8 The Planetary Stacking Order of Multilayered Crowd-AI Systems

Florian A. Schmidt

What is the impact of the high demand for artificial intelligence (AI) training data for autonomous vehicles on the working conditions of crowdworkers? This chapter will put an emphasis on the planetary dimensions of this particular outsourcing stack, its structural aspects, its layered and siloed qualities, its fractal and redundant features, as well as its fragilities and uncertainties for the different stakeholders.

The analysis focuses on the intersection of four interdependent developments that culminated in 2018: First of all, there was the race of several dozen very well financed automotive and technology companies to be the first to bring self-driving cars onto the streets. Second came the ensuing unprecedented demand for vast amounts of highly accurate training data necessary to teach the cars how to navigate traffic based on vision. Third, there was the restructuring of large crowdsourcing platforms in order to cater to the specific needs of the automotive industry and orchestrate the required workforce for them. And finally came the crash of the Venezuelan economy, because of which Venezuelans inadvertently ended up providing the brunt of the work for this mammoth task of manual image labeling in the service of AI development.

The main argument of this chapter is that this interplay of planetary economic forces and events that unfolded in 2018—their synchronicity—was not merely a freak occurrence but instead refined and laid bare the mechanics of how capital can conjure up, train, and dismiss hundreds of thousands or even millions of digital workers as needed—much more rapidly than would be possible with production sites bound to a physical location.

As the chapter will show, the long-existing infrastructure of crowdsourcing platforms underwent a significant update through this succession of events and has since become much more capable through its adaptation to very demanding and deep-pocketed clients from the automotive industry. Future spikes in demand for digital labor may have nothing to do with self-driving cars, and next time it probably will not be Venezuela that happens to have the cheapest workforce on offer globally. But

the development of the structures described here demonstrates how to match excess capital with an oversupply of labor just in time, to high standards, and without having to commit to brick-and-mortar factories and local workers on the ground—let alone national labor laws and regulations. Thus, labor has become *almost* as liquid as capital. I write "almost" because, as I will also show, the same economic desperation that can make labor matchlessly cheap and flexible on the planetary digital market can also eventually inhibit the free flow of labor by preventing access to said market through crumbling infrastructure, blackouts, and embargos.

First, I will briefly show why the push for full automation, self-driving cars, and self-learning algorithms has somewhat counterintuitively led to a surge in the demand for manual labor, mostly in the form of crowdsourced image annotation. Then I will explain how the degree of accuracy necessary for this type of work has forced the crowdsourcing industry to fundamentally restructure its processes—from a very direct, or flat, model that provided clients with little more than direct access to a crowd available for completing general tasks to a model in which the platform orchestrates every detail of a highly complex and multilayered process while the client buys the end results without ever coming into contact with the crowd. Here, the focus is on making visible how this shift from *general-purpose crowdwork* to what I will call a *specialized, full-service crowd-AI stack* has wide implications for the legal classification of workers as independent contractors and for the working conditions on these platforms.

This new generation of platforms is experimenting with various stacking orders of human labor—AI support and control systems, subautomation, and suboutsourcing. The crowdworkers both train AI systems and are trained by AI systems, with humans and machines working together in ever-more-complex structures. The chapter will conclude with a description of the turbulent working conditions that Venezuelan crowdworkers had to endure in 2018 and 2019 in order to show that high demand for digital labor and a well-designed digital workplace are not enough to ensure a reliable work environment. Unfortunately, it appears that the only constant in this highly dynamic planetary outsourcing stack is the continued precarity of the working conditions.

Speaking of the planetary dimensions of this phenomenon, however, it must be mentioned that this chapter is only able to look at publicly crowdsourced labor. This is the more visible side of training data production commissioned by clients based in the Global North, done mostly by workers in the Global South, especially in Venezuela now. The Chinese stack for the production of training data unfortunately remains entirely in the dark, as does any work not done on publicly advertised platforms that are open to crowdworkers. A clear view of the developments sketched out in this chapter is further obstructed by the secrecy of the industry, which is concerned about preserving both its

technological trade secrets and its competitive advantage gained through circumventing national labor laws and regulations via global outsourcing.

The research in this chapter is based on interviews conducted in 2018 and 2019 with CEOs of training data producers (Mighty AI, Hive, understand.ai, Clickworker, Playment, and Crowd Guru) and crowdworkers (mostly from Mighty AI and mainly from Venezuela).[1] Other crucial sources have been my direct observation of the platforms over the years; the platforms' communication with workers and clients through forums, advertising, and press releases; and experience of using their annotation tools, partly by logging on as a crowdworker myself (Schmidt 2016, 2019). From the photos I annotated in that role in 2018, it became apparent that the German car industry was among the most important clients of the platforms at that time. But client companies could not be confirmed officially due to the nondisclosure agreements preventing the platforms from disclosing the names of their clients.

Why Manually Annotated Images Continue to Be in Such High Demand

The question of whether the ambitious goal of bringing fully autonomous vehicles onto regular urban streets will be achieved in the foreseeable future remains open. Although impressive progress has been made, various difficult problems have emerged that have dampened the optimism of the engineers working on this challenge. The euphoria over the promise of *fully* automated vehicles (defined by the Society of Automotive Engineers as SAE Level 5) may have been overly optimistic, but semiautonomous vehicles (SAE Level 4) too need millions of labeled images as ground-truth data to model and predict what is happening around them in traffic. By 2018, 55 companies had secured licenses to test autonomous vehicles in California, while others were operating cars on the streets of Arizona (DMV CA, Department of Motor Vehicles California n.d.).

