Platform Authority and Data Quality: Who Decides What Counts in Data Production for Artificial Intelligence?

Julian Posada
Yale University*

I. Platform Uses and Abuses

Recent applications of OpenAI's Generative Pre-trained Transformer (GPT) family of large language models that predict word sequences have sparked debate on the potential uses and abuses of a technology that produces plausible text on demand. The technology has been tested academically[1] and via court judgments,[2] and questions have been raised about the content and quality of the artificial agent's output and the dangers of reproducing societal harm, as technologies such as facial recognition have done in the past.

In developing its products, the technology industry follows a "big is better" model, thinking that the more data models are trained on, the more accurate their output will be. However, additional training for data models also drives increased computational and human resources.[3] The latest of OpenAI's language models, GPT-4, and its applications, such as ChatGPT, use billions of parameters.[4] These data were scraped from easily accessible internet sites and platforms, but this approach raises concerns over potentially harmful content in the source material.

OpenAI has faces the same problem as many other companies producing data-intensive technologies. As the current paradigm of artificial intelligence (AI) requires large amounts of data, it is unclear how companies can ensure the quality for their products. In other words, how can they prevent their technology from reproducing societal harms? Many companies have hired workers to address this issue. Social-media platforms such as Facebook have hired workers to take down specific posts, and more recent AI companies have hired them to generate and transform their datasets. Many companies see outsourcing labor to lower-income countries as a cost-effective solution that reduces production costs. However, in doing so, they are once again prioritizing data quantity over its quality.

I focus on the relationship between labor and data quality, especially in instances where the generation, annotation, and verification of data is outsourced through digital platforms. My main argument is that higher quality data requires better working conditions because engaged employees whose labor rights are respected provide feedback for improving data quality. In the first part of this paper, I explain the significance of labor in producing data for AI. Then, I will discuss how the industry conceives of the epistemic problem of data production and how the power imbalances in platform labor play a role in this process. The final part of the paper presents recommendations for circumventing epistemic authoritarianism in data production and increasing the quality of data produced through labor outsourcing.

II. Outsourced Labor as the Hidden Ingredient for Artificial Intelligence

In 2005, Amazon launched Mechanical Turk, the first major outsourcing platform, as a form of "artificial AI" intended to distribute tasks related to data production.[5] Its name comes from an 18th-century automaton that seemed capable of playing chess but was in fact controlled by a human concealed inside.[6]

This platform, and those that came later, responds to the need of AI techniques such as machine learning for data and evaluation. Supervised learning requires labeled data, and reinforcement learning requires evaluation. The technology industry, therefore, relies on humans to provide data, annotate it, and verify algorithm outputs.[7]

The need for contemporary technology companies to reduce production costs pushes many to rely on business process outsourcing (BPO) companies or digital platforms for their data work.[8] The former is not the focus of this paper but is worth mentioning. BPO companies are popular mainly for content moderation and algorithmic-verification tasks, and they provide physical infrastructure and workspaces for their employed data workers. One example is the company Sama (previously named Samasource), which employs workers in Kenya and counts OpenAI among its clients.[9] Data-production BPO companies are located all over the world, including India and the Philippines,[10] Argentina and Bulgaria,[11] and the US.[12]

Platforms are primarily headquartered in countries with advanced economies but hire workers from around the world and specialize in different aspects of data production. Some are internal to major technology companies. In addition to Amazon's Mechanical Turk, there are Google's Raterhub and Microsoft's Universal Human Relevance System. Other major players in this ecosystem include Australia's Appen, Canada's Telus.ai, Germany's Clickworker, and the US-based Scale, which recruit workers and provide them with annotation tasks from other platforms, such as those offered by Google and Microsoft.

