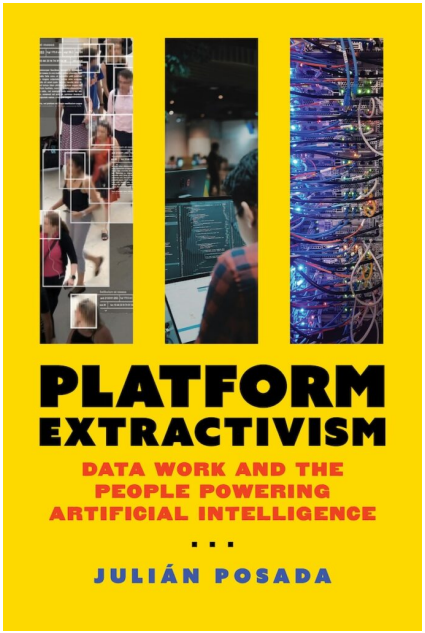
JUST TECH

Data Work and Platform Extractivism

By: Julian Posada

LABOR & ECONOMY



This book grew out of years of research on data work and platform labor that began during my doctoral studies at the University of Toronto and continued after I joined the faculty at Yale University. What started as a study of data work across Latin America during the Covid-19 pandemic came to center on Venezuela, where most of the workers I encountered were based. Combining interviews with 38 data workers with analyses of web traffic and annotation instructions, I examined how digital platforms organize the labor that sustains artificial intelligence. As someone from the region, I also approached this research as an intellectual project: bringing Latin American critical thought into dialogue with workers' experiences and scholarship on digital labor and critical data studies. Thinking through extractivism allowed me to connect contemporary data production to longer histories of labor exploitation, resource appropriation, and colonial power. The book asks what is new about the platformization of labor and data production for AI, and how these arrangements remain embedded in enduring inequalities and struggles. What follows are the opening pages of the first chapter of Platform Extractivism: Data Work and the

People Powering Artificial Intelligence *(University of California Press, October 2026).*

Reprinted from Platform Extractivism: Data Work and the People Powering Artificial Intelligence *by Julián Posada, courtesy of the University of California Press.* *Copyright 2026.*

Before becoming data workers for the artificial intelligence (AI) industry, María, Rogelio, and their children left Venezuela in search of better opportunities, seeking a new life in neighboring Colombia, my home country. Their move reflected a common response to one of the world's most severe economic crises, which had produced rampant inflation and scarce job prospects. The family settled in Bogotá, where María, trained as an administrative technician, found work in a nail salon. Rogelio, a computer science teacher, became a street vendor, while their 14-year-old daughter and 9-year-old son enrolled in public school. For a time, their new life appeared stable. However, when the Covid-19 pandemic struck, both parents lost their livelihoods. Facing the possibility of losing their home in Venezuela to government expropriation and having no income in Colombia, they made the difficult decision to return to Venezuela. After crossing the border and enduring several harrowing days in a government-run reception center, they returned home to find their former lives erased.

Through a friend, the family learned about two digital platforms, Hive and Remotasks, on which many Venezuelans were earning money online. They could work from home, logging into their computers to annotate images, tag objects, and enter data, with payments made in US dollars. Setting up accounts required a significant effort, so they initially looked elsewhere for income, but their then 13-year-old son took the initiative. He opened an account on Hive and began labeling images of cars, signs, buildings, and roads using an old Canaima computer distributed to schoolchildren by the Hugo Chávez government a decade earlier. Eventually he earned 18 cents. Recognizing the opportunity, Rogelio joined in and earned just over US$2 on his first day. He transferred the money to PayPal, hoping to confirm its authenticity. This was no scam. Soon, the entire family began producing data from their computers, with the children particularly involved due to pandemic-era school closures.

To improve their efficiency, the family turned to YouTube tutorials, each member honing their skills in different tasks. The children specialized in annotating road images, María identified supermarket products, and Rogelio focused on labeling warehouse items. One day they encountered a task called "Beast Mode," which involved labeling objects in apartment settings using ground-level camera footage. It offered higher pay than any work they had previously done. In December 2020 they earned US$20 in a single day—their highest earnings yet—just before the Christmas holiday. However, by early 2021 their income began to dwindle. Although tasks remained, the pay had sharply decreased. Beast Mode on Hive disappeared, and Remotasks ended its hourly wage program. When I interviewed the family in summer 2021, they expressed gratitude for the income and opportunities the platforms had once provided but acknowledged that the situation had become unsustainable. With their work on Hive no longer even meeting the minimum cash-out threshold of $5 per week, they were forced to seek other means of support.