Traditional auto manufacturers' fear of being left on the sidelines by Google's Waymo or by Tesla has triggered what might turn out to be an investment bubble or at least a risky gamble on whether this technology will ever actually become viable. Billions of dollars keep being invested in the development of autonomous vehicles, and hundreds of millions of those dollars trickle down the supply chain into the crowdsourced production of AI training data. Market leader Waymo was able to raise $3 billion in two venture capital rounds within just a month in 2020 (Etherington 2020). Volkswagen, having already invested $2.6 billion in the autonomous vehicle company Argo AI in June 2020, announced the expenditure of another $91 billion on the development of electric autonomous vehicles in its new unit the Volkswagen Autonomy GmbH through 2025. In the meantime, Scale AI became the first training-data provider to reach a valuation

of over $1 billion in August 2019, turning it into a so-called unicorn. In April 2021, it reached a valuation of $7.3 billion, after it had raised a total of $600 million in venture capital (Kahn 2021).

Not all production of training data revolves around image annotation for autonomous vehicles, but at least for the time being this is where the money is for the crowdsourcing platforms. Typically, the images are stills taken from videos shot in traffic, which are then manually annotated so that a machine can recognize every object within the frame. Humans have to, for example, draw two-dimensional bounding boxes around cars, draw three-dimensional cuboids to orient cars in space, or assign descriptive labels such as "truck," "tree," "school bus," "bicycle," and so forth to every pixel in a video frame. This so-called full semantic segmentation is currently the most common and most time-consuming of the various forms of image labeling: unassisted by automation, a worker may take up to two hours to complete a full semantic segmentation of a traffic scene.

At first glance, one is tempted to assume that the manual annotation of training data is a transitory phenomenon, a job that will soon be either finished or entirely automated. How hard can it be? But one crucial factor behind the sheer amount of work is that it is being done in a highly redundant fashion. The various clients need very similar sets of annotated images; it would be much more efficient for the automotive companies to draw their training data from a collectively produced pool of annotated images (different stakeholders now increasingly pay lip service to this idea). However, because of the competitiveness of the market, such sharing currently seems out of the question in most cases.

One interrelated consequence that has not yet been taken sufficiently into consideration in the debate around the public safety of self-driving vehicles is that by training cars based on these siloed datasets and opaque procedures, the manufacturers will end up with vehicles that react differently in extreme situations, or so-called edge cases. From the perspective of other participants in traffic, or regulators, for that matter, such an outcome cannot be tolerated. It therefore seems plausible, though speculative at this point, that manufacturers will eventually be obliged to all use the same training datasets or even the same algorithmic models as a common standard for the critical decisions that their vehicles constantly have to make in traffic. If this is going to be the case, it truly becomes a winner-take-all game for the car manufacturers to compete to be the one to set this standard. The fact that the quality of these algorithmic models is highly dependent on the quality of the training data gives all the more reason why so much money is being invested in the type of outsourcing system described here.

Moreover, while the training data itself could to some extent indeed be shared, the more important market, according to the CEOs interviewed for this study, is or will be

the one for validation data, in which human cognition is needed to evaluate retroactively the decisions that the autonomous systems have already made in traffic. This, of course, is highly case-specific, legally delicate, and confidential information—data that cannot be shared with competitors.

Last but not least, human labor continues to be in high demand because of how messy reality turns out to actually be. The systems need to be constantly retrained on ever-evolving traffic scenarios. One example would be the sudden large-scale rollout of commercial e-scooters (driven, just like the development of autonomous vehicles, by venture capital speculation) in 2017 (Hawkins 2018). Not only did this rollout lead to emergent behavior patterns in traffic—new silhouettes crossing the streets at unexpected speeds—but the scooters also fall under equally emergent legal restrictions, which are different and subject to change in every European metropolis.

Another vivid edge case example that illustrates geographic variety is that autonomous vehicles have difficulties processing jumping kangaroos (Ackerman 2017). As Mark Mengler, CEO of understand.ai, described in an interview for this study, it is quite likely that autonomous vehicles will have to be geofenced, meaning that they will not be able to cross invisible digital boundaries, such as between countries. Before being allowed to cross into another geographic zone, they will have to download extensive updates. Since the peculiarities even of neighboring countries are so data-intensive, it would not be possible or practical to have them all preinstalled onboard. As the next section shows, more and more annotation tasks can be done automatically, but it seems likely that the machines will have to be continuously trained by humans in the loop, especially for new tasks, new edge cases, and retroactive data validation.

From Legacy, General-Purpose Crowdwork to Specialized Full-Service Crowd-AI Stacks

Even before the first pedestrian was killed by a self-driving car in 2018, it was clear that in contrast to accuracy in annotating images of cats or food via crowdsourcing, accuracy in the production of ground-truth data for autonomous vehicles is a matter of life and death.[2] The clients buying this data demand an accuracy of at least 99 percent, to be guaranteed by the producer of the training data; the need to match that goal is the single most important reason for a cascade of structural shifts in how crowdsourcing platforms operate. It has led both to the emergence of new and specialized full-service producers of training data, organized as layered crowd-AI stacks designed to cater exclusively to clients at the intersection of the automotive industry and AI research, and to the transformation of older, hitherto flatter general-purpose crowdsourcing platforms toward that end.

