## TECHGLOBAL INSTITUTE

# Labor in the Shadows

*Rights and Risks for Asia's Data Workers*

Disha Verma

*August 2025*

# Table of Contents

# Introduction

Digital technologies—particularly social media or user-generated content platforms and artificial intelligence services—are powered by vast global supply chains of labor, often hidden from public view. At the heart of these systems are thousands of workers who perform content moderation, data annotation, and other forms of digital labor that are essential for training and maintaining technological infrastructures. This labor is predominantly extracted from the Global Majority, facilitated by favorable policy environments and an abundance of low-cost human resources. Other parts of the tech supply chain, such as mineral extraction for hardware components and electronics manufacturing, are similarly marked by exploitation.

By and large, these labor-intensive operations are controlled by multinational tech corporations, mainly located in the Global North, and are routinely outsourced to lower-cost jurisdictions—often not directly to workers, but to third-party vendors, who in turn hire and manage workers.

Offshoring has become a popular model, as experts observe, because companies can locate screen-based activities anywhere they find the best bargain of skills and productivity. India and the Philippines, since in the late 1980s and 1990s respectively, emerged as global hotspots for tech corporations to outsource back-office functions—such as customer support or IT-enabled services—owing to their vast supply of cheap specialized labor. Now, these functions include content-related digital platform labor, including but not limited to content moderation and data annotation. Business process outsourcing companies presently employ 1.1 million people in India and 1.2 million in the Philippines. Similarly, other countries in Asia, like Pakistan, Bangladesh, Lebanon, Malaysia, and Indonesia have sizable workforces engaged in various forms of digital platform labor, including gig work.

In these markets, outsourcing models relegate workers to the bottom of the supply chain—they often have little to no access to, information about, or bargaining powers with the platforms

or projects they work for. Meanwhile, local vendors compete to secure contracts from tech corporations in a race to the bottom, each striving to provide the cheapest possible service, thereby perpetuating exploitative labor practices themselves and playing a role in keeping the labor wages low. This has created transnational value chains embedded in a planetary market, where corporations are seen to be taking advantage of borderless contractual arrangements that often escape domestic legal scrutiny, sidestep labor rights, and neglect worker well-being. Pressures from regulators or civil society have proven only partly effective as platforms often find covert ways to sustain these models, discussed at length in this essay, signaling not only complicity, but a systemic and organized appetite to exploit labor.

Practices reminiscent of colonial repression have shaped and sustained offshoring businesses such as the business process outsourcing industry, where economies of the Global North are able to appropriate disproportionately high amounts of labor from the Global Majority without an equitable exchange. Creative and sophisticated aspects of tech production and sales—like direction, production, marketing, public relations, or campaigning—are almost exclusively restricted to company headquarters in

**India and the Philippines, since in the late 1980s and 1990s respectively, emerged as global hotspots for tech corporations to outsource back-office functions, which now has evolved to include content-related digital platform labor.**

in the Global North, while the unglamorous work of delivering end-mile logistics, training artificial intelligence models, or cleaning up content on platforms is leased out to exploited markets in the South. Taken together, these systems reveal how the promise of technological innovation is frequently built on labor regimes that externalize social costs onto already marginalized populations, while remaining largely unaccountable to them.

The present state of digital platform labor is marked by a wide spectrum of harms, warranting platform accountability and policy reform. In this