As María and Rogelio's family struggled to sustain themselves through data work, the AI industry was entering a new phase. My initial interviews with data workers from Latin America,[1] particularly those from Venezuela (who represented the majority of those I encountered), coincided with OpenAI's release of its first version of DALL-E. By the next year, competitors such as Midjourney and Stability AI had emerged, and OpenAI introduced ChatGPT, a conversational system built on generative pretrained transformer models, sparking a boom in applications built on foundation models. The spotlight shifted from earlier trends, such as self-driving vehicles and blockchain, to generative AI. Over the next several years, alternatives and competitors appeared, including Meta's Llama, Anthropic's Claude, and Microsoft's Copilot, alongside models from non-US companies, such as China's DeepSeek and France's Mistral AI. These developments were made possible in part by the often-overlooked labor of workers such as María, Rogelio, and their children, whose contributions formed part of the data infrastructures that sustain contemporary AI.

María and her family may have done much of their data work on a similar Canaima computer.

Source: Cristóbal Alvarado Minic/Flickr.

This family's story is not exceptional; rather, it represents the largely untold experiences of countless workers whose data production efforts form part of the backbone of contemporary AI.[2] Their labor powers innovations across industries, from self-driving vehicles to content moderation systems to foundational AI models like GPT. These workers are embedded in extensive supply chains that weave together internet users, data brokers, vendors, and AI developers across the globe. Rather than representing a revolutionary break, AI continues a historical trajectory of industrialization and, more recently, platformization, entangled in the legacies of capitalism and colonialism. The industry depends on complex supply chains, including data production, that rely on countless workers across each stage of the process.

Focusing on data production for machine learning during the Covid-19 pandemic,[3] this book argues that the AI industry is enmeshed in a form of extractivism: a transnational system of accumulation, labor exploitation, and power concentration that commodifies knowledge and labor through digital platforms. This process depends on the centralization of power through platforms that mediate between workers and AI developers. As critical infrastructures, these platforms structure work arrangements for data production while perpetuating a system of exploitation that accumulates wealth for a minority. For workers like María, her family, and many others, data work platforms became one of the few available means of survival at a time of crisis. These arrangements, dependent on digital infrastructures, profoundly affect communities and territories in sites of extraction. This book explores these dynamics at the intersection of contemporary data work, power, and profit, interrogating their consequences for the individuals and communities that depend on the AI industry for their livelihoods.

As a framework, extractivism highlights two unfulfilled promises surrounding AI's development, particularly regarding the labor of data production, one of its essential components. The first, made to outsourced workers, is freedom *of* work: the gig economy promise that individuals can work on their own terms, free from traditional employment or government control, at any time and place. The second promise is freedom *from* work: the vision that autonomous agents will eventually take on so-called lower-skilled jobs and eventually all jobs, liberating society from labor. Instead of granting these freedoms, AI data production relies on a global extractive model that commodifies both labor and the data it processes within platform-based systems. This form of labor may also offer a glimpse into the future of work: contingent, outsourced, and precarious.

It is hardly surprising that a significant share of outsourced data work, including the labor examined in this book, takes place in regions with colonial histories. "Extractivism" long predates digital technologies. In Latin America, where the critical vocabulary of extractivism gained particular force, scholars and activists have used the term to criticize development models organized around the large-scale export of minimally processed raw materials. International institutions, including the World Bank,[4] have long framed oil, gas, and mining through the language of "extractive industries." In recent decades, Latin American scholars and activists have used it to criticize governments' overreliance on rapid, unsustainable expansion through resource extraction at the expense of lives and the environment.