For the purposes of contrast, it is best to look back briefly at Amazon Mechanical Turk (MTurk), the oldest and most prototypical crowdsourcing platform for microtasks. Founded in 2005, with a workforce of about half a million people, mostly from India and the US, MTurk is the prime example of what new competitors such as Playment (founded in 2015 in Bangalore, with around 300,000 workers) have started to refer to as "legacy crowdsourcing platforms" (Magistretti 2017). MTurk is a generalist platform with an application programming interface that allows paying customers to give "human intelligence tasks" directly to a distributed crowd by addressing them like a general-purpose computer. It is flat in the sense that the client gets its results directly from the crowd, without any hidden processing or quality management layers.

While Amazon does exert some influence on the organization of the work, it frames itself as a marketplace, or infrastructure provider, that can be held accountable for neither the quality of the results nor the working conditions on the platform. The client customers of MTurk select, train, and pay the individual crowdworkers; they are also responsible for the description of tasks, the development of specialized tools, and the quality control of the results. In this legacy form of crowdwork, the platform manages to externalize most risks and responsibilities to the other two parties involved—that is, the clients and the workers.

The clients from the automotive industry reversed this logic and externalized the burden (and potentially the legal liabilities) of quality control to the training data providers. Because the new specialist companies that have emerged maintain an internal crowdsourcing platform, they can now offer fully managed data labeling or end-to-end project management. Nothing is done by the client, they promise.

The following are the most prominent examples among a dozen of these companies (data regarding investments via https://www.crunchbase.com/). The Seattle-based Mighty AI, founded as Spare5 (app.spare5.com) in 2014 and backed by $27.3 million in venture capital, was acquired by Uber for an undisclosed sum in 2019. In 2018, it had a workforce of about half a million people, three quarters of whom lived in Venezuela at the time. The San Francisco–based company Hive (thehive.ai, hivemicro.com), founded in 2013 and backed by $20 million in venture capital, had a workforce of over a million people from supposedly 150 different nations (though between 60 and 75 percent lived in Venezuela in 2020; more on the origin of the workers below). The above-mentioned Scale AI (scale.com), founded in San Francisco in 2016 and backed by $118 million in venture capital (a lot, by comparison), had a workforce of 30,000 people in 2018 (small, by comparison).

These training data producers operate as multilayered, full-service black boxes without direct contact between their clients and the crowdworkers (see also Tubaro and Casilli 2019). This new and much deeper structure of partly hidden image-processing

layers involving both human labor and AI automation constitutes a consequential departure from the MTurk model of legacy crowdsourcing for all three parties involved. The clients no longer have to develop their own tools for data annotation, train the crowd, or sort and evaluate the results provided. For this convenience, they have to pay substantially higher prices (which is why some experienced clients continue to use MTurk and do the quality control themselves).

For the workers, this shift means that they have to learn new software tools less often; that the tools are more reliable, more convenient to use, and constantly developed as proprietary assets of the platform; and also that the task descriptions are less faulty and easier to understand. There is a lock-in effect here: switching platforms means not only losing one's reputation and qualifications but also the accumulated skill of handling proprietary tools, at least to some extent (after all, the tasks are very similar). Still, for the workers, it is much safer and more reliable to deal only with platform providers instead of having to adapt to ever-changing clients—especially when it comes to getting paid reliably at the end of each week instead of having to fear late payments, disputes about inferior results due to faulty tools and descriptions, or even wage theft (all of which have been grievances for workers on MTurk for years).

For the platform providers, however, the shift to full-service solutions not only means much higher investments, it also entails the looming legal risk of getting sued for the potential misclassification of their workers as independent contractors rather than employees of the platform, as happened to CrowdFlower (*Otey v. CrowdFlower*, class action lawsuit filed in 2012, settled in 2015).[3] More rudimentary legacy platforms like MTurk, which is notoriously unresponsive to the grievances of its workforce, can very plausibly argue that their workers are merely freelancers or hobbyists whom the platform helps by connecting them to external clients via a marketplace. But full-service stacks, in which workers have no direct contact with clients, will find it much more difficult to defend the current classification of their workers as independent contractors. Although the new system is in certain ways more reliable for the workers, the entire business model is built on shaky ground as regards their legal status. In the following section, I will show in what other ways these full-service crowd-AI stacks have changed how work is being orchestrated, and how this is the result of having to deliver high volumes of training data with guaranteed high accuracy.

Platforms Investing in Tailor-Made AI Tools and Handpicked Crowds

Scale AI used to offer a price calculator on its website. In mid-2018, the price for just nine annotations in a single image was $1.00—provided the client bought 10,000 images.

The price for the much more complex semantic segmentation of an entire image, with a guaranteed accuracy of 99.2 percent, was $6.40; if the client opted for "express urgency," the same service cost $16.00 per image. Given how many well-funded automotive companies desperately need millions of this type of annotated images, it is not surprising that training data providers have become so attractive to venture capital.

