Hans Böckler Stiftung

Mitbestimmung · Forschung · Stipendien

WORKING PAPER FORSCHUNGSFÖRDERUNG

Number 155, August 2019

Crowdsourced Production of AI Training Data

How Human Workers Teach Self-Driving Cars How to See

Florian Alexander Schmidt

FORSCHUNGS-
FÖRDERUNG

Author:

Dr. Florian Alexander Schmidt is professor for conceptual design and media theory at the University of Applied Sciences HTW Dresden. Since 2006 he has been studying the structure and mechanics of digital platforms. His latest book, “Crowd Design”, was published by Birkhäuser in 2017.

This paper is condensed version of a 70-page-long report in German, published in February 2019:
Schmidt, Florian A. (2019): Crowdproduktion von Trainingsdaten. Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge. Study Nr. 415, Düsseldorf: Hans-Böckler-Stiftung. Online: https://www.boeckler.de/pdf/p_study_hbs_417.pdf.

ISSN 2509-2359

Contents

Executive Summary

Since 2017 the automotive industry has developed a high demand for ground truth data. Without this data, the ambitious goal of producing fully autonomous vehicles will remain out of reach. The self-driving car depends on so-called self-learning algorithms, which require large amounts of "training data". The production of this training data or "ground truth data" requires vast amounts of manual labour in data annotation, performed by crowdworkers across the globe. The crowdworkers both train AI systems and are trained by AI systems. Humans and machines work together in ever more complex structures.

An end of this work is not in sight; according to interviews with experts conducted for the study, the demand for this type of labour will continue to grow rapidly in the foreseeable future. However, as the study also shows, while this type of labour creates a new class of skilled crowdworkers, the precariousness of this work remains high because individual tasks are continuously under threat either of being automated or of being further outsourced to an even cheaper region in the world. As the study shows, 2018 saw an influx of hundreds of thousands of crowdworkers from Venezuela specialising in these tasks. On some new platforms, this group now makes up 75 per cent of the workforce. These recent geographical shifts in the supply of labour are a symptom of deeper structural changes within the crowdsourcing industry.

Prototypical microtasking platforms such as Amazon Mechanical Turk ("MTurk"), here described as "established generalists", typically serve as intermediaries that allow their clients to directly pitch any kind of tasks to a distributed crowd. While the platform does exert some influence on the organisation of work, its preferred role is that of an infrastructure provider that does not want to be held responsible for the quality of the results or the working conditions. Within this old system, however, the established generalists, or "legacy platforms," can't deliver the degree of accuracy required by automotive clients for autonomous driving applications. This has led to the emergence of a number of crowdsourcing platforms designed to cater almost exclusively to clients from the intersection of automotive industry and AI research. Prominent examples for the "new specialists" are Mighty AI, Hive (.ai), Playment, Scale (.ai) and understand.ai. The new specialists are well funded, fast growing, and have quickly gathered substantial crowd sizes – several hundred thousand workers each.

Crucially, they guarantee their clients at least 99.x per cent accuracy of the data. To be able to achieve this, they must invest in new, often AI-enhanced, custom-made production tools that both automatically sup-

port and control the workforce, they must furthermore invest in the preselection and training of the crowdworkers, in more community support for the workers, and in complex layers of quality management and sub-outsourcing.

For clients, this offers an expensive but reliable full-service package. For crowdworkers, this offers to some extent better working conditions, because they no longer have to deal with a diversity of clients with heterogenous tools and demands; instead, they are now reliably paid directly by the platform operator. However, this new arrangement raises far-reaching questions regarding the classification of workers as independent contractors.

Some established generalists have started to transform their services towards producing AI training data; these include Appen (which now also owns Figure Eight, previously CrowdFlower), CloudFactory, Samasource, and Alegion. MTurk, clickworker and Crowd Guru, on the other hand, continue to follow a generalist approach.

More and more digital labour platforms now market themselves as AI companies. The term "crowd" is pushed into the background. This development is also reflected in the fact that the new specialist platforms have "two faces": they have a client-facing company name, website and appearance focused on "AI" – and an entirely different, crowd-facing name, platform and appearance, promising prospective workers easy money through microtasks. Because clients and workers now access separate websites, it has become easier to analyse the constellation of the respective workforces and their fluctuation between the platforms.

Methodology and Research Questions

This paper is condensed version of a 70-page-long German report conducted in 2018 and published in early 2019 (Schmidt 2019). In late 2017 a group of German crowdsourcing platforms committed to the "Crowdsourcing Code of Conduct" (http://crowdsourcing-code.com/), a set of principles developed in cooperation with the trade union IG Metall, observed that the demand for microtasking crowdwork was shifting. Quite suddenly, demand had begun to revolve around the production of vast amounts of high quality training data for autonomous vehicles. This study is an investigation into what these shifts in demand mean for the crowdsourcing industry in general and the crowdworkers in particular.

At the time of writing, there were hardly any academic publications about the intersection of the automotive industry, crowdsourcing and AI research. Because the study was conducted while this highly dynamic phenomenon was unfolding, it does not draw on a literature review (for that see previous HBS Studies 323, by Leimeister et al. 2016, and 391, by Gegenhuber et al. 2018, as well as Schmidt 2016 and 2017).

Instead, the study is based on direct observation of the platforms, their communication with crowdworkers, their community forums, their press releases and advertising, and their tools. This was partly done by logging in as a crowdworker. Other sources were trade shows, business reports, journalistic articles, and interviews in publications such as *Wired* and *TechCrunch*. Some company information, especially about venture capital investments, stems from the database Crunchbase (www.crunchbase.com). Amazon's web traffic analysis tool Alexa (www.alexa.com/siteinfo/) proved to be exceptionally useful for following the ebb and flow of crowdworkers "migrating" between the new platforms.

The study is also based on six qualitative interviews with CEOs of crowdsourcing platforms in the field: Daryn Nakhuda of Mighty AI (https://mighty.ai/); Kevin Guo of Hive (https://thehive.ai/); Siddarth Mall of Playment (https://playment.io/); Christian Rozsenich of clickworker (www.clickworker.com/); Marc Mengler of understand.ai (https://understand.ai/); and Hans Speidel of Crowd Guru (www.crowdguru.de).