Beyond mere extraction, extractivism denotes a transnational system of labor exploitation and resource appropriation that enables the accumulation of commodities across a vast network of social actors, institutions, and territories. This system involves the extraction, preliminary processing, and refinement of commodities through the exploitation of local labor. In these contexts, the lives of workers, families, and communities are radically shaped. In the current climate emergency, this system, which has generated wealth and investments and improved living conditions in some contexts, has also depleted life and become an unsustainable pillar of global capitalism. Extractivism as a system transcends natural resources, embedding itself in institutions and social actors, encompassing life itself, with depletion as a defining feature.[5]

With the consolidation of digital platforms as global infrastructures, AI companies extend their reach into new regions to access labor treated as cheap, replicating earlier extractivist patterns. However, what is extracted through these platforms is not raw material in the conventional sense, but a fictitious, constructed data commodity. Contrary to metaphors like "data mining," data never exists in a "raw" state; it is always abstracted and curated through a production process that companies partially outsource via platforms. The labor that produces this data is itself part of the AI preprocessing chain, as the resulting data is later used to train and refine models at a subsequent stage. For AI companies, the resulting data becomes a valuable resource to collect, appropriate, and extract.

Digital platforms extract and (pre-)process data for the AI industry, but in doing so, they also commodify the exploited labor and knowledge of the many participants involved. This knowledge is both encoded in the data and applied by workers throughout the production cycle. As these platforms consolidate into global communication infrastructures, they entrench themselves in the extractivist patterns long characterized by international markets.[6] I term this system "platform extractivism." Rooted in capital accumulation and power centralization, this concept emphasizes the infrastructural role of platforms while linking them to older extractivist legacies within capitalism and coloniality.[7]

This book examines how platforms, as digital labor infrastructures, enable AI companies to expand their data production into regions with sufficient online work infrastructure. This reach allows platform companies to target populations in nations with adequate internet access and labor markets that support online work. For example, workers like María and her family worked full-time for Scale's platform Remotasks because the infrastructure in their country made participation possible. At the same time, hyperinflation in Venezuela, high unemployment, and pandemic-era restrictions positioned digital platforms among the few available sources of income.

Yet that narrative does not present the full picture. This book argues that digital labor involves not only individual workers but also their territories. Procuring labor that appears cheap in countries like Venezuela is possible because of the support provided by workers' families and communities, who collectively bear the socioeconomic risks of freelance work. Ultimately, investigations of platform extractivism move beyond the issue of platform work to interrogate how digital platforms affect not only individuals but also the communities in which they are embedded.

This book argues that digital labor involves not only individual workers but also their territories. By "territory," I refer not only to geographic space but also to the infrastructures, households, communities, and social relations that sustain workers' participation in platform labor. Procuring labor that appears cheap in countries like Venezuela is possible because of the support provided by workers' families and communities, who collectively bear the socioeconomic risks of freelance work.

Furthermore, this book argues that algorithms and interfaces systematically erase these workers' contributions through design choices that silence dissent and ensure compliance with the demands of platform clients in the "Minority World." Crucially, extractivist processes function in multiple directions. The enduring legacies of capitalism and colonialism are intertwined in a double helix of extraction and imposition. When Spanish conquistadors arrived on the shores of what they deemed a "New World," they implemented a system of *encomiendas* and *repartimientos* to exploit local populations, accompanied by evangelization efforts and cultural genocide.[8] These entangled processes continued well into the era of modern and independent national governments, shaping lives into the present.[9] The double helix symbolizes the intersection of material and immaterial forces, economic production and relations, on the one hand, and the epistemic realm on the other, where capital accumulation and epistemic dominance are inseparable.

In today's data-driven world, this takes shape as a platform system that not only extracts data but also instructs workers how to interpret the world and encode client-defined values into datasets and algorithms. As a result, algorithms may excel at recognizing objects in a California home but fail in a rural Latin American one, and they may operate fluently in English while neglecting languages spoken by smaller populations.[10] These challenges stem from the rapidly expanding data production system that fuels AI and shapes lives across the globe. The exploitation of workers is rationalized by shifting ideologies in the tech industry, such as technoliberalism, an ideology that rose to prominence in the late twentieth century and equated technological progress with freedom in a postracial, postlabor world. By the mid-2020s, however, a deeper and more overt turn among some technology leaders toward more open support of authoritarianism and right-wing populism offered companies moral cover to pursue technological development regardless of human cost, for what was framed as the future benefit of humanity.

