

Volume 6 \ Issue 1 \ 05-19-2026

DOI 10.34669/WI.WJDS/6.1.6 \ ISSN 2748-5625

Licensed under Creative Commons Attribution 4.0 (CC BY-SA 4.0)

VOICES FOR THE NETWORKED SOCIETY

Challenging the Myth of AI Autonomy

The Convenient Fiction of Autonomous Intelligence

Uma Rani^{*1}  \ **Morgan Williams**¹ 

¹International Labour Organization (ILO), Geneva, Switzerland

*Corresponding author: amara@ilo.org

ABSTRACT

A common misconception is that artificial intelligence (AI) functions magically, without human intervention. In reality, AI operates within a socio-technical system reliant on vast amounts of human labor throughout its lifecycle. AI depends on two forms of human labor: “algorithmic worker” (the workers responsible for coding and fine-tuning models) and “data worker” (those responsible for labeling, cleaning, and expanding datasets to train AI models). Both are crucial, but data worker often remains invisible to the end-user, leaving workers vulnerable to decent work deficits. This piece examines how these two forms of labor are interconnected and discusses the working conditions of data workers in India and Kenya based on recent ILO surveys conducted in 2022–23. Finally, it explores measures for enhancing transparency and accountability in AI development to ensure that this often-undervalued work and these invisible workers receive the recognition they deserve.

KEYWORDS

Artificial Intelligence \ platforms \ platform work \ data work \ business process outsourcing \ AI regulation

1 Introduction

The prevailing narrative of artificial intelligence is one of technological magic, of autonomous systems that think, learn, and act independently of human intervention through sophisticated engineering. This narrative is not merely a marketing tool; it is a fundamental misunderstanding of the technology's lifecycle. In reality, the "intelligence" in AI is not a purely technological phenomenon. It is a system of *heteromation*, a term coined by Ekbia and Nardi (2018) to describe the complex interplay in which automated systems rely on a vast, often invisible and sustained substrate of human labor to function.

From the initial data collection and cleaning for fine-tuning large language models (LLMs) to the real-time remote assistance of "autonomous" vehicles, human input is the lifeline of the AI ecosystem. The AI lifecycle is supported by an invisible infrastructure comprised of two interconnected, but vastly different, forms of labor: algorithmic work, that is, the programming, model tuning, and debugging performed by data scientists, and data work, that is the repetitive, often grueling tasks of tagging images, transcribing audio, and filtering graphic content. Yet, only the former is celebrated, prestigious, and highly compensated.

The industry is built on a stark hierarchy that idealizes the "algorithmic worker" while making the "data worker" entirely invisible. To build a future that is both ethically sound and technically reliable, we must dismantle the techno-centric myth and recognize AI as a socio-technical system sustained by a marginalized global workforce, which is worthy of recognition. To grasp the depth of human reliance and human hands behind the development of AI tools and technologies, specific examples are needed:

- \ LLMs: The success of systems like ChatGPT is often attributed to raw computing power. However, the lifecycle of an LLM involves intensive human intervention. In the pre-training phase of LLM development, software developers design the model's underlying architecture and language capabilities, and massive datasets are scraped from the internet, requiring extensive cleaning to remove bias, misinformation, and harmful content (Huyen, 2023). Because automated filtering tools often struggle to detect nuanced violations, human content moderators are essential for refining these datasets. In this initial stage, the LLM's performance and safety are directly determined by the quality of the human labor used to curate its training data. During the supervised fine-tuning phase, human workers must craft "ideal" responses to complex prompts to teach the model how to behave. This is followed by RLHF (*reinforcement learning from human feedback*), where annotators rank multiple AI-generated responses based on subjective preferences to train the model to act more humanlike (Ouyang et al., 2022). The process is thus inherently human; it embeds the cultural biases, linguistic nuances, and personal values of the annotators into the model's core (Barnhart et al., 2025).