Complementary to this, five qualitative interviews with crowdworkers from Venezuela, Brazil and Italy were conducted for the study. All five of them are workers from Mighty AI's platform Spare5 (https://app.spare5.com). Throughout the study, this platform played a prominent role, partly because it was one of the first large providers to specialise entirely on clients from the automotive industry, but also because it was the most open, accessible and supportive of this study. Last but not least, the emergent phenomenon of large groups of crowdworkers from Venezuela could be observed on Mighty AI/Spare5 early on.

The study was initially guided by four questions:

1. Which car companies are using which crowdsourcing platforms?
2. How does the demand of the automotive industry affect the crowdsourcing industry?
3. What is the impact on the working conditions of the crowdworkers?
4. Is the crowd-production of training data a short-lived phenomenon or does it offer long-term economic prospects for crowdworkers?

Unfortunately, the first question could not be answered. From the sets of photos the crowdworkers had to annotate in 2018, it was obvious that the German car industry was among the most important clients of the platforms at that time – plenty of images were shot in the vicinity of the headquarters of auto manufacturers ("OEMs") in the southern part of Germany. Yet no companies could be confirmed officially because this information is considered a trade secret. Established OEMs are secretive about their progress in the development of self-driving cars and also don't want to be associated with crowdsourcing. Accordingly, the platform providers are bound by strict non-disclosure agreements and are careful not to mention any of the bigger automotive brands by name. The remaining three questions will be discussed in the following, condensed, analysis.

The Crowdsourced Production of AI Training Data

In recent years several dozen large and well-funded companies have entered into a fierce race to bring fully autonomous vehicles onto the streets. It is unclear, however, if this goal can actually be achieved in the foreseeable future. Although impressive progress has been made, various hard to solve problems have emerged that have dampened the optimism of the engineers working on this challenge. Nonetheless billions of dollars are currently invested in the development of the technology. Competitors in the field are not just the long established automotive OEMs and their tier-one suppliers, but also tech firms hitherto only active in the computer industry, among them heavyweights like Intel, Nvidia, Google and Baidu. In addition to these there are many new companies from Silicon Valley as well as China determined to disrupt the mobility sector, with autonomous vehicles being at least part of their strategy. It is a spectrum that reaches from Uber and Tesla with billions of dollars, to many smaller, highly specialised firms. By 2018 55 companies had already secured licences to test autonomous vehicles in California. All of them need millions of labelled images as ground truth data to teach their algorithms how to see.

Typically, these images are stills taken from videos shot in traffic. They are then manually annotated to make them machine-readable. Two of the most common among various forms of image labelling are so-called bounding boxes and semantic segmentation maps. Bounding boxes roughly mark the position of individual objects; for semantic segmentation maps every pixel in an image has to be covered by a descriptive label.

The annotations need to be as detailed as possible so that eventually the algorithm learns not only to recognise objects without human supervision but can also predict how vehicles or people will behave in traffic. The reliability of the predictive machine-learning models is directly dependent on the precision of the training data and the production of this data can only partly be automated. A full semantic segmentation of an image can take a human up to two hours to complete, so for the automotive companies who need this data in bulk, fast and with high precision, this work can quickly get very expensive, especially if done in-house.

That is why most automotive companies outsource this work to specialised firms who are either organised as crowdsourcing platforms or Business Process Outsourcing firms (BPOs). Platforms rely on large groups of external freelancers who do the work remotely from their homes from all over the world. BPOs train and employ people locally, typically in developing countries. What BPOs do is not crowdsourcing because the number of workers is limited by their physical presence on site and their workers don't just choose tasks themselves as part-time hobbyists. Hence BPOs are less flexible in dealing with the extreme fluctuations in demand typical for this type of microtasking, but they have more control over their workers and can operate as an intermediary to a workforce that only speaks a regional language – i.e., they can train workers that could not access an international platform on their own because they lack the means in regard to language and technology.

Crowdsourcing platforms and BPOs produce ground truth data that is highly redundant. Their various clients need very similar sets of annotated images and it would be much more efficient for the automotive companies to draw their training data from a collectively produced pool of annotated images. Yet, because of the competitiveness of the market, such a "sharing" model currently seems to be out of the question. In addition, while the training data could to some extent be shared, the more important market, according to the CEOs interviewed for this study, is or will be the market for "validation data." For this type of data, human cognition is needed to evaluate the decisions that machine-learning systems have made in traffic. This, of course, is highly specific and confidential information, data that cannot be shared with competitors.

All CEOs interviewed for this study were convinced that, if anything, the demand for crowdworkers in this field would only rise in the foreseeable future. The machines need to be constantly trained on traffic scenarios that keep growing in complexity in regard to geographic variety, weather conditions, and particularities of traffic rules in different countries. Machines will be able to do more and more of the annotation tasks, but will have to be continuously trained for new tasks and new edge cases. The platforms thus aim for a moving target that continues to require many "humans in the loop" for the machine learning process.

Established Generalists and New Specialists

Amazon Mechanical Turk (www.mturk.com/), or MTurk, founded in 2005, is the oldest and most prototypical microtasking platform and its crowd can be used for various types of tasks. It is the prime example of a generalist platform where the provider, Amazon, serves as an intermediary that allows its clients to directly access a distributed crowd.

While the company does exert some influence on the organisation of work, its preferred role is that of an infrastructure provider who does not want to be held responsible – neither for the quality of the results nor the working conditions. The selection of crowdworkers, the description of tasks, the development of specialised tools, the training of the crowd, the quality control of the results and the payment of the crowd – all this lays in the hands of the client. Indeed some clients from the automotive industry continue to use this established type of general-purpose crowdsourcing platform to produce training data for their autonomous vehicle projects.

However, the specific demand for ground truth data – its unprecedented volume and level of precision – has led to the emergence of a number of new specialist crowdsourcing platforms designed to cater almost exclusively to clients from the intersection of automotive industry and AI research. The most prominent examples of these new specialists are Mighty AI, Hive, Playment, Scale (https://scale.ai/) and understand.ai.