This book emphasizes the link between extraction and the erasure of the obscured human contributors whose labor is vital for producing data and enabling the development of AI products. The material aspects of platforms, including interfaces and technological infrastructures, amplify the epistemic dominance of clients who dictate data validity and algorithmic meaning. At the same time, prevailing ideologies within the tech sector justify the exploitation of workers and their communities by powerful entities. Together, these dynamics suggest that the profitable creation of datasets for AI depends on cheap labor and control mechanisms, ideologically backed processes that embed dominant global perspectives within the data. Hence, the material dimensions of labor production are inseparable from the immaterial characteristics of data production. Extractivism requires erasure and epistemic dominance; in turn, these elements perpetuate extractivism.

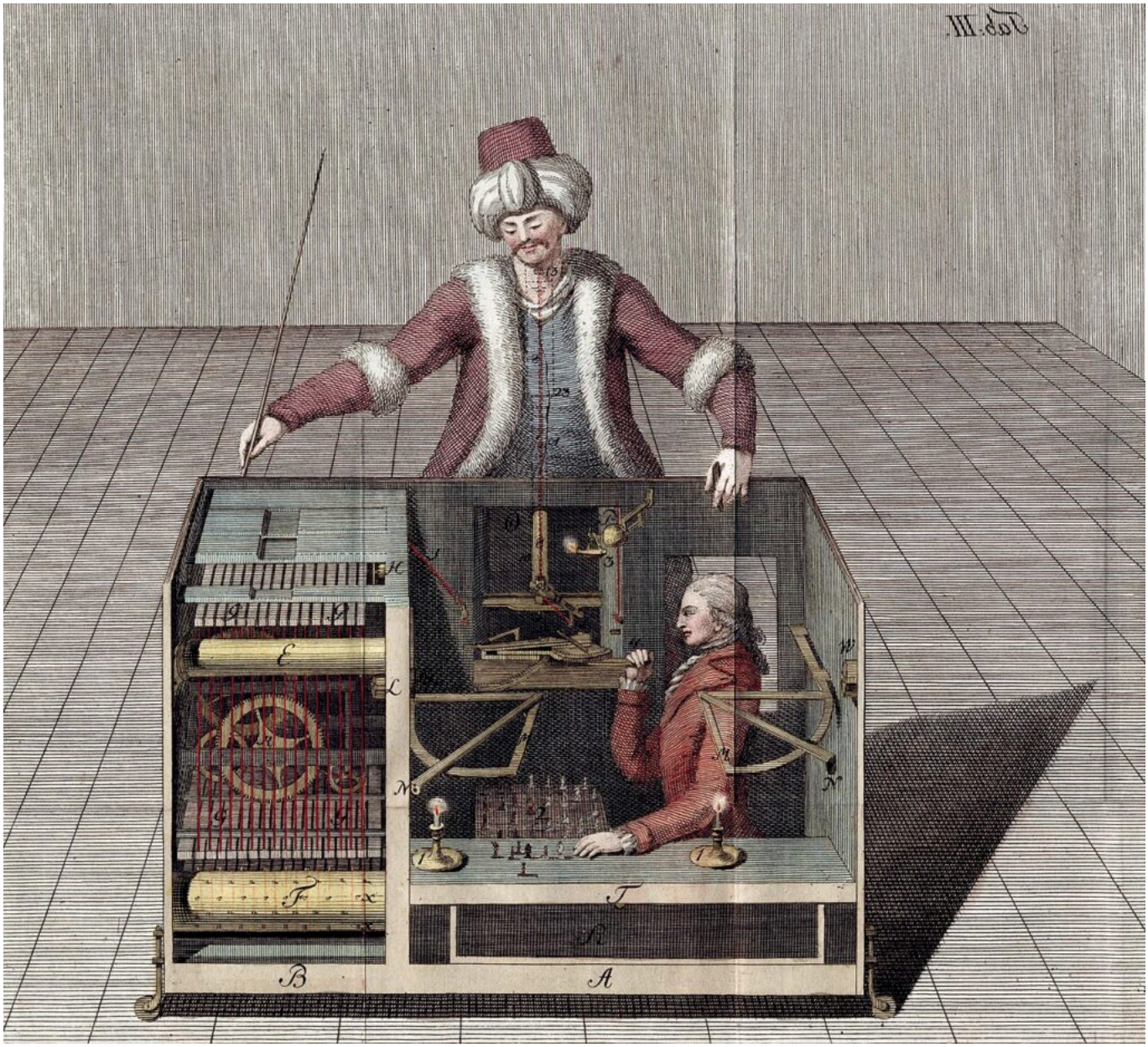


Figure 1. Engraving by Joseph Friedrich Freiherr von Racknitz, representing one of the first credible attempts to expose the Mechanical Turk as a hoax by depicting how a human operator could be concealed within the cabinet to manipulate the chess pieces. Source: Joseph Friedrich Racknitz, Ueber den Schachspieler des Herrn von Kempelen und Dessen Nachbildung (J. G. I. Breitkopf, 1789), 65.

The stories of the workers examined in this book are closely tied to the platforms that made their labor possible. A key predecessor, Amazon Mechanical Turk (MTurk), launched publicly in 2005 and was described at the time by the company's CEO as a form of "artificial artificial intelligence," a simulated form of AI.[11] The platform's name refers to an eighteenth-century automaton dressed in orientalist costume that toured Europe, creating the illusion of a machine capable of defeating challengers at chess. In reality, a hidden human operator controlled its moves; an apt metaphor for the concealed labor behind the façade of automation (see figure 1).[12]

In Venezuela, three platforms emerged as primary tools for workers. Appen, an Australian company founded in 1996, was less commonly used by the workers I interviewed but has maintained the longest

presence in the country through its acquisition of FigureEight (formerly CrowdFlower), active in the region since the mid-2010s. The other two platforms, Hive and Remotasks, employed many Venezuelans, including María and Rogelio's family. Hive, formerly Castle Global, is a San Francisco–based company founded in 2012, while Remotasks is operated by Scale AI, also based in San Francisco and established in 2016.[13] These platforms count among their clients major tech firms such as Alphabet, Meta, and Microsoft, as well as companies in other industries, including logistics and automotive manufacturing.

