Learners in the loop: hidden human skills in machine intelligence

Paola Tubaro

Learners in the loop: hidden human skills in machine intelligence

Paola Tubaro, Centre de Recherche en Economie et STatistiques (CREST), Centre National de la Recherche Scientifique (CNRS), Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE), Institut Polytechnique de Paris (IPP). 91120 Palaiseau, France. Email: paola.tubaro@cnrs.fr*

* This is a pre-print version of an article published in *Sociologia del lavoro*, Autumn 2022, n. 163, pp. 110-129, DOI: 10.3280/SL2022-163006. I use data from the research project TRIA (*Le TRavail de l'Intelligence Artificielle*), co-funded by the MITI programme of CNRS and by MSH Paris-Saclay. I gratefully acknowledge the contribution of the TRIA data collection team: M. Gómez Diéguez, J.L. Molina, A. Ortega, J.E. Rodriguez, M.S. Romero Goldar and J. Torres Cierpe. A preliminary version of this paper was presented at the Re-Situating Learning Conference of the CRC Media of Cooperation, University of Siegen, 2021. I am indebted to A.A. Casilli, J. Torres Cierpe, and two anonymous reviewers of *Sociologia del lavoro* for helpful comments on a previous version of this paper.

Abstract

Today's artificial intelligence, largely based on data-intensive machine learning algorithms, relies heavily on the digital labour of invisibilized and precarized humans-in-the-loop who perform multiple functions of data preparation, verification of results, and even impersonation when algorithms fail. Using original quantitative and qualitative data, the present article shows that these workers are highly educated, engage significant (sometimes advanced) skills in their activity, and earnestly learn alongside machines. However, the loop is one in which human workers are at a disadvantage as they experience systematic misrecognition of the value of their competencies and of their contributions to technology, the economy, and ultimately society. This situation hinders negotiations with companies, shifts power away from workers, and challenges the traditional balancing role of the salary institution.

Keywords

Digital labour platforms, artificial intelligence, skills, learning, misrecognition, Spanish-speaking countries.

Introduction

The current hype around machine learning mirrors the dream of an autonomous technical artefact, capable of constantly improving itself like the human mind. The spectacular advances of this technology in the past decade have revived efforts to harness the potential of artificial intelligence in areas as diverse as medical imagery diagnostics, hiring processes, border security and scientific discovery. The language of *learning* refers to the capacity of these information systems to infer patterns from (large masses of) data, without every step being explicitly coded. Put differently, a machine learning algorithm is *trained*, not programmed (Heuer et al., 2021) – and is often equated to the human process of progressively acquiring new understandings or capabilities based on experience.

This does not mean that increasingly efficient automation can dispense with human labour, despite widespread, but yet unproven fears of technological unemployment. A great deal of human intervention is needed at every step of the production of an algorithm – from training, testing and fine-tuning to deployment, commercialization, and sometimes updating and re-training. Beyond the highly qualified, well-remunerated engineers and computer scientists who oversee the whole process, inputs are needed from largely invisibilized, anonymous, and usually underpaid support workers. Tubaro et al. (2020a) outline the numerous entanglements of these lower-level workers with computers in machine learning, as they prepare data to train models, check outputs or correct them if needed, and even impersonate algorithms when they fail. The joint contribution of humans and machines is what the artificial intelligence industry refers to as *human in the loop.*

These workers' participation to machine learning is an instance of digital labour, seen as a technology-fuelled reconfiguration of human productive activity that suits the needs of today's increasingly data-intensive economy. Also known as *micro-work* (Irani, 2015), it is characterized by extreme fragmentation and standardisation of tasks, which may consist, for example, in recording voice, transcribing or translating bits of text, labelling objects in images, and sorting lists. This emerging form of labour grows outside the (regulated) model of paid employment and accompanying social protection schemes. Digital platforms are key enablers of its emergence, leveraging algorithmic management to recruit crowds of workers on demand (Altenried, 2020). Unlike other variants of digital labour such as delivery and transport *gigs* (Casilli, 2019), micro-work can be performed remotely, and platforms match clients to one-person microproviders across the globe (Lehdonvirta et al., 2019), on occasion allowing outsourcing to countries where labour costs are lower. If micro-work was first documented in the USA in the late 2000s, it is increasingly part of offshoring chains that procure workers from places like India (Gray & Suri, 2019), Indonesia (Lindquist, 2018), Brazil (Grohmann & Fernandes Araújo, 2021), Argentina (Miceli et al., 2020) and sub-Saharan Africa (Mohtobi et al., 2018).

Micro-work contributes to the erosion of the salary institution in multiple ways. One is commodification of labour, with very little shielding from market fluctuations via regulative institutions (Wood et al., 2019), exclusion from organizational resources through outsourcing (Tubaro, 2021) and transfer of social reproduction costs to local communities to reduce work-related risks (Posada, 2022). Another is *heteromation*, the extraction of economic value from low-cost labour in computer-mediated networks, which according to Ekbia & Nardi (2017) is a new logic of capital accumulation. In micro-work, heteromation occurs as platforms' technical infrastructures handle worker management problems as if they were computational problems, thereby concealing the employment nature of the relationship, and ultimately disguising human presence. This paper highlights a third channel through which the salary institution is threatened, namely misrecognition of micro-workers' skills, competencies and learning. Broadly speaking, salary can be seen as the framework within which the employment relationship is negotiated and resources are allocated, balancing the claims of workers and employers. In general, the most basic claims revolve around skill (Tomaskovic-Devey & Avent-Holt, 2019), and in today's 'society of performance' (Chicchi, 2020), where value is increasingly extracted from intangible resources and competencies, unskilled workers are substitutable and therefore highly vulnerable. In micro-work, tight breakdown of tasks, algorithmic control and arm's-length transactions obfuscate the competence of workers and discursively undermine their deservingness, shifting power away from them and voiding the equilibrating role of the salary institution.