Although these platforms are relatively new to the market, they are well funded (these five alone collected 55 million US dollars in venture capital within four years), fast growing, and have quickly gathered substantial crowd sizes. Mighty AI and Playment started out as generalists inspired by MTurk and then quickly evolved into new specialists who now sometimes refer to older platforms in the style of MTurk as “legacy crowdsourcing.”

Many established generalists now, too, have started to transform their services towards producing AI training data. This, however, does not have to be image-based but can also include, for example, the training of voice assistants by doing audio recordings. A glimpse at the websites of Appen (https://appen.com/), Figure Eight (www.figure-eight.com/, previously CrowdFlower, now owned by Appen), CloudFactory (www.cloudfactory.com/), Samasource (www.samasource.com/) and Alegion (https://alegion.com/) shows that it has become crucial for most of the platforms in the table below to explicitly attract the big budgets of the automotive industry currently spent on the development of self-driving vehicles. MTurk, clickworker and Crowd Guru continue to take a generalist approach.

More and more platforms now market themselves as “AI” companies. Also former CrowdFlower has rebranded itself in that manner. If “AI” is not explicitly included in the company name or URL, then it is emphasised in the way companies address prospective clients. The term “crowd” is pushed into the background, presumably due to its negative connotations of cheap labour and low quality results.

This development is also reflected in the fact that the new specialist platforms have two “faces”: They have a client-facing company name, website and appearance focused on “AI” – and an entirely different crowd-facing name, platform and appearance, promising prospective workers how easy it is to make money through microtasks. Mighty AI’s crowd-platform, as mentioned above, is Spare5, Hive’s platform is Hive Work (https://hivemicro.com/), and Scale’s platform is Remotasks (www.remotasks.com/).

Clients of the training data providers are confronted with a mixed marketing message of AI automation and manual labour. All platforms discussed here operate with a combination of both. Figure Eight describes itself as a “human-in-the-loop AI platform for data science & machine learning,” Mighty AI offers “training data as a service,” Hive promises “large scale, quality training data within hours” delivered via a “manual labeling platform,” Playment calls itself a “fully managed human intelligence platform,” Scale advertises its service as “human-powered data for AI applications” and itself as a “developer API for human intelligence, and understand.ai is “training and validating your algorithms with accurate annotations made with German precision.”

Table 1: Relevant providers of training data in 2018

Company	Platform	Origin (est.)	Alexa rank	Crowd size	Funding
Amazon	MTurk	USA (2005)	5,800	500,000	–
Appen	Appen	AUS (1998)	21,000	1,000,000	(public: APX)
Figure Eight	(various)	USA (2007)	30,000	(5,000,000)	$58 Million
clickworker	clickworker	GER (2005)	35,000	1,200,000	$13.7 Million
Mighty AI	Spare5	USA (2014)	37,000	500,000	$27.3 Million
Hive (.ai)	Hive Work	USA (2013)	49,000	300,000	$18 Million
Playment	Playment	IND (2015)	168,000	300,000	$2.3 Million
Scale (.ai)	Remotasks	USA (2016)	187,000	–	$4.6 Million
CloudFactory	(BPO)	UK (2011)	(334,000)	(3,000)	$13 Million
Crowd Guru	Crowd Guru	GER (2008)	416,000	52,000	–
Samasource	(BPO)	USA (2008)	(815,000)	(7,000)	$1.5 Million
Alegion	(various)	USA (2011)	855,000	–	$4.1 Million
understand.ai	(BPO)	GER (2016)	(3.300,000)	–	$2.8 Million

Source: Own research. All figures from mid 2018. The Alexa.com traffic rank of the platforms is more significant than the crowd size claimed by the providers. The BPOs don't have a crowd, which is why their rank and size is less meaningful as a figure, but they are important competitors in the market for training data.

The producers of training data employ more and more AI technology themselves. Paradoxically, the growth of AI increases the demand for manual labour, which in turn increases the demand for AI automation. The goal of the manual labour is the training of AI models, while at the same time similar algorithms are used to support the manual labour and make it more reliable and cost efficient. Humans and AI train each other.

Tailor-Made Tools and Handpicked Crowds

The new specialists for the crowdsourced production of training data guarantee their clients at least 99 per cent accuracy, and this is probably the most impactful shift for the structure of the platforms described here. Before this, they were not at all responsible for the results of the crowd, especially not in the rudimentary platform model established by MTurk.

Scale.ai offers a price calculator on its website. In mid 2018, the price for just nine annotations in a single image was one US dollar – provided the client would buy 10,000 images. The price for the much more complex semantic segmentation of an entire image was 6.40 US dollars. If the client opted for "express urgency" the same service cost 16 US dollars per image.

Considering how many well-funded automotive companies need this type of data in high volumes, it is not surprising that these supply chain platforms have become attractive to venture capital.

Considering that an individual human without AI assistance would need up to two hours for an image that costs 6.40 US dollars in retail, 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 a high-quality standard at speed, volume and a competitive price, the new specialists have a number of instruments or strategies at their disposal.

- Invest in custom-made, AI-enhanced, semi-automatic production tools that estimate in advance the semantic segmentation and then guide the attention of the human cognitive support system specifically to areas where the system is less certain about what it has recognised in the data.
- Invest in quality management through process optimisation in regard to how jobs are split up into atomic microtasks and then reassembled – with many interlocking layers of quality control and monitoring of the results done alternately by humans and AI systems.
- Invest in training of the crowd, community management and gamification mechanisms to keep the crowd skilled, motivated, happy and efficient.
- Invest in access to a cheaper workforce by translating the tasks, the training and the community management into the language of low wage regions on the global market for digital labour or sub-outsource part of the labour to BPOs in developing countries.

The first three points on this list, i.e. tool design, process optimisation and crowd training (or crowd control) increasingly morph into one advanced software solution. Although, interestingly, the competing producers of training data try to gain a competitive edge by favouring different constellations or "stacking orders" of these quality control layers. While some, like understand.ai, invest predominantly into new AI tools and try to reduce the number of workers necessary, others, like Playment, prioritise access to new and cheap workers and (as the company name suggests) gamification mechanisms to keep them happy.