Footnotes

1 I use the term "Latin America," which originated in nineteenth-century French imperialism, to designate the region, united by shared histories and cultural and linguistic affinities as well as common struggles and forms of oppression. I use this designation because it is the most common term used today. At the same time, I draw on Arturo Escobar's concept of Latin America as a pluriverse, acknowledging that no single designation fully captures its complexity. I acknowledge that its colonial origins have led many activists in the region to reject the term in favor of other names, with "Abya Yala," a term from the Cuna language meaning "mature land," being widely preferred, particularly by many Indigenous communities. Abya Yala is commonly used by many Indigenous activists and scholars to refer to the entire landmass of the Americas, which is considered one continent in Spanish ("América") and two continents (North and South America) in English.

2 Throughout this book I refer to the processing of data through human labor, or data work, as "data production." This process entails the generation and annotation of data as well as the verification of algorithmic outputs generated through machine learning. This term is synonymous with others used in the AI industry, such as "data enrichment."

3 I started collecting data for this study in mid-2020 and conducted the last interviews in June 2025. However, since most of the data was collected between 2020 and 2022, overlapping broadly with the World Health Organization's classification of the Covid-19 pandemic as a "public health emergency of international concern" (January 20, 2020, to May 5, 2023), I define the pandemic as the timeframe delimitation for the study.

4 Andrés Liebenthal et al., *Extractive Industries and Sustainable Development: An Evaluation of World Bank Group Experience* (World Bank, 2005), https://documents1.worldbank.org/curated/en/579221468328166526/pdf/32671.pdf.

5 Here, I understand depletion in both quantitative and qualitative terms. Quantitatively, depletion implies exhaustion; it refers to the deterioration of the social, ecological, and territorial conditions that sustain life. Beyond the social burden placed on extractive sites, it is therefore crucial to consider the environmental damage extractivist industries inflict on life in local territories. In extreme cases, depletion may also involve the outright extinction of life.

6 Following Pasquinelli's theorization, I view knowledge as imbricated in the material acts and artifacts in labor. However, because platforms trade knowledge and labor separately—knowledge as data and labor through worker accounts—I refer to them distinctly. See Matteo Pasquinelli, *The Eye of the Master: A Social History of Artificial Intelligence* (Verso, 2023).