The paper proceeds in subsequent steps. To prepare the ground for discussion, it begins by assessing the educational qualifications of micro-workers. New data confirm previous findings (such as Berg et al., 2018) according to which these workers are highly educated, and extend them to a context of globalization in which tech industries increasingly draw on workers from the Global South. It appears that a remarkable human capital, from precisely the countries in which it is particularly scarce, supports the *learning* of machines. The data corroborate extant findings that highlight forms of workplace learning on micro-tasking platforms (Margaryan, 2019a; 2019b; Margarayan et al., 2020) and the existence of relatively complex micro-tasks (Schmidt, 2019), while also foregrounding *misrecognition* – a term which, following Honneth (1995), can be essentially seen as the attitudes and practices that result in people not receiving due acknowledgement for their value and contribution to society, in this case in terms of their education, skills, and skill development. Social misrecognition occurs when clients and platforms fail to affirm workers' credentials, abilities or competencies, so that workers' experience of disrespect leads them to adopt a discourse that diminishes the value of their activity and describes it as unskilled, unchallenging and (therefore) undeserving. Building on these results, the paper discusses obstacles to skill use and learning as they emerge from workers' narratives. Platform organization construes work as having little value, and creates disincentives for micro-workers to engage in more complex tasks, weakening their status and their capacity to be perceived as competent. Misrecognition is endemic on platforms and undermines workers' potential for self-realization, negotiation and professional development.

The argument is based on original empirical data from a mixed-method survey of human-in-the-loop workers in two previously under-researched settings, namely Spain and Spanish-speaking Latin America. This choice broadens the scope of analysis, beyond the English-language populations and high-profile platforms already explored in the (growing, but still limited) literature on human-in-the-loop work. Many studies have focused on Amazon Mechanical Turk, which was unquestionably central to the initial emergence of this form of labour, but may no longer be representative of the latest trends in an industry in which global participation has dramatically increased as noted above, and the number of platforms has tripled (ILO, 2021). Importantly, 75% of Mechanical Turk workers are Americans, and 16% Indians (Difallah et al., 2018), with attributes and behaviours that do not necessarily generalize to other national and linguistic groups. Spain and Latin America constitute a relevant setting to compare and contrast higher- and middle-to-lower income countries, which nevertheless share cultural and historical proximities. Latin America also includes the exceptional case of Venezuela, where extremely severe economic, political and monetary conditions (ENCOVI, 2021) increase the attractiveness of international platform work, which pays in hard currency.

Who learns in machine learning? The role of humans

Beyond the current hype, the notion that machines learn should not be taken too literally: it is a mere figure of speech for the statistical operation of synthesizing from past data, and involves much narrower capabilities than any human cognitive processes. But without overstating its meaning, it is a useful metaphor that helps pull out the actual contribution of the human workers in the loop – who act as facilitators for machines and as discussed later, are learners themselves.

The basic idea of the human-in-the-loop is to let workers do tasks that computers cannot do well, even though they may seem simple or even trivial. These tasks typically require intuition, common sense, or familiarity with some cultural and linguistic codes. Before Amazon launched its pioneering Mechanical Turk platform in the mid-2000s, it had tested this idea internally as part of an effort to remove duplicates from its catalogue – a job that had resisted efforts to automate it, but that workers could do almost effortlessly. The 'Artificial Artificial Intelligence' slogan, with which Mechanical Turk services were later advertised, is a clear illustration of the idea that human judgement is the most effective means to achieve a goal when an artificial intelligence solution would be too costly, or outright impossible.

One of the earliest large-scale experiments with this principle was the annotation of ImageNet (Denton et al., 2021), a large dataset for computer vision, for which thousands of *Turkers* were recruited in 2010-11 to associate each image to a word or concept. They provided examples, so to speak, to train machines to classify images or detect objects in them. It was with this iconic dataset that so-called *deep learning* – a type of machine learning that uses models of the neural networks in the human brain – unlocked new avenues for artificial intelligence in 2012. Progress was spectacular, and computers' object recognition capabilities reached unprecedented levels of accuracy. The humans' *teaching* was effective and some annotation tasks were later automated, using machine learning tools to extrapolate from the original human-annotated data. This does not mean that humans are no longer needed: computers still need their help whenever new cases or new categories appear for which there are no prior examples, when the available information is insufficient or ambiguous, when there are outliers or rare occurrences. Sometimes, the whole process must be done by hand for a while, before ways to automate it can be even figured out (Tubaro et al., 2020a). Industry studies suggest a rise in the demand for human data workers, which parallels improvements in the learning of machines (Cognylitica, 2020).

These workers also check the quality of machine outputs and make corrections when needed. Interesting examples are voice assistants such as Siri or Alexa, already widely marketed products of artificial intelligence. Even after the assistant has been trained, there is a constant need to ensure it correctly recognizes users' utterances, so that any change in language (like the names of new celebrities) can be taken into account. This may mean re-training. Verification and quality assurance tasks are unavoidable and unlikely to be automated, although human intervention may create discomfort in users. In Spring 2019, widespread outrage followed media revelations that human workers routinely listen to users' conversations with their voice assistants (Tubaro & Casilli, 2022). But in general, the presence of humans remains in the backstage, in order not to diminish the appeal of automation on which technology companies base their marketing. According to Newlands (2021), the curtain on these human contributions is only lifted when vendors must co-opt unpaid labour from users, for example to train localised chatbot services, and need to elicit their cooperation.