Workers on the new specialist platforms typically have to go through a longer phase of training, different paths for different types of tasks, and like in a computer game, they have to "level up" to qualify for the more sophisticated, better paying tasks. The accuracy of the individual workers is tracked constantly and they get qualitative and quantitative feedback on how well they do. The workers can specialise on certain types of tasks to level up more quickly, and this affects which types of available tasks they can see and do.

As Daryn Nakhuda of Mighty AI explained in the interview for this study, the platform also funnels incoming new tasks to certain subcrowds on the platform to train these pre-selected groups more quickly and efficiently. In this regard, the workers are not an open and unstructured crowd anymore, self-selecting incoming tasks freely. Instead, in this advanced form of crowdsourcing, there are now various degrees of hierarchy, specialisation and orchestration conducted by the platform providers.

The new specialists advertise the shift away from the rudimentary crowdsourcing model as "trained crowd labour", "known crowds", "curated crowds" or "crowd qualification". It has become important for them to communicate to clients that the work is not given to a random, anonymous and potentially incompetent mass, but to handpicked groups of experts that are trained and monitored constantly.

Quality Management and Misclassification

In another departure from the conventions of "legacy crowdsourcing", the new specialists offer their clients "fully managed" data labelling, "end-to-end project management", and that "nothing is done by you". Instead of giving their clients direct access to the crowd, they operate as full service black boxes. This in turn means that the crowdworkers on the new platforms no longer have direct contact with clients, which constitutes a consequential shift for all three parties involved.

In contrast to platforms like MTurk, the clients no longer have to develop their own tools for data annotation, train the crowd themselves, or evaluate the results provided by individual workers. For this convenience they have to pay substantially higher prices, which is the reason why some experienced clients continue to use MTurk.

For workers, the shift means that they less often have to learn new software tools, that the tools are more reliable, convenient to use and constantly developed further as proprietary assets of the platform, and also that the task descriptions are less faulty and easier to understand. This, in combination with the collection of "experience points" in the gamification systems, does have a lock-in-effect: switching to a new platform means losing one's reputation and qualifications, as well as the accumulated skill of handling proprietary tools, at least to some extent. Still, for the workers it is much more safe and reliable to only deal with platform providers – instead of dealing with 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 results due to faulty tools, or even wage theft (as is a serious concern for workers on MTurk).

Although the payment is more reliable, it is not necessarily higher than on conventional crowdwork platforms. Yet it seems that the workers on the new platforms feel better treated in comparison to conventional crowdwork, mostly because they have reliable human interaction with the platform staff in the form of good community management and direct, immediate and personal responses to questions. The platforms, due to their investment in the training of the crowd, in turn have a higher interest in keeping well-trained crowdworkers on the platform.

There is, however, a looming legal risk for the new specialists who offer their clients full service: They could be sued for misclassification of their workers as independent contractors, as happened to CrowdFlower in the past (Otey v. CrowdFlower, class action law suit filed 2012, settled 2015). Rudimentary platforms like MTurk can quite plausibly argue that their crowdworkers are freelancers working for external clients. But platforms that orchestrate the labour on a granular level, train the workforce, and assign jobs to individuals – and on which workers have no direct contact with clients, are in a more difficult position to defend the current classification of their workforce as independent contractors, especially when people work full-time on the platform.

Although the new system for quality management is in many ways more reliable for workers and they feel treated with more respect, the entire business model is built on shaky ground in regard to the legal status of the workers.

Migrant Crowds and German Platforms

The majority of microtasks needed for the training of machine learning systems can be done from all over the world, while only some notable exceptions require regional language skills or local cultural knowledge. In a connected world, digital labour thus “flows” dynamically to those people who are willing to accept the lowest remuneration at any given time – be it because they regard themselves as hobbyists, or, much more often, because they desperately have to take whatever job they can find on the internet. This is why the average hourly wage paid out by the platforms apparently levels out globally, like in a system of communicating vessels. At the time of writing, experienced crowdworkers producing training data earned between one and two US dollars per hour (estimation based on interviews with CEOs, with crowdworkers, and on forum debates among workers; inexperienced workers earn much less as this is piecework).

During the time this study was conducted it was Venezuela that played a key role as an inadvertent supplier of cheap digital labour for the production of training data. The country has a well-connected, well-educated middle class that has collectively fallen into hardship due to the economic collapse of the national oil industry. The people of Venezuela experienced an almost unprecedented hyperinflation exactly at the time that the specialised crowdsourcing platforms in service of the car industry scaled up their operation. As a consequence, this type of crowdwork became a lifeline for hundreds of thousands of Venezuelans. They became online “migrant workers”, roaming between different platforms without having to – or being able to – leave their country.

The platforms discussed in this study each have their own composition of crowdworkers from different nations, and this mix is subject to constant change. The web traffic analysis tool www.alexa.com/siteinfo turned out to be a very informative source for the constellation of the crowd, especially for those cases where the producers of training data have separated their crowd-facing platform from their client-facing operation. Here, the country of origin tracked is especially revealing.

At the end of 2018, Spare5 was ranked 167th of the most frequented websites in Venezuela; Hive was ranked 187th. The country supplied around 75 per cent of the workforce of Spare5 throughout 2018. At Hive Work, this percentage rose from 55 to 75 over the course of that year. According to Kevin Guo of Hive, the crowd of this platform grew from 100,000 to 300,000 in the first half of 2018, at a speed of up to 3,000 new registrations per day. Considering the parallel rapid growth on

Spare5, at least 200,000 people from Venezuela must have been in search of work at the time on these two platforms alone. Because of the oversupply of labour, the job seekers then went back and forth between the two (as reported by the crowdworkers interviewed for the report).

At the same time, Scale's Remotasks predominantly attracted workers from the Philippines and India, with 31 per cent of the crowd coming from each of the two countries. Playment drew its workforce entirely from India, Crowd Guru almost entirely from Germany.

