jobs. **The Hamilton Project Report**, essay 2020-14, Jul. 2020. Available at: https://www.brookings.edu/wp-content/uploads/2020/08/AutorReynolds_LO_FINAL.pdf
5. BERG, Janine; FURRER, Marianne; HARMON, Ellie; RANI, Uma; SILBERMAN, M. Six. **Digital labour platforms and the future of work**: towards decent work in the online world. Report. Geneva: International Labour Organization, 2018.
6. BEZOS, Jeff. Opening keynote and keynote interview. **MIT World**, special events and lectures, 27 set. 2006. Available at: <http://techtv.mit.edu/videos/16180-opening-keynote-and-keynote-interview-with-jeff-bezos>
7. BRUNS, Axel. From prosumer to producer: understanding user-led content creation. In: *Transforming Audiences*, 2., 3-4 Sept. 2009, London. **Proceedings** [...]. London: University of Westminster, 2009.
8. CARSON, Eric. Robots could replace humans in a quarter of US jobs by 2030. **CNET**, 24 jan. 2019. Available at: <https://www.cnet.com/news/robots-could-replace-humans-in-a-quarter-of-us-jobs-by-2030/>
9. CASILLI, Antonio A. Da classe virtual aos trabalhadores do clique: a transformação do trabalho em serviço na era das plataformas digitais. **MATRIZES**, v. 14, n. 1, p. 13-21, 2020.
10. CASILLI, Antonio A.; TUBARO, Paola; LE LUDEC, Clément; WAHAL, Elinor. En la trastienda de la inteligencia artificial: una investigación sobre las plataformas de micro-trabajo en Francia. **Arxius de Ciències Socials**, v. 41, p. 85-108, 2019.
11. CASSANO, Jay. How uber profits even while its drivers aren't earning money. **Vice**, 2 Feb. 2016. Available at: https://motherboard.vice.com/en_us/article/wnxd84/how-uber-profits-even-while-its-drivers-arent-earning-money
12. CELLAN-JONES, Rory. Uber's self-driving operator charged over fatal crash. **BBC**, 16 Sept. 2020 Available at: <https://www.bbc.com/news/technology-54175359>
13. DUJARIER, Marie-Anne. **Le travail du consommateur**: de MacDo à eBay, comment nous coproduisons ce que nous achetons. Paris: La Découverte, 2008.

14. EKBIA, Hamid R.; NARDI, Bonnie A. **Heteromation, and other stories of computing and capitalism**. Cambridge: MIT Press, 2017.
15. ETHERINGTON, Darrell. Ford disguised a man as a car seat to research self-driving. **TechCrunch**, 2 Sept. 2017. Available at: <https://techcrunch.com/2017/09/13/ford-disguised-a-man-as-a-car-seat-to-research-autonomous-driving/>
16. FREY, Carl B.; OSBORNE, Michael A. **The future of employment**. The Oxford Martin Programme on Technology and Employment Working Paper, 2013. Available at: <https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf>
17. GOOS, Maarten; KONINGS, Jozef; VANDEWEYER, Marieke. Employment growth in Europe: the roles of innovation, local job multipliers and institutions. **SSRN**, Preprint, 1 Oct. 2015. <https://dx.doi.org/10.2139/ssrn.2671765>
18. GRAHAM, Mark; HJORTH, Isis; LEHDONVIRTA, Vili. Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. **Transfer – European Review of Labour and Research**, v. 23, n. 2, p. 135-162, 2017. <https://doi.org/10.1177%2F1024258916687250>
19. GRAY, Mary L.; SURI, Siddharth. **Ghost work: how to stop Silicon Valley from building a new global underclass**. Boston: Houghton Mifflin Harcourt, 2019.
20. HALDANE, Andy. Will robots take over your job? **Financial Times**, 13 Nov. 2015. Available at: <https://www.ft.com/content/af9684ec-8a04-11e5-9f8c-a8d619fa707c>
21. HALL, Jonathan; KENDRICK, Cory; NOSKO, Chris. The effects of Uber's surge pricing: a case study. *In: Designing the Digital Economy*, 24-25 Oct. 2015, Cambridge, MA. **Proceedings** [...]. Cambridge: Microsoft Research New England, 2015. Available at: http://andrewchen.com/wp-content/uploads/2016/01/effects_of_ubers_surge_pricing.pdf
22. HARA, Kotaro; ADAMS, Abi; MILLAND, Kristy; SAVAGE, Saiph; CALLISON-BURCH, Chris; BIGHAM, Jeffrey. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. **ArXiv**, 1712.05796, 14 Dec. 2017. <https://arxiv.org/abs/1712.05796>
23. HATMAKER, Taylor. Facebook wants content reviewers back ASAP, slows return plan for most employees. **TechCrunch**, 16 Apr. 2020. Available at: <https://techcrunch.com/2020/04/16/facebook-content-moderators-wfh-when-will-facebook-employees-return/>
24. HERSHBEIN, Brad; KAHN, Lisa B. Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. **American Economic Review**, v. 108, n. 7, p. 1737-1772, 2018.
25. HOGAN, Bernie. From invisible algorithms to interactive affordances: data after the ideology of machine learning. *In: BERTINO, Elisa; MATEI, Sorin A. (eds). Roles, trust, and reputation in social media knowledge markets*. New York: Springer, 2014. p. 103-117.