What type of tasks do micro-workers do? The case of ImageNet portrayed annotation micro-tasks as something that anyone could do – whereby thousands of different workers would be just equivalent to one another, regardless of individual specificities (Denton et al., 2021), let alone advanced or unique skills. This view is reflected in the design of micro-tasking platforms, where workers are usually anonymous because for the purposes of doing very basic tasks, they are seen as by and large substitutable to each other. Accordingly, micro-tasks are commonly described as elementary, short and repetitive. However, what was true at the time of ImageNet may no longer be an accurate description of reality. There is preliminary evidence that in domains such as the annotation of traffic images for the development of self-driving cars, tasks may have become more complex and demanding (Schmidt, 2019; Tubaro & Casilli, 2019). Research in computer science has developed tools to handle non-trivial tasks and even to map them to specific skills (Mavridis et al., 2016), although primarily for use in citizen science rather than commercial applications. While the ratio of complex vs. simpler micro-tasks in today's digital industries is hard to establish, there seems to be a division of labour, where most intermediaries specialize in different types of tasks, with only few generalist ones. Overall, there is a hitherto unacknowledged ideological element behind a view that still insists on simplicity – as a way to keep prices down for cost-conscious clients.

Even when they do very basic tasks, are workers themselves unskilled? Extant evidence suggests that they are rather highly educated (Berg et al., 2018; Casilli et al., 2019). Behind this apparent mismatch (ILO, 2021, p. 22), there may be a range of life course factors, trajectories and motivations that lead people to platform work irrespective of its fit with their actual skill levels (Margaryan et al., 2020), although these factors are still poorly understood. The other surprising finding (Margaryan, 2019a; 2019b) is that workers in the machine-learning loop undertake several workplace learning activities out of their own volition, appear self-efficacious, reflective and intrinsically motivated. However, the same studies indicate that these workers invest less in learning, and are less able to learn from collaboration with peers and from interactions with customers, compared to other platform workers – notably highly-qualified freelancers like designers, accountants or writers, who use digital technologies to expand their client base.

The present paper offers a twofold contribution to this literature. First, it provides insight into task complexity and the apparent mismatch between workers' skills and the content of human-in-the-loop tasks, by linking evidence on educational backgrounds, competencies and learning processes to the functioning of the platform economy and the way it challenges the salary institution. Second, it sheds light on the situations, views and purposes of workers outside commonly researched demographics. Differences are substantial insofar as micro-tasks are mostly an accessory activity for workers in the United States and Europe, and more often the main source of income in lesser-known emerging and lower-income countries (Ipeirotis, 2010). These alternative settings reveal the effects of globalization and the persistence of inherited inequalities. It appears that human learning, as a process embedded in machine learning, is subject to exploitative practices rather than providing professional development opportunities that may stabilize and improve the situation of individual workers.

Data and methods

To advance these arguments, the paper uses data from a mixed-method study of workers from Spain and Spanish-speaking Latin America who use platforms to do tasks for machine-learning development. It was conducted in 2020-21 and includes online questionnaires (339 from Spain, 1138 from Latin America) and follow-up semi-structured interviews (5 Spain, 52 Latin America). The questionnaire and interviews were fielded as paid micro-tasks on the US-based platform Microworkers.com, which operates worldwide, is popular in Spanish-speaking countries, and is open to new registrations. The Latin American countries included are Argentina, Bolivia, Chili, Colombia, Costa Rica, Ecuador, Dominican Republic, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela.

The questionnaire (in Spanish) was long, with over 100 questions covering basic socio-demographic information, family situation, professional status, income, internet usage, practices of micro-tasking on platforms, and a specific set of questions on skills, measured through highest degree obtained, subject of specialization, student status at the time of completing the survey, fluency in foreign languages, and basic digital skills. The analysis that follows mostly uses these variables. Additionally, it leverages the qualitative interviews, which were directed at workers who had already completed the questionnaire, and aimed to better appreciate the meanings that they gave to their responses, together with their lived experience. The interviews lasted about 45 minutes on average, and were conducted through a web conferencing system, usually by two interviewers. They were all recorded and transcribed. Preliminary, reflective analysis of these data took place in the form of debriefing reports after each interview and regular meetings among the research team. After the fieldwork was finished, the interview transcripts and reports were read and excerpts coded. The codes are partly descriptive (types of tasks, personal trajectories) and partly theoretical, inspired by literature on platform labour, the social impacts of technologies, and international outsourcing.

Notions of statistical representativeness are difficult to apply to platform settings where the characteristics and boundaries of the user populations are unknown (Tubaro et al., 2020b; Kässi et al., 2021). However, demographic data are broadly in line with what has been reported elsewhere. As in Berg et al. (2018), women constitute one third of respondents, and as in Casilli et al. (2019), about half are aged 25-44. Importantly, platform micro-tasks constitute the main source of income of two out of five Latin American respondents, but of only one out of five Spaniards. Likewise, almost two out of five Latin Americans who had already earned money through micro-work at the time of the survey, used it to pay for necessities (for example food, rent, utility bills), or to repay debt, against one out of four Spanish earners. In an optional comment to this question, one respondent volunteered 'in Venezuela, the priority is food, seriously', and another 'I put the money aside for my son's operation'.

This article uses questionnaire and interview data jointly. Questionnaire data are used to provide a broad overview of the micro-working universe. Key variables are interpreted by comparison with general population data, within and across countries, in order to gauge the specificities of these workers. Interview data are then used to augment the quantitative findings, adding depth and detail based on the experience of workers, the discourse they use, and their feelings and motivations. Focus is on the codes that are closest to the questions addressed by the quantitative analysis, in order for the former to help interpret the latter and put it in context.

University-educated learners

Micro-tasking platforms do not usually require any evidence of formal degrees to register, so that educational attainments are not part of workers' profiles. This practice has the merit of levelling the field, making platforms more widely accessible and avoiding thorny degree comparability problems. However, it implies that information on workers' education can only be obtained by surveying them – and that it is therefore subject to the well-known limitations of self-reported data. Another consequence is that workers cannot use their educational credentials to assess their suitability for a job or task, or to negotiate conditions and remunerations, as they would normally do in conventional labour markets. This is a form of misrecognition, that removes standard benchmarks and deprives workers of a bargaining tool.