For training data providers without a crowd platform separate from the client website, the traffic is not representative of their workforce because it mixes the origins of clients with that of workers. Clickworker, for example, drew up to 27 per cent of its traffic from the US and 18 per cent from Germany.

The BPOs only have workers where they also have offices. Samasource gets 29 per cent of its traffic from the US, which probably represents mostly traffic from clients, and in addition gets 23 per cent of its traffic from Kenya and 16 per cent from India. CloudFactory produces in Kenya and Nepal, from where it got 41 and 24 per cent traffic respectively in 2018. The German platform understand.ai does not have an open crowd platform either, but operates with a mix of German students who work part-time for the company, and a BPO-like service provider in India.

If training data can be produced independently of the cultural background of the workers, as is the case with most image labelling tasks, providers can get a competitive advantage by accessing workforces from developing nations with substantially lower wage levels. By the same token, it is tough if not impossible for a platform like Crowd Guru with an all-German crowd and a commitment to fair payment to compete with platforms that maintain a crowd from Venezuela for a tenth of the cost for labour.

However, the situation is completely different for platforms like clickworker and Appen, who are strong in producing training data that teaches voice assistant systems to understand various dialects. For this type of task, a diverse crowd is key. The majority of the work is not exported to developing countries but to those regions where consumers have the greatest purchasing power. The tasks are also much better paid. But because they are spread widely across people with various dialects, the demand is too sporadic for these individuals to make a living from it. Here, we indeed find the successful hobbyist crowd who only works on occasion.

Workers' Perspective: Five "Fives" from Spare5

Five crowdworkers from Spare5 were interviewed for this study: a 58-year-old female nursery school teacher from Brazil, a 55-year-old female fashion retailer from Italy, a 34-year-old female prospective dentist from Venezuela, a 22-year-old male electrical-engineering student from Venezuela, and a 20-year-old male mechanical engineering student from Venezuela. The platform provider had recommended most interviewees because they were particularly successful "Fives", as the workers on Spare5 are called.

This small sample is of course not representative, but the explorative qualitative interviews with these highly active workers made it possible, to some extent, to counterbalance the positions of the CEOs with information about the experiences of the crowdworkers. The interviews were conducted via Skype.

In summary, all five workers had a high level of education; most had found their way to the platform through active online search and had already had experience with less satisfactory working conditions on other crowdwork platforms; all five were generally very satisfied with Spare5, especially in comparison with other providers of microtasking. The most important reason was that they felt they were dealing with a trustworthy company ("not a scam") and community managers who treated them with respect and replied quickly when problems occurred. The user-friendly interface design of the tools was mentioned frequently as well.

The workers also experienced the payment on Spare5 as slightly higher and more reliable than on other platforms. The two older women who had been on the platform a little longer described that their earnings had gone down significantly (from 100 US dollars per week to 50 in one case) since the arrival of large groups of people from Venezuela, not only because the price per task went down (from 5 cents to 2 cents in some cases) but also because fewer tasks were available. Both women mentioned they felt lucky they didn't rely on the tasks as their main income and saw them instead as a hobby that generated some additional money "on the side."

In stark contrast, the three workers from Venezuela had grown economically dependent on the tasks. They earned between 20 and 50 US dollars per week, depending on the availability of the tasks; per hour they earned on average 1.50 to 2 US dollars. They were perceived as relatively affluent by their family and neighbours because of this income, and they all recruited and trained friends and relatives in doing this type of crowdwork to help them get through the crisis. Yet they were very aware that by doing so they would likely lower the price of labour even further. The younger engineering student said that some colleagues want to keep “the goose that lays the golden eggs” a secret, but he felt morally obliged to tell others about it.

All five “Fives” emphasised that they enjoyed the work. They experienced it as intrinsically rewarding and in addition they were proud of the quantified (and gamified) feedback they got in the form of experience points and special ranks.

Yet they were also frustrated that at times they didn’t see – or were not shown – any tasks they could do on the platform, which they found especially irritating when they knew of colleagues who did see tasks. For workers, the reasons why some saw tasks while others didn’t remained opaque. This lack of transparency was a constant source of concern because they could only speculate whether this was because their previous work was not deemed accurate enough, whether they didn’t have enough experience points or whether there was another reason unrelated to them personally, such as that the platform had decided – by management or algorithm – to funnel work only to a smaller group of workers to train them more efficiently. “Why don’t I see any tasks?” was the most common concern in company forums and among workers at the time of writing.

Dealing with Fluctuations in Labour Demand

Arguably the most important function of crowdsourcing platforms is to provide clients with a buffer for rapid fluctuations in their demand for labour. In this the platforms resemble local companies for temporary employment, only that the frequency and the volume with which crowd platforms can mobilise and dismiss supporting staff is by orders of magnitude higher than would be possible for location-based services rooted in the physical world (including BPOs).

For the car companies with their sudden demand for high volumes of digital labour it would in most cases make no sense to build up a workforce of the required size in-house because the labour is not needed constantly but in waves.

The crowd platforms, by serving multiple clients at the same time, should theoretically be able to level out the waves of requests of their multiple clients into a constant stream of demand, but this is often not the case. And since the platforms don't have to constantly pay their workforce either, only in the very moments they are actually working, they can use their reserve army of part-time crowdworkers as a buffer.

Theoretically, or in the original idea of crowdsourcing, these people are only activated when a larger workforce is required; they are hobbyists, happy to earn some extra cash on the side, but do not depend on it. However, as the examples above have shown, in times when there is a lot of work available, hundreds of thousand of people from poor regions of the world can quickly grow dependent on this new source of income.

The platform-based outsourcing companies neither have to physically open up shop abroad nor do they even have to actively attract a workforce – the people find their way through word-of-mouth, through worker forums, through YouTube, through constant online search for jobs like these. All the platforms have to do is make the work available in a language spoken by a lot of underemployed people with internet access.

Many of the people looking for this type of work are in desperate economic situations, like the workers from Venezuela who sometimes even have to feed an entire family through microtasking. For them, the extreme fluctuations in the availability of tasks are a serious concern, which they in turn try to buffer by virtually migrating back and forth between different digital labour platforms.