An individual human without AI assistance would need up to two hours to process an image that costs $6.40 retail. Thus, the platforms could not be profitable if they paid their workers a minimum wage at the standard of Western industrial nations. To be able to deliver high-quality service at speed, volume, and a competitive price, they have therefore developed a number of instruments and strategies:

1. Invest in custom-made, AI-enhanced, semiautomatic production tools that heuristically do the semantic segmentation in advance and then guide the attention of the crowdworkers to areas where the system is less certain about what it has recognized in an image.
2. Invest in quality management through process optimization regarding how granular the jobs become when split up into microtasks and how they are best reassembled afterward, with successive layers of quality control (monitoring both the workforce itself and the results) done alternately by humans and AI.
3. Invest at the same time in automated training of the workforce and gamification mechanisms to incentivize the ambition, focus, and skills of workers as well as in labor-intensive human community management.
4. Invest in access to a cheaper workforce either by offering task descriptions, automated training tutorials, and human community management in languages spoken in low-wage regions of the planetary market for digital labor; or by suboutsourcing part of the labor to business process outsourcing (BPO) firms that have a local workforce on site in developing countries and that can make the work accessible to people with language skills not supported at scale by the platform itself.

Together, these decisions on investments shape what I think is best described as a crowd-AI stack, a complex proprietary solution with alternating layers of human management, algorithmic management, AI automation, and manual human labor. The competing producers of training data try to gain a competitive edge by favoring different stacking orders of these data annotation and quality control layers. While some, like understand.ai, a training data company based in Karlsruhe, Germany, are investing heavily in new AI tools in order to reduce the number of workers necessary; others, like Playment in Bangalore, prioritize access to a cheap workforce and, as the company name suggests, gamification mechanisms to keep that workforce focused and motivated. Both

use workers from India, but for understand.ai this is a BPO layer alternating with local student temp workers in Germany and involving as much automation as possible.

The fact that the producers of training data do not merely supply AI development services but are also developing and employing their own AI technology to streamline processes leads to an interesting paradox: the growth of AI increases both the demand for manual labor and the demand for AI automation. The goal of the hard-to-automate manual labor is the training of AI models, while, at the same time, AI automation is used to train and support the manual labor that makes it more reliable and cost efficient. In short, humans and AI train each other.

Although they are reliant on crowdworkers, the training data providers market themselves as AI companies, while the term *crowd* is pushed into the background—possibly due to its negative connotations of cheap labor and low-quality results or maybe simply because it is yesterday's buzzword. This development is reflected in the fact that some of the specialist platforms have a client-facing company name, website, and appearance that emphasizes AI, along with an entirely different crowd-facing name, platform, and appearance promising prospective workers that they can make money quickly through microtasks. Mighty AI's crowd platform is Spare5, Hive's platform is Hive Work (the worker platform of thehive.ai is also known as Hive Micro), and Scale AI's platform is Remotasks.

But even with this two-faced approach, the new specialists must convey to their clients a seemingly contradictory message that advertises both a high degree of automation from quasimagical AI as well as human precision from well-trained and handpicked crowdworkers. On their respective websites, these specialist platforms advertise the shift away from rudimentary crowdsourcing as "trained crowd labor," "known crowds," "curated crowds," or "crowd qualification." The message for clients is that the work is given not to a random, anonymous, and potentially incompetent mass but to handpicked groups of experts who are trained and monitored constantly.

Workers on the new platforms must go through a longer training phase, in which they specialize in certain types of tasks and, as in a computer game, level up to qualify for more sophisticated, better-paying tasks. This specialization, however, negatively affects which tasks they can see in the future (a serious source of stress for the workers, discussed below). The accuracy of the individual workers is tracked constantly, and they get qualitative and quantitative feedback on how well they do, partly automatically and partly from management.

As Daryn Nakhuda, cofounder and former CEO of Mighty AI, explained to me, the platform funnels incoming tasks in bulk to preselected subgroups to train them more quickly and efficiently. Thus, the workers are not an open and unstructured crowd

that self-selects incoming tasks freely. Instead, they are subject to various degrees of hierarchy, specialization, and orchestration conducted by the platform providers. For example, normal workers are called "Fives," but there are also "SuperFives" with access to better-paying tasks, more direct rapport with community management, and the privilege to beta-test new tools early.

If a worker gains access to tasks, the payment is generally more reliable than on legacy crowdwork platforms, though not necessarily higher. In interviews I conducted in 2018, the workers on Spare5 felt much better treated than on legacy crowdwork platforms, mostly because they had reliable human interaction with the platform staff in the form of good community management and direct, immediate, and personal responses to their questions. They were proud to be part of the company. This positive view changed, however, after Uber acquired the platform in 2019.

Fluctuations in Labor Demand and Migrant Crowdworkers

A crucial function of crowdwork is to provide employers with a buffer for rapid fluctuations in labor demand. In this, the platforms resemble temporary staffing agencies, only the frequency and volume with which they can mobilize and dismiss workers is orders of magnitude greater. For the car companies, which need waves of data annotation labor rather than a constant amount, it would, in most cases, make no sense to build up a workforce in-house. By serving multiple clients, the platforms should theoretically be able to level out the waves of demand into a constant stream, but this is often not the case. And even though the demand for manual image annotation has risen substantially since 2017, we still see a constant oversupply of labor (as became apparent in the various interviews I conducted with workers in 2018 and 2019; see also below). The platforms discussed here attract, train, and put on hold—in other words produce—far more workers than they regularly need, and yet this oversupply of labor is necessary to swiftly cope with peaks in demand. It is not a bug but a feature. The problem is well known in other areas of the gig economy, where it is managed through strategies like Uber's "surge pricing," a technique that Hive and others have started to use as well.