Workers on these platforms perform four types of tasks, as originally described by Tubaro et al.[13] and detailed by Miceli and Posada[14] in our analysis of over 280 task instructions received by workers. First, platforms provide data *generation*, where workers perform tasks ranging from inputting data to capturing photos, videos, and sound recordings of their surroundings. For example, workers can be tasked with taking photos of themselves in certain positions to train an algorithm to identify them. Second, platforms provide data *annotation*, where workers categorize and give meaning to data. One common task workers perform for autonomous vehicles, for example, is to identify bodies, such as pedestrians, buildings, and other vehicles, that can be encountered while driving. Third, platforms provide *evaluation* of algorithmic outputs by, for example, having workers moderate data for ChatGPT.[15] Fourth, platforms provide *impersonation* of artificial agents, as in the observed case of a worker impersonating a chatbot for a major social-media company.

The persistence of labor in creating and regulating autonomous agents poses several questions. First, do the many workers involved in the data-production pipeline work under decent conditions? Second, are there high standards of data security and privacy, particularly when data are transferred among different users globally? Third, does the data-production process prevent the propagation of harmful content, especially when data generation, annotation, and verification create meanings or ways of seeing the world that will later be distributed by algorithms? In the next section, I will explore the interrelations between labor rights, data-security standards, and harm avoidance, arguing that companies cannot achieve high-quality data without paying close attention to the social processes involved in producing data.

III. The Ground Truth Problem and Platform Power

When reading dozens of instruction documents for data work, my research team and I were struck by how many of them included managerial elements constantly reminding the workforce that, if they did not perform the tasks according to the clients' design, they would be banned or expelled from the project. In interviews, workers led us to realize that, though most of the tasks were easy and straightforward, some generated disagreement. For example, when moderating social media posts, a Latin American worker disagreed that anti-immigration rhetoric advocating the removal of all immigrants from the US should be protected under freedom of speech. We also encountered complex issues, such as determining the boundaries for "adult content" and issues related to the racial and sexual classification of humans. Thus far, companies can largely self-govern. Although they follow some legal and commercial guidelines, classification decisions are almost left entirely to their discretion.

In managing workers algorithmically and including threats in their instructions, platforms aim at reducing worker bias, a particular problem the data-production industry faces when distributing tasks around globally. Managerial algorithms ensure that results do not deviate from what clients consider correct information or *ground truth*, a term used in the computation and information field. However, as we observed, data classification, especially human and social classification, is subjective and potentially contestable and harmful. Even seemingly straightforward classifications, such as a person crossing the street in an image for autonomous vehicles, can have different labels. Companies tend to classify them as "pedestrians," but such a generic label could preclude manufacturers and their vehicles from considering the particular needs of people who could be labeled as a "child," "person in a wheelchair," or "elderly person."

Managerial algorithms try to reduce worker bias by imposing specific conceptions of ground truth; they also try to reduce risks arising from alienating workers and discouraging feedback. One key difference between smaller data production in BPO companies, which generate data with in-house labelers, and larger platforms, which generate data with freelancers, is that BPO worker engagement and feedback reduce errors and improves data quality and security.

A recent article in *MIT Technology Review* reported that workers in Venezuela hired by US-based Scale AI leaked photos of individuals in private settings, such as in their home bathrooms, taken by development versions of the Roomba, iRobot's robot vacuum cleaner.[16] The company equips these robots with a camera for visualizing their surroundings, and photos from test sites were sent to data workers so they could label the objects in houses. I documented a potentially related episode in which workers were not told what the images they were working on were for and flocked to unmoderated forums on social media to denounce and comment on potential privacy concerns without risking retaliation from their employers.[17]

The fear of being "deactivated" or "banned," terms many platforms use for dismissing a freelance worker, is constant among the dozens of workers we interviewed in Latin America. As platforms' operations are largely unregulated due to the international nature of their transactions, governments and clients do not necessarily compel them to abide by labor laws and regulations. Workers, who are usually located in low-income countries and paid a few cents per task, can be fired without recourse or explanation. Some platforms, such as Australian-based Appen and Canadian-based Telus.ai, are starting to implement contracts with some workers, but this practice is far from being the norm in the sector.