- \ \ Autonomous vehicles (AVs): The development of AVs relies on a complex synergy between sophisticated algorithmic work and intensive data work, challenging the popular image of a self-driving car navigating a city independently. While algorithmic work provides the vehicle's "logic," that is, the decision-making, path planning, and motion-control systems, these functions are ineffective without the annotation work performed by data workers. Human annotators label massive datasets, identifying critical objects like pedestrians and road signs to create the "golden datasets" necessary for safe navigation (Wang et al., 2022). Consequently, the reality of AV functionality relies heavily on a hidden, labor-intensive human infrastructure that includes both real-time remote assistance and the large-scale outsourcing of data labeling to lower-cost countries, proving that "full" autonomy is still deeply dependent on human intervention (Andersson et al., 2024; Parr et al., 2024).
- \ \ Content moderation: AI-driven content moderation is perhaps the most disturbing example of hidden labor. While algorithms can flag keywords, they struggle with satire, cultural nuance, and intent. Human moderators must fill this gap, viewing thousands of hours of graphic violence, self-harm, and hate speech to "clean" the digital environment for others (Gorwa et al., 2020). The work is psychologically harmful in itself, but these effects are exacerbated by the "algorithmic management" systems, as they often track every second of a worker's activity and leave no room for them to process the trauma they are experiencing.
- \ \ Medical diagnostics: Medical diagnostics is a high-stakes field of AI medical imaging, and the reliability of AI-assisted medical diagnostics is fundamentally determined by the quality of human labor throughout the system's lifecycle, from initial data labeling to clinical deployment. Even after an AI is trained, human oversight is essential (Patel et al., 2019; Bodén et al., 2021; Rädtsch et al., 2023). Medical professionals must validate AI recommendations to ensure that they align with clinical judgment. This "human-in-the-loop" system creates a vital feedback loop but comes with specific challenges, such as "automation bias," where clinicians might become over-reliant on the software (Patel et al., 2019; Bodén et al., 2021). Although expert annotation is essential for creating accurate diagnostic tools, cost-cutting measures often lead to the outsourcing of data work to crowd-workers without specialized medical training (Rädtsch et al., 2023). This practice reduces accuracy and can lead to "data cascades" and, subsequently, to misdiagnosis and a direct risk to patient safety due to the crowd-workers' lack of expertise and incentives for speed over precision.

The term "data cascades" was first introduced by Sambasivan et al. (2021) to describe how upstream data quality issues ripple through an AI pipeline, compounding into issues that undermine the reliability of the AI tool, as seen in the medical diagnostics example. In this process, the work carried out by both algorithmic and data workers is distinctly responsible for shaping the quality of the AI tools. Therefore, maintaining decent working conditions for all workers along the AI supply chain is integral to developing trustworthy AI systems.

2 Working Conditions for Data Workers

Nevertheless, existing research highlights that data work, which is typically performed on digital labor platforms and through business process outsourcing (BPO) companies, is frequently outsourced to countries in the Global South, such as Kenya, India, and the Philippines (Casilli, 2021). There, lower labor costs and weaker institutional protections create a “decent work deficit.” This workforce is quite well educated. For instance, results from ILO surveys conducted in 2022-23 in India and Kenya show that approximately 55–56% of data workers in India and up to 50% in Kenya possess science, technology, engineering, and mathematics education. Many hold university degrees or even PhDs, yet they find themselves trapped in a cycle of “deskilling,” where their advanced cognitive abilities are reduced to mechanical validation, annotation, or content moderation tasks. This points to a systematic underutilization of human capital. This skill mismatch not only hinders the socioeconomic development of these countries through a potential “brain drain” but also introduces ethical risks, for instance, when workers without formal medical training are tasked with annotating high-stakes medical data.

ILO survey data from 2022-23 indicates that despite data workers’ high levels of qualifications, their working conditions in India and Kenya are characterized by a profound gap between formal employment status and actual job security. BPO workers typically hold formal contracts, unlike microtask platform workers who are classified as independent contractors, but these contracts often fail to protect them from arbitrary dismissals or unexplained wage deductions. In Kenya, despite high contract rates, employment is frequently tied to short-term client projects, leaving workers in a cycle of precariousness. Furthermore, gender disparities persist: in India, women are less likely to have formal contracts than men, whereas the opposite is true in Kenya.