26. HULL, Dana. Tesla engineer says Musk overstated the capability of autopilot. **Bloomberg**, 8 May 2021. Available at: <https://www.bloomberg.com/news/articles/2021-05-07/musk-s-autopilot-claims-don-t-match-reality-tesla-engineer-says>
27. HUWS, Ursula. **The making of a cybertariat**: virtual work in a real world. New York: Monthly Review Press, 2003.
28. ILO – INTERNATIONAL LABOUR ORGANIZATION. **The role of digital labour platforms in transforming the world of work**. Report. Geneva: International Labour Organization, 2021. Available at: https://www.ilo.org/global/research/global-reports/weso/2021/WCMS_771749/lang--en/index.htm
29. JAIMOVICH, Nir; SIU, Henry E. Job polarization and jobless recoveries. **Review of Economics and Statistics**, v. 102, n. 1, p. 129-147, 2020. https://doi.org/10.1162/rest_a_00875
30. KÄSSI, Otto; LEHDONVIRTA, Vili; STEPHANY, Fabian. How many online workers are there in the world? A data-driven assessment. **ArXiv**, 2103.12648, 2021. Available at: <https://arxiv.org/abs/2103.12648>
31. KJZZ PHOENIX. Tempe police release video in fatal Uber vehicle accident. **Youtube**, 21 Mar. 2018. Available at: <https://www.youtube.com/watch?v=pO9iRUx5wmM>
32. LEVANDOWSKI, Anthony. **Self-driving cars**. Distinguished Innovator Lecture. Berkeley: University of California, 12 Nov. 2013.
33. MICELI, Milagros; YANG, Tianling; NAUDTS, Laurens; SCHUESSLER, Martin; SERBANESCU, Diana; HANNA, Alex. Documenting computer vision datasets: an invitation to reflexive data practices. *In*: Conference on Fairness, Accountability, and Transparency (FAccT'21), 3-10 Mar. 2021, Virtual Event, Canada. **Proceedings [...]**. New York: ACM, 2021.
34. MORESCHI, Bruno; PEREIRA, Gabriel; COZMAN, Fabio G. The Brazilian workers in Amazon Mechanical Turk: dreams and realities of ghost workers. **Revista Contracampo**, v. 39, n. 1, p. 44-64, 2020. <https://doi.org/10.22409/contracampo.v39i1.38252>
35. MORTIMER, Thomas. Lectures on the Elements of commerce, politics, and finance. London: T.N. Longman & O. Rees, 1801.
36. OJANPERA, Sanna; GRAHAM, Mark; STRAUMANN, Ralph K.; SABBATA, Stefano De; ZOOK, Matthez. Engagement in the knowledge economy: regional patterns of content creation with a focus on Sub-Saharan Africa. **Information Technologies and International Development**, v. 13, p. 33-51, 2017.
37. RICARDO, David. **On the principles of political economy and taxation**. London: John Murray, 1821.
38. RIFKIN, Jeremy. **The end of work**: the decline of the global labor force and the dawn of the post-market. New York: Tarcher/Putnam, 1999.
39. ROBERTS, Sarah T. **Behind the screen**: content moderation in the shadows of social media. New Haven: Yale University Press, 2019.

40. ROCHWERGER, Alyssa S. Crowd survey reveals how Appen platform helps during global pandemic. **Appen**, 14 Apr. 2020. Available at: <https://appen.com/blog/crowd-survey-reveals-how-appen-platform-helps-during-global-pandemic/>
41. ROSENBLAT, Alex; STARK, Luke. Algorithmic labor and information asymmetries: a case study of Uber's drivers. **International Journal of Communication**, v. 10, p. 3758-3784, 2016.
42. SEAGER, Charlotte. After the robot revolution, what will be left for our children to do? **The Guardian**, 11 May 2016. Available at: <https://www.theguardian.com/careers/2016/may/11/robot-jobs-automated-work>
43. SRNICEK, Nick. **Platform capitalism**. Cambridge: Polity Press, 2017.
44. TERRANOVA, Tiziana. Free labor: producing culture for the digital economy. **Social Text**, v. 18, n. 2, p. 33-58, 2000. http://dx.doi.org/10.1215/01642472-18-2_63-33
45. TUBARO, Paola; CASILLI, Antonio A. Micro-work, artificial intelligence and the automotive industry. **Journal of Industrial and Business Economics**, v. 46, n. 3, p. 333-345, 2019.
46. TUBARO, Paola; CASILLI, Antonio A.; COVILLE, Marion. The trainer, the verifier, the imitator: three ways in which human platform workers support artificial intelligence. **Big Data & Society**, v. 7, n. 1, 2020. <https://doi.org/10.1177%2F2053951720919776>
47. UNITED STATES. District Court for the District of Massachusetts. **Gabriela Rojas-Lozano v. Google Inc.**, Cas 3:15-cv-10160-MGM, 2015.
48. URE, Andrew. **The philosophy of manufactures**. London: Charles Knight, 1835.
49. VON AHN, Luis; MAURER, Benjamin; MCMILLEN, Colin; ABRAHAM, David; BLUM, Manuel. ReCAPTCHA: human-based character recognition via web security measures. **Science**, v. 321, n. 5895, p. 1465-1468, 2008.
50. WILLIAMS, Oscar. Carl Benedikt Frey: Covid-19 will accelerate automation. **New Statesman**, 2 Sept. 2020. Available at: <https://www.newstatesman.com/spotlight/coronavirus/2020/09/carl-benedikt-frey-covid-19-will-accelerate-automation>
51. WOLF, Marty J.; MILLER, Keith W.; GRODZINSKY, Frances S. Why we should have seen that coming: comments on Microsoft's Tay "Experiment," and wider implications. **The ORBIT Journal**, v. 1, n. 2, p. 1-12, 2017. <http://dx.doi.org/10.1145/3144592.3144598>

Received: 17 May 2021.

Accepted: 09 July 2021.