The oversupply of labor is a constant source of stress for the workforce. "Why don't I see any tasks?" was the most common concern in forums and conversations among workers in 2018. For the workers, it is especially irritating when colleagues are shown tasks when they are not, and the lack of transparency regarding the distribution of tasks leads to a lot of second-guessing about potential correlations with one's work history, performance, levels of accuracy, or experience points. What is causing the stress is not only the volatility of the total amount of work available, but that the workers, being

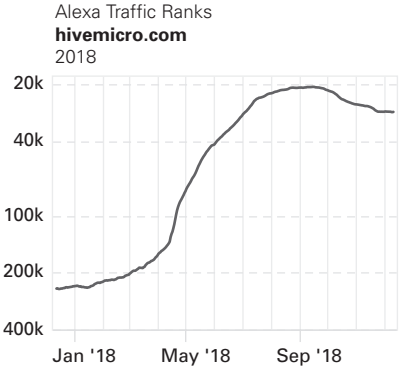
very well connected via external forums and social media, do know of collogues who have access to tasks while they themselves don't. The workers do not know whether they are not receiving tasks due to a decision by human management or by algorithm.

Most importantly, the constant oversupply of labor (see figure 8.1) erodes any negotiating power for better pay because for every task there is already a long line of people willing to do it for less money. On top of this, the oversupply leads to an indirect deterioration of average hourly earnings due to the unpaid downtime of waiting and searching for tasks across platforms—time spent desperately hitting the refresh button in the hope for more work. As in the rest of the gig economy, there is little flexibility or autonomy left if you have to jump at any opportunity to do a gig or tasks before someone else scores, and this competition is much more extreme if the work is not location-based, like food delivery or chauffeur rides, but can be done from anywhere in the world.

Where the producers of training data operate in the form of BPO firms, we can still see a slower, more conventional model of taking advantage of different prices for labor on a global market because the employers open shop locally (e.g., in Kenya or Nepal), where labor happens to be cheap. For crowdsourced microtasks, however, the disenfranchised workforce has begun to virtually roam the globe in search of tasks—almost like migrant agricultural workers.

What we see unfolding in the realm of manual data annotation is a truly planetary market for digital labor in which the tasks dynamically flow to those people who are willing to accept the lowest remuneration at any given point in time. As in a system of communicating vessels, the average hourly wage paid out by the platforms for relatively unskilled labor seems to level out globally. In 2018, experienced data annotators earned between $1 and $2 per hour (an estimate based on my interviews with CEOs and crowdworkers, and on forum debates among workers). As this is piecework, inexperienced workers of course earn much less. A large quantitative study from 2018 found that workers on MTurk, too, earned about $2 per hour on average (Hara et al. 2018).

As explained above, Hive, Mighty AI, and Scale AI have separate websites for their clients and their workforce; because the workers enter the virtual factory through a separate entrance, it is easy to follow their ebbs and flows, and also their countries of origin, simply by using Amazon's web traffic analysis tool, found at alexa.com/siteinfo. That is, one can use the web traffic of the crowdwork platform as an approximate measure of worker supply. I followed this data in 2018 and 2019, triangulating it with various interviews with crowdworkers. The two most important findings were that, in those two years, several hundred thousand workers were moving back and forth between platforms (especially between Mighty AI and Hive) desperately looking for tasks—and that up to three-quarters of them were based in Venezuela.

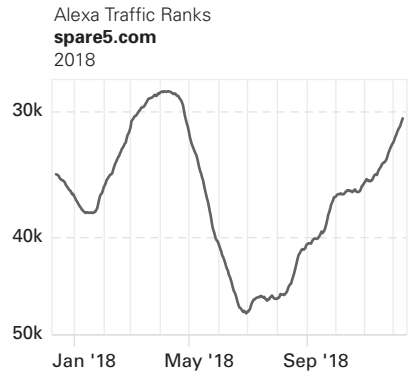


Global Rank: 27.941
Rank in Venezuela: 187

Visitors by Country:

Venezuela	75.9%
India	6.0%
Vietnam	1.8%
USA	1.7%
Philippines	1.6%

Data retrieved on 12 December 2018 via https://www.alexa.com/siteinfo/hivemicro.com

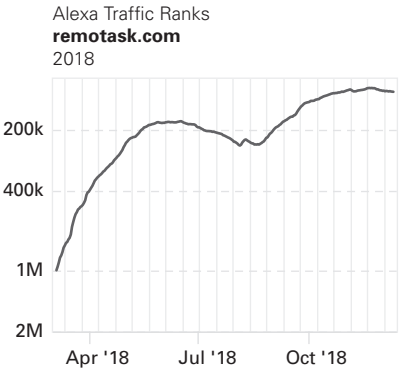


Global Rank: 30.531
Rank in Venezuela: 162

Visitors by Country:

Venezuela	76.2%
Colombia	4.1%
Mexico	2.5%
India	2.1%
Argentina	1.4%

Data retrieved on 12 December 2018 via https://www.alexa.com/siteinfo/spare5.com



Global Rank: 128.938
Rank in the Philippines: 2.129

Visitors by Country:

Philippines	31.1%
Venezuela	15.0%
Kenya	10.5%
Nigeria	9.2%
India	7.2%

Data retrieved on 12 December 2018 via https://www.alexa.com/siteinfo/remotasks.com

Figure 8.1
Graphs showing the fluctuations in web traffic on the crowdwork platforms belonging to Hive, Mighty AI, and Scale AI in 2018.
Source: alexa.com/siteinfo.