So, what do questionnaire data reveal about micro-workers' education? Among Spanish respondents, 70% have some higher education, a proportion that increases to almost 80% among Latin Americans. To see how this compares to the general population, it is useful to restrict the analysis to the 25-64 age range, corresponding to adults who are likely to have completed their education, and for whom there are national statistical data that can be used as benchmarks. Table 1 reports educational levels for study respondents (left panel), and general population (right), within this age range. Because general population data are available at country level, focus is on the four Latin American countries with the largest numbers of respondents in the desired age range. General population data are from OECD's *Education at a Glance* report (2021), except for Venezuela which is not covered by OECD studies, and for which data are from the national ENCOVI, *Encuesta Nacional de Condiciones de vida* (2018) survey, conducted by Universidad Católica Andrés Bello.

As mentioned above, the figures for micro-workers are self-reported and should therefore be interpreted with some caution, all the more so as samples are small for some Latin American countries in the table (a large proportion of participants from these countries being aged 18-24). Nonetheless, a clear and consistent picture emerges: among these workers, those with higher education are over-represented compared to the general population of similar age. If fewer human-in-the-loop workers hold Master's and PhD degrees in Latin America than in Spain, the gap with respect to the general population is larger in Latin America. This result means that in countries where access to higher education is relatively limited (to, for example, less than 20% of the population in Mexico, or less than 12% in Venezuela), it is largely within the small subpopulation that has had this opportunity, that workers are recruited for the needs of the artificial intelligence industry. A significant human capital, from the countries in which it is particularly scarce, contributes to the *learning* of machines that are mostly conceived in, and controlled from, higher income-countries.

Table 1: Percentage of human-in-the-loop task workers aged 25-64 (left) and general population in the same age range (right) with a given level of education as the highest level attained, in 2020. A short HE degree corresponds to at least two years in full-time higher education; a Bachelor's degree to three to four years; and a Master's or PhD degree to five to eight years. Interpretation: Among the 249 Spanish micro-taskers aged 25-64, 46% have a bachelor's degree, against 11% in the general Spanish population within the same age range. Source: author's elaboration (left), and OECD (2021) (right), unless otherwise specified.

<table>
<tr><th></th><th colspan="4">Micro-taskers (age 25-64) (%)</th><th colspan="3">General population (age 25-64) (%)</th></tr>
<tr><th></th><th>n</th><th>Short HE degree</th><th>Bachelor's</th><th>Master's or PhD</th><th>Short HE degree</th><th>Bachelor's</th><th>Master's or PhD</th></tr>
<tr><td>ES</td><td>249</td><td>9</td><td>46</td><td>24</td><td>12</td><td>11</td><td>17</td></tr>
<tr><td>AR</td><td>89</td><td>20</td><td>39</td><td>9</td><td>14</td><td>20</td><td>1</td></tr>
<tr><td>CO (1)</td><td>133</td><td>18</td><td>65</td><td>8</td><td>9</td><td colspan="2">25</td></tr>
<tr><td>MX</td><td>84</td><td>19</td><td>63</td><td>13</td><td>1</td><td>17</td><td>2</td></tr>
<tr><td>VE (2)</td><td>157</td><td>16</td><td>54</td><td>6</td><td colspan="2">11</td><td>0.5</td></tr>
</table>

Notes:
(1) General population data for Colombia (OECD, 2021) report bachelor's and higher degrees jointly.
(2) Due to lack of benchmarks, data for Venezuela are not limited to the 25-64 age range. General population data have been obtained courtesy of the ENCOVI (2018) project team. Data for all University degrees below master's level are reported jointly.

ILO (2021, p. 141) finds a similar, though less extreme result in a worldwide survey of online platform workers (not limited to those that serve the artificial intelligence industry), with more highly-educated workers in lower-income countries, and explains it as an effect of lack of opportunities in local labour markets. However, relatively few of the micro-workers studied by Margaryan et al. (2020), all Europeans, cite lack of opportunities as their main reason to undertake this activity. The Spain vs. Latin America comparison undertaken here suggests important differences between higher- and lower-income countries. While the first reason for micro-working is 'need of money' for two thirds of respondents (a finding that echoes Casilli et al., 2019, in the case of France), as many as 41% of Latin Americans, against only 22% of Spaniards, indicate that 'the political/economic situation of my country does not allow me to find a job' as their first, second or third reason. The gap widens if Spain is contrasted not to the whole of Latin America, but to the extreme case of crisis-stricken Venezuela, where this answer is chosen by 76% of respondents.

While their education is not formally visible on platforms, do workers make any use of it? To see this, it is useful to leverage data from qualitative interviews, where participants discussed at length their experience, motivations and feelings. Here, misrecognition appears insofar as workers' experience of disrespect leads them to adopt a discourse that frames human-in-the-loop tasks as so simple that they do not require any specific education. Diana, an arts student in Lima, Peru, who uses Microworkers.com intensely to pay her tuition fees and living expenses, says that 'the platform doesn't need a lot of science. It's not that complicated once you get to grips with it'. However, when asked to elaborate more on their activities, interviewees admit that even apparently straightforward tasks do require knowledge and skills. Jack, a twenty-year-old who works for the transport authorities of the Dominican Republic and has been using platforms since he was 13, thinks that good writing and reading comprehension are essential. Also, he stresses, it is necessary to know how to effectively use search engines, and to figure out how to do tasks, to understand what they are about. Rafael, a twenty-year-old Venezuelan who dropped out of University to micro-work full time in order to help his family, explains that a good understanding of the clients and the ways in which tasks serve the artificial intelligence industry helps to gauge customers' goals, and limits the risk that a task might be rejected (and, therefore, not paid). Cecilia, a mechatronics engineering student and former web developer from Quito, Ecuador, who started micro-tasking as the COVID-19 pandemic reduced offline work opportunities, and experimented with a number of different platforms, says that at least an intermediate level of English is necessary.

