JUST TECH

Data Work and its Layers of (In)visibility

By: Adrienne Williams and Milagros Miceli

LABOR & ECONOMY

Williams, Adrienne, and Milagros Miceli. "Data Work and its Layers of (In)visibility." Just Tech. Social Science Research Council. September 6, 2023. DOI: doi.org/10.35650/JT.3060.d.2023.

No technology has seemingly steam-rolled through every industry and over every community the way artificial intelligence (AI) has in the past decade. Many speak of the inevitable crisis that AI will bring. Others sing its praises as a new Messiah that will save us from the ails of society. What the public and mainstream media hardly ever discuss is that AI is a technology that takes its cues from humans. Any present or future harms caused by AI are a direct result of deliberate human decisions, with companies prioritizing record profits, in an attempt to concentrate power by convincing the world that technology is the only solution to societal problems.

AI's development so far has been based on the exploitation of workers and users around the world,[1] performing what anthropologist Mary L. Gray and computational social scientist Siddharth Suri call *ghost work*.[2] This term refers to the undervalued human labor utilized to develop and maintain the automation of websites and apps. Ghost work is characterized by on-demand, short-term projects or tasks performed globally by precarized workers through platforms like Amazon Mechanical Turk and specialized companies like Sama. These workers, usually vulnerable people from Asia, Latin America, and Africa, are paid less than $2 per hour to generate and label data that trains AI models. Moreover, users who validate algorithmic outputs or help perfect systems usually do it for free. ¬Ghost work is often outsourced, hidden, or rendered invisible by the tech companies who request it. As Noopur Raval argues, we must ask ourselves how and for whom this work is invisible and what happens when workers are finally seen.[3] In light of these questions, we delve into three nuanced layers of invisibility that pervade data work: the unpaid work performed by users, human workers pretending to be AI systems, and the different forms of exploitation of vulnerable communities globally. Finally, we explore potential avenues for not only rendering this labor visible but also transforming the material conditions under which it takes place.

Users' Unpaid Labor

AI has launched a new era in unpaid labor, one in which everyone participates in building the foundations for new companies and technologies while remaining completely oblivious to their contributions. ReCAPTCHA is an example of a crowdsourcing model that forces users to perform unpaid work to train AI under the guise of providing proof that one is "Not A Robot." This model of having visitors to a website perform small unpaid tasks that support the growth and success of a business has been applied by several companies. Meta, for instance, is famous for leveraging their users' experience to gain profitable data. Most recommender systems embedded in platforms such as Netflix or Spotify ask users to rate their recommendations. Even OpenAI's ChatGPT asks us to give a thumbs up or down to the system's output. Such feedback is used to improve models, which save companies millions in algorithmic verification tasks that are otherwise performed by human workers. Still, the role of users in validating and perfecting AI systems is hardly ever rendered visible or regarded as work within public discourse.

Users of recreational technology, like social media and online games, can appear willing participants. But, when companies use employees, students, or refugees without their consent, forcing a captive audience to build and update their tech without payment: Is that still willing participation or is it forced labor?

> *Infractions from the Mentor app require the completion of unpaid, off-the-clock homework, which constitutes a unique form of wage theft.*

For instance, Amazon delivery drivers are forced to become users of two surveillance technologies while on the job—the Mentor App, which monitors hard stops, aggressive turns, speeding, and other app-determined driving infractions, and the AI Netradyne cameras inside and outside their vans. Both technologies surveil drivers' behavior and detect infractions, which can include, for example, yawning, often penalizing drivers for occurrences far beyond their control. Infractions from the Mentor app require the completion of unpaid, off-the-clock homework, which constitutes a unique form of wage theft. Moreover, drivers have reported that ignoring the homework can result in termination. The behavioral and biometric data collected on drivers is used to train and perfect the same AI technologies surveilling them. Moreover, Netradyne is suspected of using this data to create high-definition 3D mapping for autonomous vehicles. In the words of the company's CEO, David Julian, "Netradyne's Driveri® vision-based driver recognition safety program has captured and analyzed one billion minutes of driving video data and 500 million miles, 3D mapping the most US roads in history." Amazon delivery drivers are not Amazon employees. They work for third-party Delivery Service Providers (DSP). Neither the drivers nor the DSPs who employ them have a voice in how or if they are used to create 3D maps. Amazon drivers are not described as AI trainers on their job descriptions or the Netradyne consent forms they must sign to obtain their jobs. Yet, perform a second, invisible, entirely unpaid job, building the foundation for society's next giant leap forward in vehicle innovation.