At the end of 2018, Hive Work was ranked #187 of the most frequented websites in Venezuela (up to #123 in October 2020), and Spare5 (Mighty AI's work platform) was ranked #162 (down to #333 in October 2020). The country supplied around 75 percent of the workforce of Spare5 throughout 2018. At Hive Work, this percentage rose from 55 to 75 percent over the course of 2018 (and back down to 55 percent in late 2020). As I learned in my interview with Kevin Guo, CEO of Hive, the first half of 2018 saw the number of workers on his platform grow from 100,000 to 300,000, at a speed of up to 3,000 new registrations per day (at the end of 2020 it stood at more than a million workers). Considering the parallel rapid growth on Spare5, at least 200,000 people from Venezuela must have been searching for work at the time on these two platforms alone. In the first half of 2018, Scale's Remotasks predominantly attracted English-speaking workers from the Philippines and India, with 31 percent of the workers coming from each (at the end of 2020, its workers mainly came from Venezuela, the Philippines, and Kenya, which provided 36, 22, and 11 percent of the workforce, respectively). Playment drew its workforce of 300,000 people entirely from India, and continues to be entirely focused on this country.

For the image annotation platforms to stay competitive, it matters where the workers live, in regard to how expensive the labor is, but for older companies, it is not easy to adapt. Crowd Guru, founded 2008 in Berlin, is a platform with an almost entirely German-speaking workforce. When it comes to microtasks that can be done by anyone anywhere, it is impossible for a company like Crowd Guru to compete with platforms that can mobilize workers from the Global South for a tenth of the cost. In 2018, the CEO of Crowd Guru, Hans Speidel, told to me in an interview for the original report (Schmidt 2019) that from time to time he was approached by small BPO firms who offered him their workforces as a form of suboutsourcing. However, as Speidel explained, the inclusion of a cheaper external workforce would alienate or antagonize the German workforce, which had developed over time a sense of community important for social cohesion and loyalty to the platform.

The situation is completely different for platforms like Appen (appen.com), founded 1998 in Australia and with a current workforce of over a million, and Clickworker (clickworker.com), founded 2005 in Germany and with a current workforce of over two million, who produce a large quantity of training data for voice assistants. Since this type of work entails teaching the machines to understand various dialects, a culturally diverse workforce is key. Most audio tasks are distributed to regions where consumers have great purchasing power. It can be lucrative if the AI voice assistant understands, say, a thick Bavarian accent to ensure a smooth purchase via voice control. These tasks are much better paid, but because they are spread widely among people with various dialects, the demand is too sporadic for those individuals to make a living from it. Here,

we indeed find the hobbyist crowdworkers who only work on occasion without getting exploited.

Learning from the Venezuelan Workers

In 2018, Venezuela inadvertently became the key supplier of cheap digital labor in the field of data annotation for the automotive industry. For this it was advantageous that the country has a well-connected, well-educated middle class and a reasonably good technical infrastructure—remnants of more fortunate times. The socioeconomic collapse of Venezuela had been a decade in the making, but hyperinflation and mass starvation became particularly acute between 2017 and 2019. During this period of collective hardship, crowdwork became a lifeline for hundreds of thousands of Venezuelans. It is not that the crowdsourcing platforms discussed in this chapter deliberately decided to go to Venezuela. They merely decided to offer the work in Spanish, and Venezuelans, in search of a means to make money online, found the work through Google, YouTube, Reddit forums, and word of mouth.

The average hourly earnings of $1.50–$2.00 paid out by Spare5 at the time were perceived as a pittance by crowdworkers from the Global North, but for the Venezuelan workers, such an income meant the difference between starving and being able to provide for a family. Those who worked for Spare5 were perceived by their neighbors as relatively affluent and quickly recruited more friends and family. One Venezuelan engineering student explained to me that some of his colleagues wanted to keep "the goose that lays the golden eggs" a secret but that he felt morally obliged to tell many others.

The Venezuelan workers from Spare5 that I interviewed in 2018 were very satisfied with the platform, especially in comparison with other providers of microtasking crowdwork. They felt they were dealing with a trustworthy company that treated them with respect. They valued the user-friendly interface and tools and most importantly the reliable weekly payments via PayPal. They experienced the work as intrinsically rewarding, were proud of the quantified and gamified feedback they received in the form of experience points and special ranks, and were proud to be part of the company. This is important because it shows that doing this type of work can actually be quite a positive experience in and of itself—if it weren't for the extreme precarity.

As explained above, the unpredictable availability of tasks had been a systemic source of stress for workers within the Spare5 platform. However, their uncertainty reached new heights in June 2019, when Uber ATG (Advanced Technology Group, Uber's branch specialized on autonomous vehicles) acquired Mighty AI and shortly thereafter informed the Venezuelan workers, via a popup window on Spare5, that their lifeline had been suspended. As Andrea, a longtime SuperFive—one of the few hundred most productive

and accurate workers—and hitherto always in close contact with management, recalled: "When we realized that us losing access was not a technical error, but an intentional action to leave us on the sidelines for an indefinite time, it was as if the floor under our feet disappeared." The SuperFives in particular, who strongly identified with their special role and with the company, were deeply disappointed by how they were treated.

It seems that Uber's sudden exclusion of Venezuelan workers was due to US sanctions against Maduro. Around the same time, Adobe discontinued access to its cloud service for Venezuelan users, directly referencing compliance with an executive order of the US against Venezuela as the reason (Lee 2019). The Spare5 workers eventually got paid three weeks later, after having to sign a statement confirming that they were not affiliated with the government. Finally, Uber also restored access to their workplace while, curiously, applying a new geoblocking strategy—now excluding workers from regions such as Europe. By October 2020, as inferred from the Alexa web traffic tool, the share of Venezuelan workers on Spare5 had risen to over 88 percent (with India following at 4.4 percent), but the overall volume of traffic had dropped.