Financial compensation in the sector is often lower than expected. Surprisingly, BPO wages tend to be comparable to, or even lower than, the piece-rates found on globally competitive microtask platforms. For instance, Indian platform workers earn significantly more on microtask platforms (US\$3.9 per hour) than BPO workers (US\$2.2 per hour), while Kenyan workers in both sectors earn US\$1.10 per hour. These low wages, which are typically supplemented by stressful performance-based incentives and bonuses for night shifts, force workers to prioritize speed over accuracy. This economic pressure contributes to “data cascades,” where the quality of AI training data is compromised by workers who are trying to earn a decent living wage.

Even when workers hold formal full-time positions, access to basic benefits and social security remains remarkably limited. In both countries, a significant portion of BPO workers lack access to paid holidays, sick leave, and maternity benefits. Social security coverage is equally sparse: in India, over half of the surveyed BPO workers have no coverage at all, while insurance for workplace injuries or retirement is rare in Kenya. Even when medical insurance is provided, it is often restrictive and paid out-of-pocket, leaving low-earning workers vulnerable, especially in the event of an illness or injury.

The physical and psychological demands of the work further erode workers' well-being. Data workers in both countries average over 40 hours per week and often perform rotational shifts that include regular night work. Content moderators, in particular, face extreme pressure, with some expected to process one task every 50 seconds. This high-intensity environment is compounded by the psychological trauma of reviewing graphic and violent content (Ahmad and Krzywdzinski, 2022; Cherry, 2016; Roberts, 2014). In Kenya, this has led to reports of severe health issues, such as chronic insomnia and panic attacks. Compounding these risks, the threats imposed in non-disclosure agreements can restrict workers' access to independent healthcare, forcing them to rely only on company-approved doctors. Much like the work itself, the poor working conditions of data workers remain largely invisible to end-users of AI tools. Therefore, it is imperative to increase transparency requirements for AI developers and deployers in the development of AI.

3 AI Development in a Regulatory Vacuum

While global regulatory frameworks increasingly focus on the ethical deployment and safety of AI outputs, the development phase remains largely overlooked, especially the working conditions of those who build these systems. Most current AI legislation prioritizes the protection of the end-user while largely ignoring the "upstream" human labor required for data annotation. This creates a regulatory vacuum where exploitative practices and psychological risks go unchecked. The disparity between strict rules for AI application and the lack of oversight for AI production means that although the technology is highly regulated in the market, the workers in the Global South who train these models often remain invisible to regulators. This lack of oversight incentivizes a "race to the bottom" in wages and working conditions, as the focus stays on the final product rather than the precarious human labor required to clean and categorize the data that powers it.

In these environments, workers are often compelled to sign non-disclosure agreements that prevent them (particularly content moderators) from seeking independent mental health support, effectively "locking" the human cost of AI within corporate silos (Blackwell, 2025). Further, the regulatory vacuum regarding cross-border digital labor and the absence of regulations governing digital work contracts allows for the unchecked use of non-disclosure agreements, which privatize the human cost of AI development and insulate corporations from liability for worker well-being.

To address these systemic failures, a new regulatory paradigm is required, similar to "fair trade" labels in agriculture or "conflict mineral" disclosures in electronics. AI companies should be mandated to disclose the origins of their training data and undergo independent audits of working conditions at BPO centers and microtask platforms. If part of an AI model is branded as "trained in Kenya," the company developing the AI tool should be required to provide evidence of living wages and safety standards. This is particularly important for content moderation, as workers face severe psychological trauma, including insomnia and panic

attacks. Regulations must establish legal limits on exposure to traumatic content and mandate 24/7 access to qualified, independent mental health professionals. Protections should extend for at least 2 years post employment to account for the delayed onset of post-traumatic stress disorder.