Human Labor Masking as AI

Another way in which labor in AI development is rendered invisible concerns the types of tasks data workers are required to perform. While the term ghost work is used mainly in reference to data annotation, which involves data curation, labeling, keywording, and semantic segmentation, the extent of workers' involvement in other types of tasks is relatively unknown to the general public.

For instance, collecting "raw data" for AI is commonly described as scraping publicly available information from the internet. What is often concealed is that, in many cases, data workers are asked to generate the data themselves. Some tasks literally require workers to record their voices reading text passages or upload selfies, pictures of friends and family, or images of rooms or objects in their own homes to enrich datasets.[4] Similar to the case of Amazon drivers, these data workers are not informed about the nature of the systems that will be trained on their data or the privacy consequences of having their picture enrich a training dataset.

> *In reality, what lies behind these cameras is not AI. It's a workforce consisting of highly unstable, underpaid, hidden labor.*

Furthermore, companies that have positioned themselves as being "AI first" often resort to hiring workers to impersonate AI systems,[5] like chatbots. This behavior is usually driven by pressure from venture capitalists to incorporate technologies into their products. For instance, workers in Madagascar are the "algorithm" behind supposed AI-powered smart cameras, as reported by researchers Maxime Cornet, Clément Le Ludec, and Antonio Casilli. These workers monitor CCTV feeds and identify anomalies during shifts that span both day and night. Similar instances of human labor being disguised as AI systems have also been documented concerning workers in Syria.[6] In reality, what lies behind these cameras is not AI. It's a workforce consisting of highly unstable, underpaid, hidden labor.

Using data workers instructed to "think and act like machines"[7] is not rare, nor is it a result of flawed technology. It is part of the business design of many start-ups. In fact, a study conducted by a London-based investment firm discovered that despite being described as "AI-focused," there was no evidence of artificial intelligence applications in 40 percent of the 2,830 AI start-ups surveyed in Europe, which further emphasizes the prevalence of human labor being concealed behind the façade of AI.[8]

Tech Labor Conditions in the Global South

Like many other industries before, tech corporations frequently exploit workers from the Global South, capitalizing on lax regulations and conveniently disregarding workers' safety and well-being to maximize profits. They seize opportune moments, such as times of war, natural disasters, and economic crises,[9] to prey on vulnerable people who have little option but to comply with and accept the unfavorable conditions imposed on them. Among these is assigning tasks that workers in Silicon Valley would

otherwise refuse at the fraction of what a Global North worker would get. These tasks often inflict severe damage on mental health including anxiety, depression, PTSD, and even suicide. These conditions are detrimental to the overall well-being of workers, perpetuating a cycle of exploitation.

Investigations have uncovered a wide range of concerning issues within AI supply chains, including payment discrepancies and challenging conditions. Such reports have exposed the stark differences in tech labor conditions between the Global North and South. For example, contractors in the United States received a wage of $15 per hour for their contributions to powering ChatGPT, a figure considered very low in many places of the Global North. In stark contrast, workers in Kenya, who perform similar tasks for the same client, were compensated with less than $2 per hour. What makes these conditions even more alarming is that the work outsourced to Kenya involved the laborious task of confronting and sorting through "toxic" data, including graphic content like child sexual abuse, murder, suicide, and torture.

Despite the glaring evidence of exploitation,[10] a persistent myth among those who commission these tasks is that a meager hourly wage of $2 constitutes fair compensation in countries like Kenya, Syria, or Venezuela, and paying more would disrupt local economies. The truth remains that miserable wages are inherently unjust, irrespective of geography. Data workers in the AI supply chain are paid barely enough to survive,[11] trapped in a perpetual struggle to make ends meet. These wages offer workers no means to plan for their futures or provide their children with educational opportunities, perpetuating the cycles of poverty that keep workers vulnerable.[12]

What Comes after Visibility?