Before the Uber acquisition, Mighty AI was an outsourcing service popular with various automotive companies, among them some large German car manufacturers. Afterward, Uber discontinued the client-facing brand, Mighty AI, while its crowd platform, Spare5, now serves only its internal program of developing autonomous vehicles, which of course makes the pool of tasks available to the workers much smaller.[4]

The Venezuelan workers have developed several strategies to survive the extreme volatility of their digital workplace. In addition to switching back and forth between various platforms, they are also renting out valuable, leveled-up accounts with access to good tasks to others while they are not using them. They are doing this either locally with friends and neighbors or via members-only Discord groups where renters have to pay a monthly fee to get access. As a result, and ironically, the far ends of the tenuous suboutsourcing stack are effectively unknowable to each other. Just as the workers can never be quite sure which car company currently employs them, the employers cannot be sure who is actually doing the work.

While the Venezuelan workers are hyperconnected via Slack and Discord, they are often physically stuck in abject poverty with outdated computing equipment. Their livelihood is threatened by blackouts, corruption, organized crime, and food shortages on their side of the screen and by the capriciousness of algorithmic management, venture capital, and political sanctions on the other.

Douglas, a 21-year-old engineering student from Venezuela, explained: "The situation is better than five years ago, mainly because—this may sound crazy to you—most of the criminals have fled the country. The crime rate is still really high, but it is more secure now to go outside without getting robbed. But there is not much for me out

there anyway, because I do almost everything here on the computer." However, staying at home did not protect Douglas from being robbed: "During one of the blackouts, people climbed into my courtyard, where, right under my window I keep some livestock for extra food. They took a few chickens and climbed back. Behind my house, there is a kind of wasteland with improvised settlements, and from there people must have observed that I have chickens here."

The example of Venezuela might seem extreme, a freak occurrence in how its crash coincided with the hype around autonomous vehicles. But then again, as an ever-larger percentage of the world's population goes online while at the same time geopolitical constellations are becoming less stable, many other countries could potentially be the next Venezuela for digital labor in the years to come.

Conclusion

Returning to the initial question of this chapter, what is the impact of the high demand for AI training data for autonomous vehicles on the working conditions of crowdworkers?

On the one hand, this demand has improved the working conditions of the crowd significantly, at least in some respects, over those of flat, general-purpose, legacy crowdwork platforms (like MTurk). The workers deep within the training-data crowd-AI stack can follow a career path by gaining specialized skills; their progress within the system is tracked and documented, and they can reach more senior roles and enjoy a sense of mastery over professional tools and skills. More experienced workers benefit from training and become less exchangeable and more valuable to their employers, and the workers welcome constructive feedback and support from a responsive and human community management that treats them with respect and replies to their concerns quickly. Most importantly, the workers can rely on getting paid weekly by the platforms and do not have to deal with unpredictable payment practices of ever-changing clients treating them merely as interchangeable subhuman machine parts.

On the other hand, even though in the best-case scenario outlined here, the work experience has improved as a result of the shift toward specialized AI training-data platforms, a set of interconnected, fundamental, and potentially unsolvable problems remain for workers in the global crowdwork market: the race to the bottom in terms of wages and, maybe even worse, the constant insecurity or precarity regarding the question of whether there will be enough work the next day. What the labor is worth in monetary terms is negatively affected not only by the drive toward ever more automation but also by the fact that the work can be funneled almost instantaneously to an even cheaper workforce across the globe. Even if the platforms do their best to design a

virtual workplace that treats the individual crowdworkers with respect and pays them reliably for their work, there is no guarantee that tomorrow the tasks a worker has just specialized in will not be automated or performed by someone even more desperate.

Probably the most important lesson from studying the crowdsourced production of AI training data is that in the relatively short time between mid-2017 and the end of 2018, the automotive industry, through a supply chain of venture capital–funded platforms, was able to access—or rather create—hundreds of thousands of workers. In other words, capital conjured up a massive, globally distributed workforce almost overnight.

A crowd is a mass phenomenon in which the individual human is by definition replaceable. The crowdworkers become the equivalent of the seemingly endless mass of repetitive and interchangeable microtasks. As long as it is easy to produce workers in an instant—workers who recruit themselves and can be trained and managed automatically—the workforce has hardly any negotiating power to improve its wages or working conditions. That is to say, the problem of the crowd remains the crowd.

Finally, I want to point to the fascinating self-similarity (see figure 8.2) between the layered structure of the crowd-AI stacks described here and the layered structure of