The oversupply of labour unsurprisingly leads to a deterioration of average hourly wages on the platforms and a lot of stress for the workers who can never be sure whether they'll find work.

For the platforms, however, the oversupply of labour is not a bug but a feature. It is necessary to swiftly cope with peaks in demand and deliver results in record time. The problem is well known in other areas of the gig economy, too, and is addressed through strategies like Uber's "surge pricing", a technique that Hive Work and others have started to use as well.

Conclusion

Today there are at least several hundred thousand people across the globe producing training data as crowdworkers for the automotive industry. This is of course only a tiny niche of the labour market, and even though it has been growing rapidly over the last few years it will remain without direct consequences for most people. And yet it is instructive to pay close attention to this highly dynamic phenomenon as it is situated at the cutting edge of how humans work in close cooperation with AI machine learning systems.

The crowdsourced production of training data can give us a glimpse into a future in which, contrary to common belief, humans are not replaced by the machine, but instead find themselves in a complex and precarious relationship with it – by working within it. Humans train AI and are in turn trained by AI. Humans control the results of AI while in turn being controlled by AI.

Paradoxically, it is repetitious manual labour that makes the current progress in automation possible. The often proclaimed "end of work" is not in sight. That said, if one looks at individual tasks, the machine indeed constantly replaces workers, while at the same time new tasks keep opening up that require human cognitive capabilities.

Often enough, the machine is already pretty good at solving a task, but it is simply cheaper to have the job done by crowdworkers. However, amid the current hype around AI, it is not only the labour of the crowd but also that of the programmers developing these systems, that is downplayed or even made invisible in order to make artificial intelligence appear more impressive. Seen from this angle, crowdsourcing continues to be the "dirty secret" of automation. This is why Astra Taylor describes these labour intensive automation systems as "fauxtomation" (Taylor 2018).

Already in 2005 Amazon Mechanical Turk advertised its service as "*artificial* artificial intelligence", and for the various specialist crowdwork platforms who have now added "AI" to their company name or URL, this slogan still holds true, maybe even more than ever.

Within these new systems, the crowd workforce has become a cognitive processing layer within a much larger automation and outsourcing apparatus. Humans and machines form a cybernetic organism, built from layers of artificial and human intelligence. Interestingly, the new specialist platforms try to get a competitive advantage over each other by experimenting with different stacking-orders – alternating successions of humans and algorithms. In this, the platforms resemble not only what Benjamin Bratton has described as "the stack" (Bratton 2016);

there is also a self-similarity reminiscent of the processing layers within artificial neural networks.

Returning to the initial research questions of this study: How does the demand of the automotive industry affect the crowdsourcing industry? And what is the impact on the working conditions of the crowdworkers?

As we have seen: On the one hand, the demand of the automotive industry for highly accurate training data does change the crowdsourcing industry in ways that improve the working conditions of the crowd significantly, at least in some respects. Interestingly, the improvements are remarkably similar to what researchers around Aniket Kittur in 2013 had sketched out as an idealised “future of crowdwork” (Kittur et al. 2013). As Kittur et al. proposed in 2013, crowdworkers on the new specialist platforms today can now indeed follow career-paths by gaining specialised skills and by levelling up in a gamified hierarchy, on platforms that track and document their progress. The workers enjoy doing more skilled labour and appreciate the significantly improved task descriptions and sophisticated digital tools to work with. They welcome the constructive feedback and support by a responsive and personal community management (or crowd management) that treats the workers with respect and replies to their concerns quickly. The workers also benefit from the training by the platform, which makes them less exchangeable and more valuable for the platform providers.

Most importantly, the workers can rely on getting paid weekly by the platforms and don’t have to deal with unpredictable payment practices of ever changing clients treating them merely as sub-human machine parts (Irani/Silberman 2013; Salehi et al., 2015).

On the other hand, even though in the best case scenarios outlined here the experience of the crowdworkers has been improved substantially, a set of interconnected, fundamental and potentially unsolvable problems of a global market for crowdwork remain: the race to the bottom in terms of wages, and, maybe even worse, the constant insecurity or precariousness regarding the question whether there will be enough work the next day.

What the labour is worth in monetary terms is negatively affected by the drive towards ever more automation, and in addition to this, by the fact that the work can be accessed and accomplished from all over the world and thus can very quickly be funnelled to an even cheaper workforce. 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 can’t be a guarantee for them that tomorrow the task they have specialised on will not be automated or performed by someone more desperate.

Probably the most important lesson from studying the crowdsourced production of AI training data is that in the relatively short time of one and a half years the automotive industry was able to access hundreds of thousands of new workers, through a labour supply chain of venture capital funded platforms which sprung up like mushrooms to cater for this new demand. In other words, a massive and globally distributed crowd workforce has been conjured up by capital, almost over night. The crowd is a mass phenomenon in which the individual human is by definition replaceable. The crowd is the equivalent of the seemingly endless mass of repetitive and interchangeable microtasks.

As long as it is that easy to access new crowds, almost at an instant, a crowd that recruits itself and can be trained and managed automatically, the workforce has hardly any negotiating power with which to improve its wages or working conditions. That is to say: the problem of the crowd remains the crowd.

References and Further Reading

Arnoud, Sacha (2018): The Rise of ML in Self-Driving cars. Guest lecture for course 6.S094: Deep Learning for Self-Driving Cars. MIT Boston. Online: https://youtu.be/LSX3qdy0dFg.

Benner, Christiane (editor) (2015): Crowdwork – Zurück in die Zukunft? Perspektiven digitaler Arbeit. Frankfurt am Main: Bund-Verlag.

Bordoli, Robin (2018): Focused on the Future with a New Name. In: Figure Eight (previously CrowdFlower), 27 February, 2018. Online: https://www.figure-eight.com/focused-future-new-name/.

Bratton, Benjamin (2016): The Stack: On Software and Sovereignty, MIT Press.

Charrington, Sam (2016): TWiML Talk #6, Angie Hugeback, Generating Labeled Training Data for Your ML/AI Models, 29 September, 2016. Online: https://twimlai.com/twiml-talk-6-angie-hugeback-generating-labeled-training-data-mlai-models/.