The prevalence of labor exploitation in AI supply chains has garnered significant press attention this year, leading to a necessary shift in the narrative surrounding invisible forms of data work and the individuals who bear the brunt of its consequences. News reports have not only brought to light the stark disparities in work conditions experienced by workers worldwide but have also highlighted the varying visibility of certain issues and geographic regions.

But visibility is certainly not enough. Efforts must be directed not only toward exposing injustice but to implementing tangible measures to rectify and repair. The greatest myth of "Us vs. Them" is that there is a "Them." When communities are harmed, it is always "Us" being harmed. Ignoring the role power and greed play when discussing labor ignores the mechanism creating vulnerability. Greed is the motivation behind automating as many jobs as possible and unnecessarily hoarding resources. Greed is just an addiction to wealth and power. If corporate leaders can get away with it in one industry, they will do it in every industry. Any industry could be next on the chopping block. We are in this together. We should never assume to be so untouchable that we do not picture ourselves on the other side of the fence, in the shoes of and identifying as "the vulnerable."

> *Fair wages should be based on the nature of the work performed and the value it generates rather than exploiting geographical disparities as an excuse to underpay and perpetuate inequality.*

To start thinking of just data work, we must recognize our contributions as users for what they are: labor. This includes transparent information on how our data is utilized, the opportunity to seek fair compensation for our work, and the ability to opt in or out of assisting in AI development. Additionally, ensuring adequate working conditions in data work is essential. Fair wages should be based on the nature of the work performed and the value it generates rather than exploiting geographical disparities as an excuse to underpay and perpetuate inequality. Internationally recognized labor standards and regulations must be enforced. Not doing so will cause irreparable physical, mental, and financial harm to our world's working poor. Fairwork and Turkopticon are examples of organizations working on these solutions.

This transformative journey toward worker and user protection requires a collective effort. It is important to talk about "invisible" work. But for sustainable change, we must establish mechanisms that enable people to voice their concerns, seek protection, and actively participate in decision-making processes that affect their lives. These are the principles that guide the work of several organizations, including the DAIR Institute. Supporting organizers who advocate for workers' and users' rights can dismantle the culture of silence, creating an environment where exploitation is not tolerated. Continuous education campaigns must be implemented so the public is knowledgeable about how AI supply chains operate and how much a users' and data worker's contribution is actually worth. By fostering a collective understanding that the often-invisible labor of individuals in the AI supply chains contributes significantly to the success and profitability of the companies, we can pave the way for a future where data workers no longer suffer in silence.

Footnotes

1 Adrienne Williams, Milagros Miceli, and Timnit Gebru, "The Exploited Labor behind Artificial Intelligence," *Noema*, October 13, 2022.

2 Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (New York: Harper Collins, 2019).

3 Noopur Raval, "Interrupting Invisibility in a Global World," *Interactions* 28, no. 4 (July–August 2021): 27.

4 Milagros Miceli and Julian Posada, "The Data-Production *Dispositif*," *Proceedings of the ACM on Human-Computer Interaction* 6, no. CSCW2 (2022): 1–37.

5 Paola Tubaro, Antonio A. Casilli, and Marion Coville, "The Trainer, the Verifier, the Imitator: Three Ways in which Human Platform Workers Support Artificial Intelligence," *Big Data & Society* 7, no. 1 (2020).

6 Milagros Miceli et al., "Documenting Data Production Processes: A Participatory Approach for Data Work," *Proceedings of the ACM on Human-Computer Interaction* 6, no. CSCW2 (2022): 1–34.

7 Milagros Miceli and Julian Posada, "Data-Production *Dispositif*."

8 Tubaro, Casilli, and Coville, "The Trainer, the Verifier, the Imitator."

9 Karen Hao and Andrea Paola Hernández, "How the AI Industry Profits from Catastrophe," *MIT Technology Review*, April 20, 2022.

10 Williams, Miceli, and Gebru, "Exploited Labor."

11 Miceli and Posada, "Data-Production *Dispositif*."

12 Julian Posada, "Embedded Reproduction in Platform Data Work," *Information, Communication & Society* 25, no. 6 (2022): 816–834.