My research project on the platformization of data production has led me to conclude that labor rights and data quality are interrelated. By *quality*, I mean the capacity of data to yield insightful actionable outcomes without reproducing societal harms. Thus far, the industry has relied on underpaid and exploited workers to cost effectively produce data. Companies then utilize managerial algorithms to reduce this alienated workforce's "bias," but those algorithms reproduce the "ground truth" (i.e., the bias) of particular clients and risk the security of the data, which can carry sensitive personal information. In the next section, I present some actionable recommendations to the issue of platform authority.

IV. Recommendations for Data Quality

Advanced AI systems are continuing to perpetuate biases and societal harms, which makes the quest for high-quality data urgent. High-quality data can be achieved in many ways, but here I will underline three methods that relate to data work: ensuring fair-work principles are respected throughout the data-production pipeline; engaging a variety of voices, including those of workers, in AI development; and supporting worker-oriented enterprises.

1. "Labor is not a commodity" is the founding principle of the UN's International Labour Organization. Yet, the rise of the gig economy has enabled the unparalleled commodification of work across myriad sectors, including data production. As decent work is one of the UN's development goals, data-based technologies cannot continue to rely on precarious workforces. Thus, AI developers, platform companies, and regulators should *ensure fair-work principles are respected* throughout the data-production process. The Fairwork Project, a research-oriented initiative from the University of Oxford inspired by the Fairtrade movement, has evaluated different labor platforms across the globe according to the five principles of fair pay, conditions, contracts, management, and representation.[18] To date, none of the platforms evaluated has achieved a perfect score, meaning none of them implements the minimum standards for working conditions. Building upon Clark and Hadfield's concept of regulatory markets for AI, where independent expert institutions inform the public of compliance with regulations,[19] I argue a thorough evaluation of data-production platforms, either through government action or independent research, could elucidate the working conditions across the sector and inform different stakeholders, including AI companies, and thereby potentially induce a race to the top and thus compliance with labor rights and laws.

2. Data quality is also a question of governance. The case of outsourced data production links discussions on data, platforms, and AI governance. For

example, digital platforms have created internal governance mechanisms such as Meta's Oversight Board for content moderation policies. External governance has also come in the form of regulations such as the General Protection Data Regulation. Moreover, co-governance mechanisms, such as the Global Network Initiative and the Partnership on AI, are examples of third-party entities that can steer policies governing data-based goods and institutions.[20] I recommend governance mechanisms that *enable worker input in the data-production pipeline*. Miceli and Posada's research has shown that worker input and feedback on tasks in BPO settings are crucial in improving data quality.[21] Data workers have expertise and unique perspectives because they handle data directly. Their insights could prove crucial to identifying errors in generation, classification, and verification processes.

3. Ethical AI cannot exist without ethical data-production processes that guarantee worker well-being. However, endeavors that guarantee working standards are difficult to conceive and operationalize due to a lack of labor standards in the data-production sector and the race to the bottom that characterizes digital labor outsourcing because clients expect access to data at lower costs. Several impact source initiatives, such as CloudFactory, iMerit, and Sama, have emerged in the data-production sector in recent years. However, as Kaye stresses, the lack of governance mechanisms and involvement from civil society in these issues renders the proliferation and accountability of such ethical endeavors difficult.[22] Even supposedly ethical organizations have been criticized for their labor practices. For example, a recent *Time* article documented possible union-busting practices from Sama.[23] Standards and mechanisms of accountability should be created while *supporting worker-centered initiatives,* including impact-sourcing companies, cooperatives, and not-for-profits, that respect the standards mentioned above to mitigate the race-to-the-bottom trend in platform labor.

In this paper, I have described the importance of labor in the production of data and the subsequent development of data-based technologies such as AI. The current system is one of self-governance and increasing platform authority, where profits are prioritized over high-quality data, that is, data that produce insightful outcomes without reproducing societal harms. There cannot be high-quality data and ethical AI systems without respect for human rights—including labor rights. Therefore, the industry should strive to respect fair-work principles, enable worker feedback in the data-production pipeline, and support worker-centered initiatives backed by standards and effective governance. These initiatives allow for the broader considerations necessary to democratize digital spaces and entities, including platforms, and reduce power concentration among a few entities.