Likewise, even basic tasks enable skill development processes of which workers are conscious (although in the questionnaire, only 16% report learning goals as reasons for micro-tasking). In interviews, many say that their command of the English language (essential to do tasks that are most often posted by international, especially North American, clients) has improved. Alexis, a Venezuelan computer science student, believes that tasks broadened his English lexicon: 'by reading it more, by writing it more, I understand it more'. English is even a 'challenge' for 30-year-old Daniel from Colombia, who started from a very elementary level and is now proud to have done many tasks in this language. Other areas where many interviewees see progressive improvement fall largely within what Margaryan et al. (2020) call 'skills in being an online worker'. Daniela, an auditor in the Venezuelan oil industry who does pretty basic online micro-tasks to add to her low salary, learned to optimize her use of time:

> With simpler tasks, it's all a matter of time, and if you don't do it quickly the task expires and you lose it, sometimes you did half of the work. With practice you at least know what to expect, you have to have several tabs open and look at the preview of what they ask you to do. Then, before you accept the task, you do something at least, opening the pages they tell you and preparing the task so you can then accept it and finish before time runs out.

Similarly Colombia's Carlos, who started micro-tasking as pandemic-related restrictions prevented him from getting a better job elsewhere, learned to do a wider range of tasks in less time:

> There were tasks that when I started I saw that they were too long and I said no I won't know how to do them, I withdrew. And as time went by I saw the key... and I know, this one is done like this, I do it in less time... And then it is like, every time you are there you learn to do them faster (...) As time has gone by, I've improved a lot.

There is some recognition for these improvements insofar as most platforms publish indicators such as number of tasks executed successfully, sometimes sub-divided by type of task. They often also offer non-remunerated qualification tasks that workers undertake to prove their skills in some area (often language proficiency), and whose results remain attached to their profiles. Doing such qualification tasks increases unpaid working time (and therefore lowers hourly remunerations), but may open the way to future opportunities as clients can use these indicators to propose tasks selectively to workers who meet certain criteria. There is a large cost in terms of time: Ecuador-born, former human resources officer Gabriel, who moved to Spain shortly after turning thirty, says that 'It's work that demands a lot, a lot of time at the beginning, more than eight hours a day'. This statement may seem at odds with a common discourse that emphasizes time flexibility as an advantage of platform labour. But interviews confirm that those who do tasks on the side, occasionally, or intermittently, do not trigger the full effects of experience as a pathway toward learning. Sometimes they have no idea of what tasks are about, what their purposes are, and who demands them.

Another difficulty is that there is no portable recognition of qualifications and experience, and on each platform, workers have to re-build their credentials. Without information on formal education, and with platforms' own qualification systems operating in silos, skills information is fragmented, and misrecognition is systemic.

Engineers in the loop

So far, the evidence concerns only pretty basic or general abilities. One might ask whether workers have, and use, more advanced or specialized skills and competencies, especially those related to the technological, scientific or commercial dimensions of algorithms and artificial intelligence. Table 2 reports data for all age ranges and shows that more than half of the Spanish and Latin American micro-workers have studied a relevant discipline, notably engineering, computer science, economics and business administration.

Workers' specializations are aligned with the demands of the digital industry – regardless of the types of tasks they do. There are proportionally more engineers in Latin America, although interestingly, UNESCO data show that Spanish-speaking South American countries are not those with the highest ratio of new graduates specializing in engineering fields, lagging well behind Singapore, Germany, the United Arab Emirates and other countries (UNESCO, 2022). Again, it is a sign that the artificial intelligence industry accesses relevant and scarce human resources from lower-income countries for the needs of producers and consumers in richer parts of the world.

The technical backgrounds of so many workers suggests that they may be comfortable in human-in-the-loop environments, perhaps motivating them to develop a genuine interest in this activity at least temporarily. Indeed some – particularly numerate – interviewees make a significant personal investment in micro-tasks, even putting to use unsolicited extra skills. Hugo, an engineering student from the Dominican Republic, routinely uses his programming skills to check the coding of tasks before doing them. He proudly related how on one occasion, he alerted a customer to a bug in a task, and the customer thanked him and paid him a bonus in exchange. While such windfall earnings are uncommon, other workers draw on their technical education to give meaning to tasks. Jonathan, a Colombian industrial engineering graduate who sells motorbike parts in his family's business and uses his free time to micro-work, enjoys tasks that consist in testing software applications:

> I download the applications to my mobile phone, test them and send the review... I love that task because I feel that it is somehow close to my profile... because I am an industrial engineer and one of the things that I have done, is to evaluate systems, in a functional way – to approve them and detect faults. I like that part of the job, it has to do with my profession. It's one of my areas.

Table 2: fields of study of human-in-the-loop task workers, in percentage. Top panel: all respondents, bottom panel: only respondents with some higher education (2-year degrees or higher). Interpretation: Among the 259 Spanish micro-taskers who participated in the survey and with higher education, 19.3% have studied economics or business administration. Source: author's elaboration.

<table>
<tr><th rowspan="2" colspan="2"></th><th rowspan="2">n</th><th colspan="7">Fields of study (%)</th></tr>
<tr><th>SSH (1)</th><th>Economics, business</th><th>Computer science</th><th>Engineering</th><th>Natural science, health (2)</th><th>Other</th><th>NA</th></tr>
<tr><td rowspan="2">All</td><td>Lat. Am.</td><td>1138</td><td>19.1</td><td>20</td><td>10.3</td><td>26.9</td><td>14.1</td><td>3.2</td><td>6.3</td></tr>
<tr><td>ES</td><td>339</td><td>25.3</td><td>20.4</td><td>15.2</td><td>16</td><td>13.4</td><td>3</td><td>6.1</td></tr>
<tr><td rowspan="2">Univ.</td><td>Lat. Am.</td><td>890</td><td>19.6</td><td>21.9</td><td>11.1</td><td>29.6</td><td>15.5</td><td>2.2</td><td>1</td></tr>
<tr><td>ES</td><td>259</td><td>29.7</td><td>19.3</td><td>16.6</td><td>18.1</td><td>15.8</td><td>3</td><td>0</td></tr>
</table>

Notes:
(1) SSH: Social sciences and humanities. Includes the arts, humanities, languages, social sciences, political science and law.
(2) Natural science includes mathematics, physics, chemistry, biology and agriculture. Health includes medicine, dentistry, nursing and other health-related fields.