Furthermore, regulators should prohibit “inhumane quotas” or performance metrics that prioritize processing speed over human well-being. Since workers are often pressured to label data or content even when it is unclear, there is a critical need for “human-in-the-loop” standards that ensure that workers have the agency to flag ambiguous data without facing automatic pay deductions or quality-score penalties. In high-stakes sectors, like medical diagnostics and autonomous vehicles, authorities such as food and drug administrations or transport regulators should require a “data provenance” certification. This would guarantee that the individuals labeling the data are appropriately trained and fairly compensated and work in conditions that are conducive to the high levels of accuracy these fields demand.

However, the ethical concerns associated with AI development extend beyond labor conditions to the very data being processed. LLMs that rely on vast swathes of publicly available data without explicit consent from individuals or those who have copyright materials have created significant risks, including the potential exposure of individuals’ locations, contact details, and familial connections (Rani & Dhir, 2024). As Gal (2023) notes, there are currently few standardized procedures to ensure that AI companies do not permanently store sensitive personal information or that they delete it upon request. These concerns have already led to regulatory friction, and the AI companies face a growing wave of litigation regarding intellectual property, with authors alleging that their copyrighted works were utilized to train LLMs without authorization or compensation (David, 2023).

Compounding these privacy and legal issues is a pervasive accountability gap rooted in the “black box” nature of AI algorithms. Because these learning models are often shielded from scrutiny, it is rarely possible to evaluate them for inherent biases or to understand the logic behind their outputs. In some cases, the operations of these systems are so complex that even the companies responsible for their design cannot fully explain their internal mechanics (Schoenherr, 2023). This lack of transparency makes it difficult to hold developers responsible for the societal or ethical impacts of their technology.

Ultimately, addressing these emerging challenges requires a shift from viewing AI as a mere “software product” to recognizing it as a complex socio-technical system. For such a shift to occur, both top-down regulation and bottom-up organized resistance are necessary. As demonstrated by the Writers Guild of America strike (Silberling, 2023) and the launch of the Global Trade Union Alliance of Content Moderators in Nairobi, collective bargaining and unified worker power remain the most effective tools to ensure decent working conditions and ethical oversight of the global AI industry.

Disclaimer

This piece is based on Uma Rani, Morgan Williams and Nora Gobel (forthcoming 2026) “The human cogs in the AI machine: Exploring decent working conditions in the AI-Driven BPO sector in India and Kenya” in Research Handbook on Decent Work edited by Ishbel McWha-Hermann, Christian Yao, and Noelle Donnelly, Edward Elgar Publishing Ltd.

The authors would like to acknowledge that the views expressed or conclusions drawn in this piece represent the views of the authors and do not necessarily represent ILO views or ILO policy. The views expressed herein should be attributed to the authors and not to the ILO, its management, or its constituents.

References

- Ahmad, S., & Krzywdzinski, M. (2022). Moderating in obscurity: How Indian content moderators work in global content moderation value chains. *In Digital work in the planetary market* (pp. 77–95). Cambridge, MA, Ottawa: The MIT Press, International Development Research Centre.
- Andersson, J., Rizgary, D., Söderman, M., & Vännström, J. (2024). Exploring remote operation of heavy vehicles—findings from a simulator study. *Human-Intelligent Systems Integration*, 6(1), 15-24.
- Barnhart, L., Bafghi, R. A., Becker, S., & Raissi, M. (2025). Aligning to what? limits to RLHF based alignment. [arXiv preprint arXiv:2503.09025](https://arxiv.org/abs/2503.09025).
- Blackwell, L. (2025). Content Moderation Features. [arXiv preprint arXiv:2509.09076](https://arxiv.org/abs/2509.09076).
- Bodén A.C.S., Molin J., Garvin S., West R.A., Lundström C., & Treanor D. (2021). The human-in-the-loop: an evaluation of pathologists’ interaction with artificial intelligence in clinical practice. *Histopathology*, 79(2): 210-218. <https://doi.org/10.1111/his.14356>
- Casilli, A. (2021). “Waiting for Robots: The Ever-Elusive Myth of Automation and the Global Exploitation of Digital Labor”. *Sociologias* 23 (57): 112–33.
- Cherry, M. A. (2016). Virtual Work and Invisible Labor. In M. G. Crain, W. R. Poster & M. A. Cherry (Ed.), *Invisible Labor: Hidden Work in the Contemporary World*. University of California Press.
- David, E. (2023, September 20). *George R.R. Martin and other authors sue OpenAI for copyright infringement*. The Verge. <https://www.theverge.com/2023/9/20/23882140/george-r-r-martin-lawsuit-openai-copyright-infringement>