Endnotes

* Acknowledgement and disclaimer: The views and positions expressed in this report are solely those of the author and do not necessarily reflect the views of the Department of Foreign Affairs, Trade and Development (commonly known as Global Affairs Canada) or the Government of Canada. The report is in its original language.

1 Chris Westfall, "Educators Battle Plagiarism As 89% Of Students Admit To Using OpenAI's ChatGPT For Homework," Forbes, 2023, https://www.forbes.com/sites/chriswestfall/2023/01/28/educators-battle-plagiarism-as-89-of-students-admit-to-using-open-ais-chatgpt-for-homework/.

2 Luke Taylor, "Colombian Judge Says He Used ChatGPT in Ruling," The Guardian, February 3, 2023, https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling.

3 Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? □," in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (ACM, 2021), 610–23, https://doi.org/10.1145/3442188.3445922.

4 Open AI, "GPT-4 Technical Report," 2023, https://cdn.openai.com/papers/gpt-4.pdf

5 Mary L. Gray and Siddharth Suri, Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass (Houghton Mifflin Harcourt, 2019).

6 Antonio A. Casilli, En Attendant Les Robots: Enquête Sur Le Travail Du Clic, La Couleur Des Idées (Éditions du Seuil, 2019).

7 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 (January 2020): 205395172091977, https://doi.org/10.1177/2053951720919776; Milagros Miceli and Julian Posada, "The Data-Production Dispositif," Proceedings of the ACM on Human-Computer Interaction 6, no. CSCW2 (2022).

8 Antonio A. Casilli and Julian Posada, "The Platformisation of Labor and Society," in Society and the Internet: How Networks of Information and Communication Are Changing Our Lives, ed. Mark Graham and William H. Dutton, Vol. 2 (Oxford University Press, 2019).

9 Billy Perrigo, "Inside Facebook's African Sweatshop," Time, February 14, 2022, https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/.

10 Sana Ahmad and Martin Krzywdzinski, "Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains," in Digital Work in the Planetary Market, ed. Mark Graham and Fabian Ferrari (The MIT Press, 2022), https://doi.org/10.7551/mitpress/13835.001.0001.

11 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).

12 Sarah T. Roberts, Behind the Screen: Content Moderation in the Shadows of Social Media ; with a New Preface (Yale University Press, 2021).

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

14 Miceli and Posada, "The Data-Production Dispositif."

15 Billy Perrigo, "Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer," Time, January 18, 2023, https://time.com/6247678/openai-chatgpt-kenya-workers/.

16 Eileen Guo, "A Roomba Recorded a Woman on the Toilet. How Did Screenshots End up on Facebook?," MIT Technology Review, accessed February 22, 2023, https://www.technologyreview.com/2022/12/19/1065306/roomba-irobot-robot-vacuums-artificial-intelligence-training-data-privacy/.

17 Julian Posada, "Family Units. The Communities behind the Data Annotation Work That Powers AI," Logic Magazine, 2021, https://logicmag.io/beacons/family-units/.

18 Mark Graham et al., "The Fairwork Foundation: Strategies for Improving Platform Work in a Global Context," Geoforum 112 (2020): 100–103, https://doi.org/10.1016/j.geoforum.2020.01.023.

19 Jack Clark and Gillian K. Hadfield, "Regulatory Markets for AI Safety" (arXiv, 2019), http://arxiv.org/abs/2001.00078.

20 Robert Gorwa, "What Is Platform Governance?," Information, Communication & Society 22, no. 6 (2019): 854–71, https://doi.org/10.1080/1369118X.2019.1573914.

21 Miceli and Posada, "The Data-Production Dispositif."

22 D. Kaye, "A Human Rights Approach to Platform Content Regulation," Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, 2018, https://freedex.org/a-human-rights-approach-to-platform-content-regulation/.

23 Perrigo, "Inside Facebook's African Sweatshop."