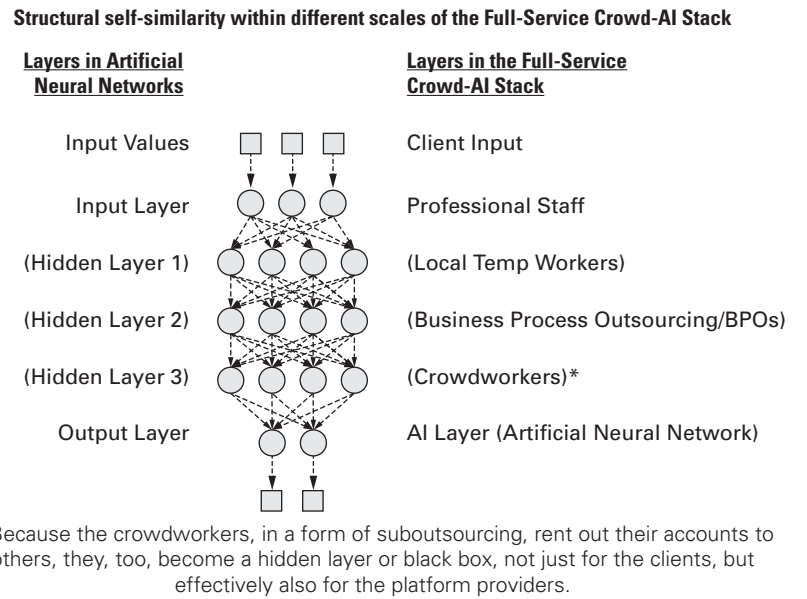


Figure 8.2
Structural self-similarity within different scales of the image processing stack.
Source: Author.

the artificial neural networks for image recognition used within those larger planetary outsourcing structures. On both the micro and the macro level, these systems are characterized by hidden processing layers—black boxes within black boxes. Although the accuracy of the end results is being guaranteed, nobody really knows in its entirety how these results, for which the crowds are providing training data, come about.

Notes

1. The project was funded by the German Hans Böckler Foundation. References within this chapter to interviews with platform CEOs are based on work published in the longer German version of the report (Schmidt 2019).

2. In June 2016, the driver of a Tesla died in a car crash in Florida while using the "autopilot" mode of the vehicle. The first pedestrian fatality happened in March 2018, when a woman was killed by an Uber test vehicle in Arizona (Schmelzer 2019).

3. "In 2014, workers sued CrowdFlower under the Fair Labor Standards Act ('FLSA') and Oregon's minimum wage law for failure to pay adequate wages. In response, CrowdFlower argued that the microtask workers were independent contractors, based on the terms in the EULA, and thus the minimum wage laws would not apply. The employee status question, however, was foreclosed by settlement before it could be decided by the court" (Cherry 2017, 1823).

4. Remarkably, Uber sold its ATG branch to the start-up Aurora in December 2020, apparently either externalizing or potentially giving up entirely on its endeavor to produce self-driving cars.

References

Ackerman, Evan. 2017. "Autonomous Vehicles vs. Kangaroos: The Long Furry Tail of Unlikely Events." *IEE Spectrum*, July 5. https://spectrum.ieee.org/cars-that-think/transportation/self-driving/autonomous-cars-vs-kangaroos-the-long-furry-tail-of-unlikely-events.amp.html.

Cherry, Miriam A. 2017. "The Sharing Economy and the Edges of Contract Law: Comparing U.S. and U.K. Approaches." *George Washington Law Review* 85 (6): 1804–1845. https://www.gwlr.org/wp-content/uploads/2018/03/85-Geo.-Wash.-L.-Rev.-1804.pdf.

DMV CA (Department of Motor Vehicles, State of California). n.d. "Autonomous Vehicle Testing Permit Holders." Accessed September 6, 2021. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-testing-permit-holders/.

Etherington, Darell. 2020. "Waymo Expands First External Investment Round to $3 Billion." *TechCrunch*, May 12. https://techcrunch.com/2020/05/12/waymo-expands-first-external-investment-round-to-3-billion/.

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P. Bigham. 2018. "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk." *Proceedings of the*

2018 CHI Conference on Human Factors in Computing Systems 449: 1–13. New York: Association for Computing Machinery.

Hawkins, Andrew J. 2018. "The Electric Scooter Craze Is Officially One Year Old—What's Next?" *The Verge*, September 20. https://www.theverge.com/2018/9/20/17878676/electric-scooter-bird-lime-uber-lyft.

Kahn, Jeremy. 2021. "Data-labeling Company Scale AI Valued at $7.3 Billion with New Funding." *Fortune*, April 13.

Lee, Dami. 2019. "Adobe Is Cutting Off Users in Venezuela Due to US Sanctions." *The Verge*, October 7. https://www.theverge.com/2019/10/7/20904030/adobe-venezuela-photoshop-behance-us-sanctions.

Magistretti, Bérénice. 2017. "Playment Raises $1.6 Million to Improve AI Training through Crowdsourced Data Tagging." *VentureBeat*, November 21. https://venturebeat.com/2017/11/21/playment-raises-1-6-million-to-improve-ai-training-through-crowdsourced-data-tagging/.

Schmelzer, Ron. 2019. "What Happens When Self-driving Cars Kill People?" *Forbes*, September 26. https://www.forbes.com/sites/cognitiveworld/2019/09/26/what-happens-with-self-driving-cars-kill-people/.

Schmidt, Florian A. 2017. *Digital Labour Markets in the Platform Economy: Mapping the Political Challenges of Crowd Work and Gig Work*. Bonn: Friedrich-Ebert-Stiftung. https://library.fes.de/pdf-files/wiso/13164.pdf.

Schmidt, Florian A. 2019. *Crowdproduktion von Trainingsdaten* [Crowd production of training data]. HBS Study no. 417, February. Düsseldorf: Hans-Böckler-Stiftung. https://www.boeckler.de/pdf/p_study_hbs_417.pdf.

Tubaro, Paola, and Antonio Casilli. 2019. "Micro-work, Artificial Intelligence and the Automotive Industry." *Journal of Industrial and Business Economics* 46 (3): 1–13. https://hal.archives-ouvertes.fr/hal-02148979/document.