- Ekbia, H., & Nardi, B. (2014). Heteromation and its (dis) contents: The invisible division of labour between humans and machines. *First Monday*, 19(6).
- Gal, U. (2023, February 8). *ChatGPT is a data privacy nightmare. If you've ever posted online, you ought to be concerned. The Conversation*. <https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 2053951719897945. <https://doi.org/10.1177/2053951719897945>
- Huyen, C. (2023, May 2). RLHF: Reinforcement Learning from Human Feedback. *Chip Huyen*. https://huyenchip.com/2023/05/02/rlhf.html#rlhf_and_hallucination
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback* (arXiv:2203.02155). arXiv. <https://doi.org/10.48550/arXiv.2203.02155>
- Parr, H., Harvey, C., Burnett, G., & Sharples, S. (2024). Investigating levels of remote operation in high-level on-road autonomous vehicles using operator sequence diagrams. *Cognition, Technology & Work*, 26(2), 207-223.
- Patel, B. N., Rosenberg, L., Willcox, G., Baltaxe, D., Lyons, M., Irvin, J., Rajpurkar, P., Amrhein, T., Gupta, R., Halabi, S., Langlotz, C., Lo, E., Mammarrappallil, J., Mariano, A. J., Riley, G., Seekins, J., Shen, L., Zucker, E., & Lungren, M. P. (2019). Human-machine partnership with artificial intelligence for chest radiograph diagnosis. *Npj Digital Medicine*, 2(1), 1-10. <https://doi.org/10.1038/s41746-019-0189-7>
- Rädsch, T., Reinke, A., Weru, V., Tizabi, M. D., Schreck, N., Kavur, A. E., Pekdemir, B., Roß, T., Kopp-Schneider, A., & Maier-Hein, L. (2023). Labelling instructions matter in biomedical image analysis. *Nature Machine Intelligence*, 5(3), 273-283. <https://doi.org/10.1038/s42256-023-00625-5>
- Rani, U. & Dhir, R. K. (2024). AI-Enabled business model and human-in-the-loop (deceptive AI): Implications for labour. In M. Garcia-Murillo, I. MacInnes, & A. Renda (Eds.), *Handbook of artificial intelligence at work*. Edward Elgar Publishing.
- Roberts, S. T. (2014). *Behind the screen: The hidden digital labor of commercial content moderation*. University of Illinois at Urbana-Champaign.

Schoenherr, J. R. (2023, March 5). *Generative AI like ChatGPT reveal deep-seated systemic issues beyond the tech industry*. The Conversation. <https://theconversation.com/generative-ai-like-chatgpt-reveal-deep-seated-systemic-issues-beyond-the-tech-industry-198579>

Silberling, A. (2023, September 27). *The writers' strike is over; here's how AI negotiations shook out*. TechCrunch. https://techcrunch.com/2023/09/26/writers-strike-over-ai/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAMrSXZPa2PmXAlINiF-HmYLIvD3vLA0Cd9duF_nJYeKGQ3KiA-oq-gE-zhhyVQpZbJUK4E9QqwOFRxt4wh0FkwSCeEIfw_UdTkn6ib79NZu8ji5eCX6FW3Fs6do-VeESbWoZ4uUU6y6rBKwZCsFFei-fPkKF_w5sMYBudRFRYk9rei

Wang, D., Prabhat, S., & Sambasivan, N. (2022). *Whose AI Dream? In search of the aspiration in data annotation* (arXiv:2203.10748). arXiv. <https://doi.org/10.48550/arXiv.2203.10748>

Date accepted: 16 March 2026