Charrington, Sam (2017): TWiML Talk #57, Daryn Nakhuda, Training Data for Autonomous Vehicles, 23 October, 2017. Online: https://twimlai.com/twiml-talk-057-training-data-autonomous-vehicles-daryn-nakhuda/ (Accessed 24.01.2019).

Cherry, Miriam A. (2016): Beyond Misclassification: The Digital Transformation of Work, 18.02.2016. In: Comparative Labor Law & Policy Journal, Forthcoming; Saint Louis U. Legal Studies Research Paper No. 2016-2. Online: http://papers.ssrn.com/abstract=2734288.

Christensen, Ben (2015): Crowdsourced Data: When to Use Curated Crowds vs. Crowdsourcing. In: Appen Blog, 17 November, 2015. Online: https://appen.com/curated-crowds-vs-crowdsourcing/.

Corporaal, Greetje F./Lehdonvirta, Vili (2017): Platform Sourcing – How Fortune 500 Firms are Adopting Online Freelancing Platforms. University of Oxford: Oxford Internet Institute (OII). Online: https://www.oii.ox.ac.uk/publications/platform-sourcing.pdf.

Davies, Alex (2017): Nissan's Path to Self-Driving Cars? Humans in Call Centers. In: Wired, 5 January, 2017. Online: https://www.wired.com/2017/01/nissans-self-driving-teleoperation/.

Davies, Alex (2018): The Unavoidable Folly of Making Humans Train Self-Driving Cars. In: Wired, 22 June, 2018. Online: https://www.wired.com/story/uber-crash-arizona-human-train-self-driving-cars/.

DMV CA (2017): Autonomous Vehicle Disengagement Reports 2017. Online: https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/disengagement_report_2017.

DMV CA (2018): Autonomous Vehicle Permit Holders 2018. Online: https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/permit.

Figure Eight (2018): What is Human-in-the-Loop Machine Learning? In: figure eight resources. Online: https://www.figure-eight.com/resources/human-in-the-loop/.

Gegenhuber, Thomas/Ellmer, Markus/Scheba, Claudia (2018): Partizipation von Crowdworkerinnen auf Crowdsourcing-Plattformen: Bestandsaufnahme und Ausblick. HBS Study Nr. 391. Düsseldorf: Hans-Böckler-Stiftung. Online: https://www.boeckler.de/5248.htm?produkt=HBS-006913

Graham, Mark/Hjorth, Isis/Lehdonvirta, Vili (2017): Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. In: Transfer: European Review of Labour and Research 23, pp. 135–162. DOI: 10.1177/1024258916687250.

Gray, Mary L./Suri, Siddharth (2017): The Humans Working Behind the AI Curtain. In: Harvard Business Review, 9 January, 2017. Online: https://hbr.org/2017/01/the-humans-working-behind-the-ai-curtain.

Irani, Lilly (2013): The Cultural Work of Microwork. In: New Media & Society 17, pp. 720–739. DOI: 10.1177/1461444813511926.

Irani, Lilly C./Silberman, M. Six (2013): Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. In: CHI ’13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 2013. New York: ACM, pp. 611–620.

Jürgens, Kerstin/Hoffmann, Reiner/Schildmann, Christina (2017): Arbeit transformieren! Denkanstöße der Kommission „Arbeit der Zukunft”. Forschung aus der Hans-Böckler-Stiftung, Arbeit, Beschäftigung, Bildung, Vol. 189. Bielefeld: transcript Verlag. Online: https://www.boeckler.de/6299.htm?produkt=HBS-006612.

Kahn, Jeremy (2018): To Get Ready for Robot Driving, Some Want to Reprogram Pedestrians. In: Bloomberg, 16 August, 2018. Online: https://www.bloomberg.com/news/articles/2018-08-16/to-get-ready-for-robot-driving-some-want-to-reprogram-pedestrians.

Kässer, Matthias, et al. (2018): Artificial intelligence as auto companies’ new engine of value. McKinsey & Company, January 2018. Online: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/artificial-intelligence-as-auto-companies-new-engine-of-value.

Kässer, Matthias/Müller, Thibaut/Tschiesner,Andreas (2017): Analyzing start-up and investment trends in the mobility ecosystem. McKinsey & Company, November 2017. Online: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/analyzing-start-up-and-investment-trends-in-the-mobility-ecosystem.

Kässi, Otto/Lehdonvirta, Vili (2018): Online Labour Index: Measuring the Online Gig Economy for Policy and Research. In: Technological Forecasting and Social Change, Forthcoming, 21 August, 2018. Online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3236285.

Katz, Miranda (2017): Amazon's Turker Crowd Has Had Enough. In: Wired, 23 August, 2017. Online: https://www.wired.com/story/amazons-turker-crowd-has-had-enough/.

Kittur, Aniket, et al. (2013): The Future of Crowd Work. In: CSCW '13 Proceedings of the 2013 conference on Computer supported cooperative work, San Antonio, Texas, USA, 2013. New York: ACM, pp. 1301–1318.

Krafcik, John (2018): Waymo One: The next step on our self-driving journey. In: Medium, 5 December, 2018. Online: https://medium.com/waymo/waymo-one-the-next-step-on-our-selfdriving-journey-6d0c075b0e9b.

Laing, Keith (2018): Few carmakers submit self-driving car safety reports. In: The Detroit News, 10 September, 2018. Online: https://eu.detroitnews.com/story/business/autos/mobility/2018/09/10/few-carmakers-submit-self-driving-safety-assessments/1076691002/.

Lehdonvirta, Vili (2017): Where are online workers located? The international division of digital gig work. In: The iLabour Project, 11 July, 2017. Online: http://ilabour.oii.ox.ac.uk/where-are-online-workers-located-the-international-division-of-digital-gig-work/.

Leimeister, Jan Marco, et al. (2016): Systematisierung und Analyse von Crowd-Sourcing-Anbietern und Crowd-Work-Projekten. HBS Study Nr. 324. Düsseldorf: Hans-Böckler-Stiftung. Online: https://www.boeckler.de/6299.htm?produkt=HBS-006379.