Platform activity often drives workers to sharpen their technical knowledge, notably to understand the platform's algorithm – to figure out, for example, at what times tasks are more likely to be published, what countries get most tasks, how many competitors they face at a given moment, how fast they should work in order to do a maximum of tasks without being misidentified as a bot and thus banned. They learn to distinguish platforms by discrepancies in their user interfaces and the methods through which they allocate or validate tasks. Beyond skills that are instrumental to effectively perform micro-work, some also acquire a broader understanding of machine learning, artificial intelligence, and their key applications such as search engines, voice recognition, and computer vision. Venezuela's Rafael, introduced earlier, knows precisely what he is doing when he annotates street traffic images: 'I teach the car by training it, so to speak, hence... I collaborate with artificial intelligence'. Importantly, they also learn to understand the digital economy and to develop solutions to reap its benefits, sometimes in view of a future professional project. Cecilia, the mechatronics engineer, says that she is 'looking for a way to develop something on the web (...) because my degree is very versatile and we learned to program and develop things on the Internet'.

Ironically, misrecognition is more pronounced here, as there is no formal mechanism to acknowledge workers' increasingly advanced understanding of machine learning and more generally, of digital technologies, through the practice of micro-work. The more technically-savvy workers even risk sanctions if they try to take maximum advantage from their grasp of platforms' algorithms. Indeed, their competencies can threaten a *statu quo* that gives more power to platforms and clients, relative to workers. This is best illustrated by the story of Alonso, a Venezuelan married man in his forties who derives all his income from online tasks. He worries because some platforms have restricted Venezuelans' access to some tasks, as a result of their fear that the massive presence of these workers might crowd out others, reducing the diversity that some customers demand. A female friend of Alonso, a systems engineer, has written a piece of code that gets around the restriction, basically by preventing an automated re-direction of the submission page, and has shared it with him and four other data workers. At the time of the interview (April 2021), the code works well, though not on the mobile version of the platform. But what if this device gets detected? There is no way to recognize the programming competencies of Alonso's friend – who if caught, will be remembered only for breaking the rules.

Difficult micro-tasks

Workers' experience with the relatively complex techniques of image labelling used in the development of, for example, self-driving cars, comes out more neatly in the interviews than in the questionnaire. Indeed almost all interviewees have used multiple platforms and have experience of such tasks. Gerardo, a divorced construction worker in his late forties who lives in the Buenos Aires province in Argentina, characterizes the latter as being about:

> framing objects in a diagram. For example, cars, you have to outline them. They have to fit exactly under the tyre. You have to take into account the edges, the little mirrors on top. So, you have to do it very precisely and sometimes you have a task where there's, like, fifteen cars in a picture. It takes a long time.

He characterizes these tasks as 'difficult' because several qualification tests must be passed before having access to them, and because many technical terms in English are used. Rafael, the above-cited Venezuelan full-time micro-worker, adds that these tasks last several hours, sometimes up to three days. Interviewees often associate difficulty and duration, highlighting again the need to devote time to this activity to reap its benefits. But inherited inequalities may make time a scarce resource, notably in the case of women with care duties. Clara, a young mother from Venezuela, started micro-working after losing her administrative job, and left a platform that offers rather complex tasks for one that specializes in simpler ones: ' I really didn't manage it very well, because you have to dedicate a lot of time and concentration to it, and I really couldn't, because apart from that, I have an 18-month-old baby and I also have to dedicate time to him'.

In this respect as argued in Tubaro et al. (2022), women seem to be as disadvantaged on micro-working platforms as they are on standard labour markets. The platform economy does not significantly alter legacy socio-economic structures, and does not redistribute earning opportunities beyond extant inequalities.

Task duration can also be a problem as a result of infrastructural obstacles – an aspect that sharply distinguishes lower- from higher-income countries. The internet infrastructure is uneven across Latin America, and especially Venezuelans often experience power cuts, slow connection, and obsolete or inadequate computing equipment. Alexander, a former computer technician, says that:

> For these tasks, you need very… very good internet signal. And sometimes I can't do them, because it's complicated, like, there are 3D models of the vehicle and I have to classify different things that I see inside it… It's not so feasible with my internet, which is not so good… sometimes it fails, because of lack of connection.

Task duration involves financial risk because connection or electricity failures may involve loss of the entire work done and therefore, of the right to claim payment. Shorter, simpler tasks that can be submitted immediately significantly reduce potential losses, and are therefore preferred by workers with precarious infrastructural arrangements. The highly competitive environments of micro-working platforms, where tasks are typically allocated to large anonymous masses on a first-come, first-served basis, exacerbate this problem compared to online freelancing platforms in which there is a one-to-one matching between worker and client.

Difficult tasks are also unattractive when they are not better paid. A commonly heard view is that complex tasks are riskier because of higher chances of making mistakes that may result in tasks being rejected – hence not paid – by clients. But sometimes, pay rates do not even differ. According to Jack from the Dominican Republic, 'The more complex the job, the better paid it is. And there are other times when it is not even complex and is well paid. That is according to... the needs of the clients'.

Remunerations follow market laws. The more tasks are open to multiple workers across countries, the more labour supply increases relative to demand (Graham & Anwar, 2019). The massive presence of qualified workers from lower-income countries – in Latin America, mainly from Venezuela – inflates supply and lowers the price of data labour. On the demand side, little effort is made to raise remunerations above market level (although some platforms invite clients to fix pay rates at minimum wage). There is misrecognition, again, when clients deny the complexity of a task, notably by under-estimating the time needed to complete it, as a way to keep remunerations down, driving a wedge between officially announced and actually practised pay rates. Overall, organization of work on platforms keeps many micro-workers out of more complex tasks, and/or underestimates the complexity of the tasks they do, thereby weakening their status and lowering negotiation opportunities.