7 My definition of “platform extractivism” is closely related to, yet distinct from, the concept of “data extractivism.” Authors such as Gago and Mezzadra, Chagnon and Hagolani-Albov, Couldry and Mejias, and Paola Ricaurte have used the latter term to describe the appropriation of data from daily life by technology corporations. Platform extractivism overlaps with this understanding in that digital platforms are also employed by corporations for these purposes. However, my focus is not on the data itself but on the platform as an infrastructure for extractivism. In addition, I consider labor, alongside data, as a fictitious commodity central to this system. I regard platform extractivism as distinct from, but complementary to, data extractivism: Each addresses different aspects of the broader phenomenon of extractivism and its entanglement with capitalism and coloniality. It is also important to recognize that the concept of platform extractivism builds on the interdisciplinary scholarship of critical data studies, including work on race, gender, and technology; coloniality; the climate emergency; and epistemology, referenced here and throughout the book, as well as the extensive literature on platform labor, also referenced below and throughout. See, for example, Payal Arora, “Creative Data Justice: A Decolonial and Indigenous Framework to Assess Creativity and Artificial Intelligence,” *Information, Communication & Society* 28, no. 13 (2025): 2231–47, https://doi.org/10.1080/1369118X.2024.2420041; Chimayi Arun, “The Silicon Valley Effect,” *Stanford Journal of International Law* 61, no. 1 (2025): 55–110, https://law.stanford.edu/publications/the-silicon-valley-effect/; Renata Ávila Pinto, *Against Digital Colonialism* (Autonomy Institute, 2020), https://autonomy.work/wp-content/uploads/2020/09/Avila.pdf; Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity, 2019); Patrick Brodie, “Climate Extraction and Supply Chains of Data,” *Media, Culture & Society* 42, nos. 7–8 (2020): 1095–114, https://doi.org/10.1177/0163443720904601; Iván Chaar López, *The Cybernetic Border: Drones, Technology, and Intrusion* (Duke University Press, 2024); Anita Say Chan, *Networking Peripheries: Technological Futures and the Myth of Digital Universalism* (The MIT Press, 2014); Beth Coleman, “Race as Technology,” *Camera Obscura: Feminism, Culture, and Media Studies* 24, no. 1 (2009): 177–207, https://doi.org/10.1215/02705346-2008-018; Catherine D’Ignazio and Lauren F. Klein, *Data Feminism* (The MIT Press, 2023); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (Picador St. Martin’s Press, 2019); Ignacio Siles et al., “Toward a Popular Theory of Algorithms,” *Popular Communication* 21 no. 1 (2023): 57–70, https://doi.org/10.1080/15405702.2022.2103140; Milagros Miceli et al., “Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM, March 3, 2021), 161–72, https://doi.org/10.1145/3442188.3445880; Andrew Iliadis and Federica Russo, “Critical Data Studies: An Introduction,” *Big Data & Society* 3, no. 2 (2016): 2053951716674238, https://doi.org/10.1177/2053951716674238; Rob Kitchin, *Critical Data Studies: An A to Z Guide to Concepts and Methods* (Polity Press, 2025); Michael Kwet, “Digital Colonialism: US Empire and the New Imperialism in the Global South,” *Race & Class* 60, no. 4 (2019): 3–26, https://doi.org/10.1177/0306396818823172; Sebastian Lehuedé and Ana Valdivia, “Peripheries on the Rise: Eco-Imperialism in the Race for Technology Resourcing,” SSRN, January 7, 2025, https://doi.org/10.2139/ssrn.5085112; Mirca Madianou, *Technocolonialism: When Technology for Good Is Harmful* (Polity, 2024); Nick Couldry and Ulises Ali Mejias, *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism, Culture and Economic Life* (Stanford University Press, 2019); Ulises A. Mejias and Nick Couldry, *Data Grab: The New Colonialism of Big Tech and How to Fight Back* (University of Chicago Press, 2024), https://doi.org/10.7208/chicago/9780226832319.001.0001; Stefania Milan and Emiliano Treré, “Big Data from the South(s): Beyond Data Universalism,” *Television & New Media* 20, no. 4 (2019): 319–35, https://doi.org/10.1177/1527476419837739; Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York University Press, 2018); Pasquinelli, *Eye of the Master*; Bruno Moreschi et al., “The Brazilian Workers in Amazon Mechanical Turk: Dreams and Realities of Ghost Workers,” *Contracampo* 39, no. 1 (2020): 44–64, https://doi.org/10.22409/contracampo.v39i1.38252; Juan Ramos-Martín, “Capitalismo informacional desde el sur. Fetichismo de la información y materialidades de la colonialidad de los datos,” *Revista latinoamericana estudios de la paz y el conflict* 3, no. 6 (2022): 58–68, https://doi.org/10.5377/rlpc.v3i6.14051; Paola Ricaurte, “Data Epistemologies, the Coloniality of Power, and Resistance,” *Television & New Media* 20, no. 4 (2019): 350–65, https://doi.org/10.1177/1527476419831640; Fernanda R. Rosa, “From Community Networks to Shared Networks: The Paths of Latin-Centric Indigenous Networks to a Pluriversal Internet,” *Information, Communication & Society* 26, no. 11 (2023): 2326–44, https://doi.org/10.1080/1369118X.2022.2085614; Nick Seaver, “Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems,” *Big Data & Society* 4, no. 2 (2017): 205395171773810, https://doi.org/10.1177/2053951717738104; Sophie Toupin and Roda Siad, “Artificial Intelligence for Development (AI4D): A Contested Notion,” *Big Data & Society* 12, no. 2 (2025): 20539517241304678, https://doi.org/10.1177/20539517241304678; Tiziano Bonini and Emiliano Treré, *Algorithms of Resistance: The Everyday Fight Against Platform Power* (The MIT Press, 2024); Christopher W. Chagnon and Sophia E. Hagolani-Albov, “Data Extractivism,” in *The European Digital Economy*, ed. Judyta Lubacha et al. (Routledge, 2023), https://doi.org/10.4324/9781003450160-14; Verónica Gago and Sandro Mezzadra, “A Critique of the Extractive Operations of Capital: Toward an Expanded Concept of Extractivism,” *Rethinking Marxism* 29, no. 4 (2017): 574–91, https://doi.org/10.1080/08935696.2017.1417087.