Leimeister, Jan Marco/Durward, David/Zogaj, Shkodran (2016): Crowd Worker in Deutschland: eine empirische Studie zum Arbeitsumfeld auf externen Crowdsourcing-Plattformen. HBS Study Nr. 323. Düsseldorf: Hans-Böckler-Stiftung. Online: https://www.boeckler.de/6299.htm?produkt=HBS-006421.

Levin, Sam/Wong, Julia Carrie (2018): Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. In: The Guardian,19 March, 2018. Online: https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe.

Lundgren, Carole (2017a): Embracing New Opportunities in the Automotive Industry: Appen Establishes Detroit Office. In: Appen Blog, 5 June, 2017. Online: https://appen.com/blog/new-opportunities-automotive-industry-appen-establishes-detroit-office/.

Lundgren, Carole (2017b): AI Requires a Human Touch: Improve Technology with Crowd Recruiting. In: Appen Blog, 31 August, 2017. Online: https://appen.com/blog/ai-requires-human-touch-appen-crowd-recruiting/.

Marshall, Aarian (2018): Why Self-Driving Cars Will Never Be „Ready". In: Wired, 8 April, 2018. Online: https://www.wired.com/story/when-will-self-driving-cars-ready/.

Marvit, Moshe Z. (2014): How Crowdworkers Became the Ghosts in the Digital Machine. In: The Nation, 5 February, 2014. Online: https://www.thenation.com/article/how-crowdworkers-became-ghosts-digital-machine/.

Redrup, Yolanda (2017): Appen to become global leader after $105 million Leapforce acquisition. In: Financial Review, 29 November, 2017. Online: www.afr.com/technology/appen-to-become-globalleader-after-105-million-leapforce-acquisition-20171129-gzv2bd.

Reese, Hope (2016): Is 'data labeling' the new blue-collar job of the AI era? In: TechRepublic,10 March, 2016. Online: https://www.techrepublic.com/article/is-data-labeling-the-new-blue-collar-job-of-the-ai-era/.

Rosenblat, Alex (2018): When Your Boss Is an Algorithm. In: The New York Times, 12 October, 2018. Online: https://www.nytimes.com/2018/10/12/opinion/sunday/uber-driver-life.html.

Ross, Philip E. (2017): 2017: The Year In Robocars. In: IEEE Spectrum: Technology, Engineering, and Science News, 20 December, 2017. Online: https://spectrum.ieee.org/cars-that-think/transportation/self-driving/the-year-in-robocars.

Rowley, Jason (2018): The well-funded startups driven to own the autonomous vehicle stack. In: TechCrunch, 18 April, 2018. Online: https://techcrunch.com/2018/05/27/the-well-funded-startups-driven-to-own-the-autonomous-vehicle-stack/.

Salehi, Niloufar, et al. (2015): We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers. In: CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 2015. New York: ACM, pp. 1621–1630.

Schmidt, Florian A. (2013): For a Few Dollars More: Class Action Against Crowdsourcing. In: APRJA peer reviewed open access journal. Online: http://www.aprja.net/?p=836.

Schmidt, Florian A. (2015): The Good, the Bad and the Ugly. Warum Crowdsourcing eine Frage der Ethik ist. In: Benner, Christiane (editor): Crowdwork – Zurück in die Zukunft? Perspektiven digitaler Arbeit. Frankfurt am Main: Bund-Verlag, pp. 367–385.

Schmidt, Florian A. (2016): Digital Labour Markets in the Platform Economy – Mapping the Political Challenges of Crowd Work and Gig Work. Bonn: Friedrich-Ebert-Stiftung, Online: https://library.fes.de/pdf-files/wiso/13164.pdf.

Schmidt, Florian A. (2017): Crowd Design: From Tools for Empowerment to Platform Capitalism. Basel: Birkhäuser.

Schmidt, Florian A./Kathmann, Ute (2017): Der Job als Gig – digital vermittelte Dienstleistungen in Berlin. Berlin: ArbeitGestalten Beratungsgesellschaft und Senatsverwaltung für Integration, Arbeit und Soziales. Online: https://www.arbeitgestaltengmbh.de/assets/projekte/Joboption-Berlin/Der-Job-als-Gig-Expertise-Digital-November-2017.pdf.

Schmidt, Florian A. (2019): Crowdproduktion von Trainingsdaten. Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge. Study Nr. 415, Düsseldorf: Hans-Böckler-Stiftung. Online: https://www.boeckler.de/pdf/p_study_hbs_417.pdf.

Solon, Olivia (2018): The rise of 'pseudo-AI': how tech firms quietly use humans to do bots' work. In: The Guardian, 6 July, 2018. Online: https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies.

Somers, James (2017): Is AI Riding a One-Trick Pony? In: MIT Technology Review, 29 September, 2017. Online: https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony/.

Stewart, Jack (2017): The Human Army Using Phones to Teach AI to Drive. In: Wired, 9 July, 2017. Online: https://www.wired.com/story/mighty-ai-training-self-driving-cars/.

Taylor, Astra (2018): The Automation Charade. In: Logic Magazine, 2 October, 2018. Online: https://logicmag.io/05-the-automation-charade/.

Venkatesh, Mothi (2017): What's the difference between Playment and Traditional Crowdsourcing? In: Playment Blog, 1 December, 2017, updated 24 September, 2018. Online: https://blog.playment.io/whats-difference-playment-traditional-crowdsourcing/.

All websites were last accessed on 4 June, 2019.

Since 2017 the automotive industry has developed a high demand for ground truth data. Without this data, the ambitious goal of producing fully autonomous vehicles will remain out of reach. The self-driving car depends on self-learning algorithms, which in turn have to undergo a lot of supervised training. This requires vast amounts of manual labour, performed by crowdworkers across the globe. As a consequence, the demand in training data is transforming the crowdsourcing industry. This study is an investigation into the dynamics of this shift and its impacts on the working conditions of the crowdworkers.

ISSN 2509-2359