Conclusions

Today's artificial intelligence, largely based on data-intensive machine learning algorithms, relies heavily on the contributions of invisible and precarious humans-in-the-loop who perform multiple essential functions. They form part of cross-national supply chains in which corporate actors mostly from the Global North access human capital generated, partly, in the Global South. Comparison between workers from the North (Spain) and South (Latin America) shows that they are all highly educated, but that the Latin Americans more often specialize in the engineering disciplines that best serve the international technology industry. The tasks that these workers do are not always as simple and straightforward as received wisdom implies, and the loop is one in which both machines and humans learn. However, the latter are at a disadvantage. Time and infrastructural resources are needed to access more complex tasks, and pay is not always higher. Workers' education is not taken into account, and their skill development is recognized only to a limited extent. The fact that workers themselves often take the rhetoric of simple tasks at face value, and admit the complexities of their activity and their learning process only upon reflection, is indicative of their experience of disrespect, due to widespread misrecognition.

The most obvious practical recommendation that derives from this analysis is to devise ways to recognize skills and competencies, so that workers can leverage their experience in the loop to build a career. Forms of certification could be designed, ideally based on standards shared by the whole industry, so as to be valid across intermediaries and platforms. Taking this seriously may be difficult, though, as recognition of skills may impact pay rates and therefore, increase the costs of human-in-the-loop algorithmic production. This consideration invites to raise the stakes and address the problem at a higher level – that of the transformations of labour that technology is bringing about, the erosion of the salary institution, and the human-machine interactions that populate imaginaries of the future of work. This paper has illustrated how misrecognition of education and skills is one mechanism through which these changes are unfolding, maintaining the balance of power in favour of platforms and clients. If the salary institution had the historical merit of providing social protection to workers spanning very diverse levels of education, the present-day techno-economic configuration paradoxically disenfranchises some of the more educated workers.

As a future perspective, it would be useful to further explore the micro-dynamics of misrecognition in other platformized contexts. Delivery couriers and drivers, who often perceive themselves as doing mere manual work, may actually be experiencing misrecognition of the digital and data-producing activities in which they are involved (Casilli, 2019).

The phenomenon under study may be linked to extant evidence of misrecognition of workers' skills in the informal economy. Recent research insists on the need to recognize informally acquired skills, but also to formally train workers in order for them to access quality jobs in the formal economy – as the only way to promote inclusive growth (Mehrotra, 2018; Palmer, 2017). While the context is very different, both cases involve working arrangements that depart from the model of the salary institution, and miss its key benefits.

In broader perspective, the analysis undertaken here contributes to debates on individual skill-building policies, which are widely believed to be part of countries' essential response to technological change and globalization, and an opportunity to achieve or preserve full employment. While sharing the consensus that there are strong linkages between education, work and social inclusion, concerns have been voiced that the employment-generating power of individual skill improvements is limited (Crouch, 1999; Crouch et al., 2001). The present paper adds further evidence that education, skills, and learning alone do not suffice to guarantee fair working conditions and remunerations in the absence of an adequate institutional, regulatory and organizational framework. The key underlying question is that of the technology we want, of the model of society that should embed it, and of the human price we are ready (or not) to pay for it.

References

Altenried M. (2020). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, 44(2): 145-158. DOI: 10.1177/0309816819899410

Berg J., Furrer M., Harmon E., Rani U. & Silberman M.S. (2018). *Digital labour platforms and the future of work: Towards decent work in the online world*. Geneva: ILO Report. Available at: https://www.ilo.org/global/publications/books/WCMS_645337/lang—en/index.htm, accessed 31 March 2022.

Casilli A. (2019). *En attendant les robots: Enquête sur le travail du clic*. Paris: Seuil.

Casilli A., Tubaro P., Le Ludec C., Coville M., Besenval M., Mouhtare T. & Wahal E. (2019). *Le micro-travail en France. Derrière l'automatisation, de nouvelles précarités au travail?* Paris, Report of the Digital Platform Labor (DiPLab) project. Available at: https://hal.inria.fr/hal-02139528/, accessed 5 January 2022.

Chicchi F. (2020). Beyond the 'salary institution': on the 'society of performance' and the platformisation of the employment relationship. *Work Organisation, Labour & Globalisation*, 14(1): 15-31. DOI: 10.13169/workorgalaboglob.14.1.0015

Cognilytica (2020). Data preparation and labeling for AI 2020. Report CGR-DLP20. Washington DC: Cognilytica.

Crouch C. (1997). Skills-based full employment: the latest philosopher's stone. *British Journal of Industrial Relations*, 35(3): 367-391.

Crouch C., Finegold D. & Sako M. (2001). *Are Skills the Answer? The Political Economy of Skill Creation in Advanced Industrial Countries.* Oxford: Oxford University Press.

Denton E., Hanna A., Amironesei R., Smart A. & Nicole H. (2021). On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society,* 8(2). DOI: 10.1177/20539517211035955

Difallah D., Filatova E. & Ipeirotis P. (2018). Demographics and dynamics of Mechanical Turk workers. *Proceedings of WSDM 2018: the Eleventh ACM International Conference on Web Search and Data Mining*, ACM: 135-43.

Ekbia H.R. & Nardi B.A. (2017). *Heteromation, and Other Stories of Computing and Capitalism*. Cambridge (MA): MIT Press.

ENCOVI (Encuesta Nacional de Condiciones de Vida) (2018). Universidad Católica Andrés Bello. Caracas.

ENCOVI (Encuesta Nacional de Condiciones de Vida). Documento técnico (2021). Universidad Católica Andrés Bello. Caracas.

Graham M. & Anwar M. (2019). The global gig economy: towards a planetary labour market? *First Monday*, 24(4). DOI: 10.5210/fm.v24i4.9913

Gray M. & Suri S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* Boston, MA: Houghton Mifflin Harcourt.