8 The encomienda and repartimiento systems were labor practices imposed by Spanish colonizers in the Americas. Under the encomienda system, Spanish settlers were granted control over Indigenous communities, forcing them to work in exchange for supposed protection and Christian instruction. The repartimiento system replaced encomienda in the mid-sixteenth century, requiring Indigenous peoples to continue providing labor for a set number of days with nominal pay. Although less brutal than encomienda, repartimiento nonetheless perpetuated exploitation. Both systems contributed significantly to the oppression and decline of Indigenous populations in the Americas.

9 This account is written with the historical context of Canada's residential school system and the American Indian Boarding Schools in the United States in mind. These institutions forcibly removed Indigenous children from their families, stripping them of cultural identity under the guise of education. In Canada, Gordon Residential School in Saskatchewan closed in 1996. In the United States, the Indian Child Welfare Act of 1978 responded to the continuing removal of Native children from their families through foster care and adoption systems, following earlier assimilationist policies that included boarding schools.

10 These examples echo the findings of Terrance DeVries and colleagues, who evaluated early computer vision models for object recognition. Their research showed that these models were more effective at identifying objects common in wealthier households in higher-income countries (e.g., the United States) than those found in lower-income contexts. Since then, newer datasets and models have been developed to address such disparities. See Terrance DeVries et al., "Does Object Recognition Work for Everyone?" (paper presented at IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019).

11 Lilly Irani, "The Cultural Work of Microwork," *New Media & Society* 17, no. 5 (2015): 720–39, https://doi.org/10.1177/1461444813511926.

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

13 In the early stages of my research, I anonymized the companies Appen, Hive, and Remotasks, referring to them as Clickrating, Workerhub, and Tasksource. This was intended to ensure participant anonymity and support ongoing research. However, as this study reaches its conclusion and this book represents a final presentation of results, I determined that naming the companies no longer created meaningful additional risk for participants, provided that individual workers remained anonymized.