Grohmann R. & Fernandes Araújo W. (2021). Beyond Mechanical Turk: The work of Brazilians on global AI platforms. In P. Verdegem (ed.), *AI for Everyone? Critical Perspectives*, University of Westminster Press, pp. 247-266.

Heuer H., Jarke J., & Breiter A. (2021). Machine learning in tutorials – Universal applicability, underinformed application, and other misconceptions. *Big Data & Society,* 8(1). DOI: 10.1177/20539517211017593

Honneth A. 1995 [1992]. *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Cambridge: Polity Press.

ILO (International Labor Organization). 2021. *World Employment and Social Outlook 2021: The role of digital labour platforms in transforming the world of work.* Report. Geneva: ILO.

Ipeirotis P. (2010). *Demographics of Mechanical Turk.* NYU Working Paper, CEDER-10-01.

Irani L. (2015). The cultural work of microwork. *New Media & Society*, 17(5): 720-739. DOI: 10.1177/1461444813511926

Kässi O., Lehdonvirta V. & Stephany F. (2021). How many online workers are there in the world? A data-driven assessment [version 4; peer review: 4 approved]. *Open Research Europe*, 1:53. DOI: 10.12688/openreseurope.13639.4

Lehdonvirta V., Kässi O., Hjorth I., Barnard H. & Graham M. (2019). The global platform economy: A new offshoring institution enabling emerging-economy microproviders. *Journal of Management*, 45(2): 567-599. DOI: 10.1177/0149206318786781

Lindquist J. (2018). Illicit economies of the Internet: Click farming in Indonesia and beyond. *Made in China Journal*, 3(4): 88-91.

OECD (Organization for Economic Cooperation and Development). (2021). *Education at a Glance 2021*. Report. Paris: OECD.

Margaryan A. (2019a). Comparing crowdworkers' and conventional knowledge workers' self-regulated learning strategies in the workplace. Human Computation, 6: 83-97. DOI: 10.15346/hc.v6i1.5

Margaryan A. (2019b). Workplace learning in crowdwork: Comparing microworkers' and online freelancers' practices. *Journal of Workplace Learning*, 31(4): 250-273. DOI: 10.1108/JWL-10-2018-0126

Margaryan A., Charlton-Czaplicki T. & Gadiraju U. (2020). *Learning and Skill Development in Online Platform Work: Comparing Microworkers' and Online Freelancers' Practices*. CBS Report.

Mavridis P., Gross-Amblard D. & Miklós Z. (2016). Using hierarchical skills for optimized task assignment in knowledge-intensive crowdsourcing. *Proceedings of the 25th International Conference on World Wide Web*, 843-853.

Mehrotra S. (2018). From the informal to the formal economy: Skills initiatives in India. In A. Sakamoto & J. Sung (eds.), *Skills and the future of work: Strategies for inclusive growth in Asia and the Pacific.* Geneva: International Labour Organization (ILO), pp. 364-390.

Miceli M., Schuessler M. & Yang T. (2020). Between subjectivity and imposition: Power dynamics in data annotation for computer vision, *Proceedings of the ACM on Human-Computer Interaction*, 4, CSCW2, 115, DOI: 10.1145/3415186

Mothobi O., Gillwald A. & Schoentgen A. (2018). *What Is The State Of Microwork in Africa? A View from Seven Countries*. Report, Research ICT Africa.

Newlands G. (2021). Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. *Big Data and Society*, 8(1). DOI: 10.1177/20539517211016026

Palmer R. (2017). *Jobs and skills mismatch in the informal economy*. Geneva: International Labour Organization (ILO).

Posada J. (2022, online first). Embedded reproduction in platform data work. *Information, Communication & Society.* DOI: 10.1080/1369118X.2022.2049849

Schmidt F. (2019). *Crowdproduktion von Trainingsdaten: Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge.* Report, Düsseldorf: Hans-Böckler-Stiftung.

Tomaskovic-Devey D. & Avent-Holt D. (2019). *Relational Inequalities: An Organizational Approach*. Oxford: Oxford University Press.

Tubaro P. & Casilli A.A. (2019). Micro-work, artificial intelligence and the automotive industry. *Journal of Industrial and Business Economics,* 46(3): 333-345. DOI: 10.1007/s40812-019-00121-1

Tubaro P., Casilli A.A. & Coville M. (2020a). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, 7(1). DOI: 10.1177/2053951720919776

Tubaro P., Le Ludec C. & Casilli A.A. (2020b). Counting 'micro-workers': societal and methodological challenges around new forms of labour. *Work Organisation, Labour & Globalisation*, 14(1), 67-82. DOI: 10.13169/workorgalaboglob.14.1.0067

Tubaro P. (2021). Disembedded or deeply embedded? A multi-level network analysis of online labour platforms. *Sociology*. 55(5): 927-944. DOI:10.1177/0038038520986082

Tubaro P. & Casilli A.A. (2022). Human listeners and virtual assistants: privacy and labor arbitrage in the production of smart technologies. In *Digital Work in the Planetary Market*, F. Ferrari & M. Graham (eds.), Cambridge (MA): MIT Press, pp. 175-190.

Tubaro P., Coville M., Le Ludec C. & Casilli A.A. (2022). Hidden inequalities: the gendered labour of women on micro-tasking platforms. *Internet Policy Review*, 11(1). DOI: 10.14763/2022.1.1623

UNESCO (United Nations Educational, Scientific and Cultural Organization) Institute for Statistics. (2022). Sustainable Development Goals Data. Available at: http://data.uis.unesco.org/, accessed 7 January 2022.

Wood A.J., Graham M., Lehdonvirta V., Hjorth I. (2019). Networked but commodified: The (dis)embeddedness of digital labour in the gig economy. *Sociology*. 53(5): 931-950. DOI:10.1177/0038038519828906