context, our research examines how systemic mechanisms within tech supply chains, particularly content moderation and data annotation, enable sustained labor exploitation in the Global Majority. Drawing on case studies from Asia with supporting evidence from Kenya, Ghana, and Venezuela, it maps the patterns of harm embedded in the outsourcing model. The final section assesses these practices against international and regional labor standards on decent work and worker protection, such as those of the International Labor Organization (ILO), and evaluates domestic measures to combat harm.
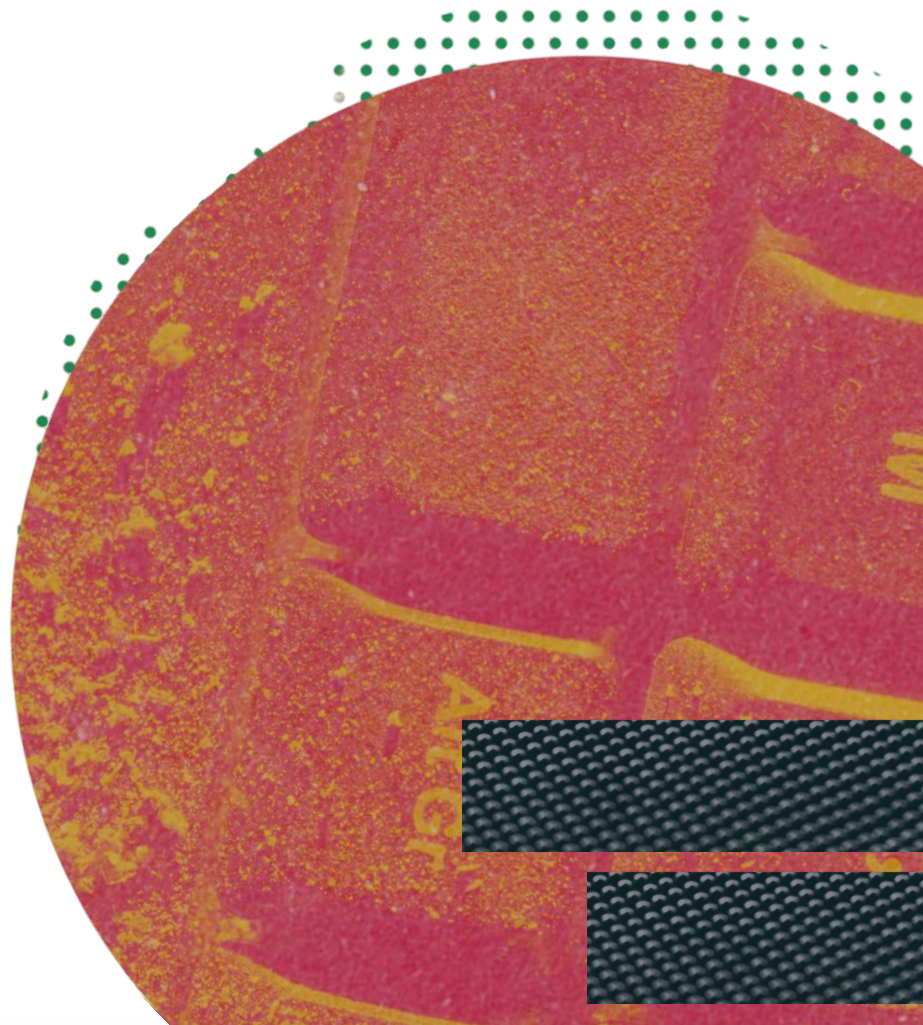
# Background

Content moderation refers to the structured process of screening user-generated content on websites, social media, and other online platforms to assess its adherence to legal mandates or community guidelines for a particular site, community, or jurisdiction. This process can be automated, where tools such as Microsoft's PhotoDNA or YouTube's Content ID can be used to manage content at scale, or it can be manual. Human moderators—variously referred to as digital janitors, the cleaner, hidden soldiers, internet's guardians, or internet's frontline defenders—remain central to this system, performing tasks that are, as experts note, "*resource intensive and relentless.*"

A key challenge with examining this environment is the opacity surrounding moderation practices, often treated as industrial secrets. Technology companies cite several reasons for this secrecy, including protection of workers' identities and safeguarding proprietary technology. As a result, academic scholarship is limited, and public knowledge largely stems from

media exposés, worker testimonies, and litigation focused on the psychological and legal toll of this work.

Data labelling or annotation refers to the structured process of reviewing, tagging, classifying, or categorizing raw data—such as text, images, audio, or video—to make it usable for machine learning and artificial intelligence systems. It involves significant human labor to train algorithms to recognize patterns, make decisions, and perform tasks such as facial recognition, natural language processing, autonomous navigation, and content recommendation. Variously described as microworkers, click-workers, data workers, hidden soldiers, or ghost workers, humans are constantly in the loop to vet the voluminous datasets that feed these models.

Given the volume and the characteristics of the tasks, these workers, comparably to content moderators, often carry out highly repetitive, cognitively taxing, invisible, and often underpaid labor that is essential to generate data for training

and validating artificial intelligence systems. A large amount of this workforce comes from economically marginalized groups in South and Southeast Asia, Europe, West Asia, East Africa, and South America; the data annotation market in India could exceed $7 billion by 2030, with a potential workforce of 1 million workers. But, also like content moderators, the workforce operates without formal contracts, social protections, or avenues for redress.

Trust and Safety (T&S) in the tech industry refers to the work of protecting users and digital communities from harm by setting rules, moderating content, improving user-to-user interactions, and building safer products. It includes preventing abuse like harassment, scams, disinformation, or exploitation. Although there is growing recognition of Trust and Safety (T&S) professionals —through increased visibility of their voices and acknowledgement of their role in shaping safer online experiences —the industry still marginalizes the contributions of those at the very frontlines of this work: content moderators, data annotators, and labelers (collectively referred to as "data workers" in this report).

This imbalance reflects a deeper bias: while the Global North benefits from safer, curated digital spaces, the Global Majority bears the human cost of the labor that enables them without being afforded dignity, transparency, or power in shaping the systems they sustain.

# Asia's Data Worker Landscape

**India and the Philippines are two of the biggest hubs for data workers**

🔴 **Third-party data workers**  🔵 **Company T&S unit**

**JORDAN, LEBANON**
both have third-party data labelers and content moderators

**PAKISTAN**

**UNITED ARAB EMIRATES**

**BANGLADESH**

**INDIA**
only country with both trust & safety teams and multiple third-party data worker units

**CAMBODIA**

**THE PHILIPPINES**

**MALAYSIA**

**SINGAPORE**
has the largest trust & safety teams in the region

**INDONESIA**
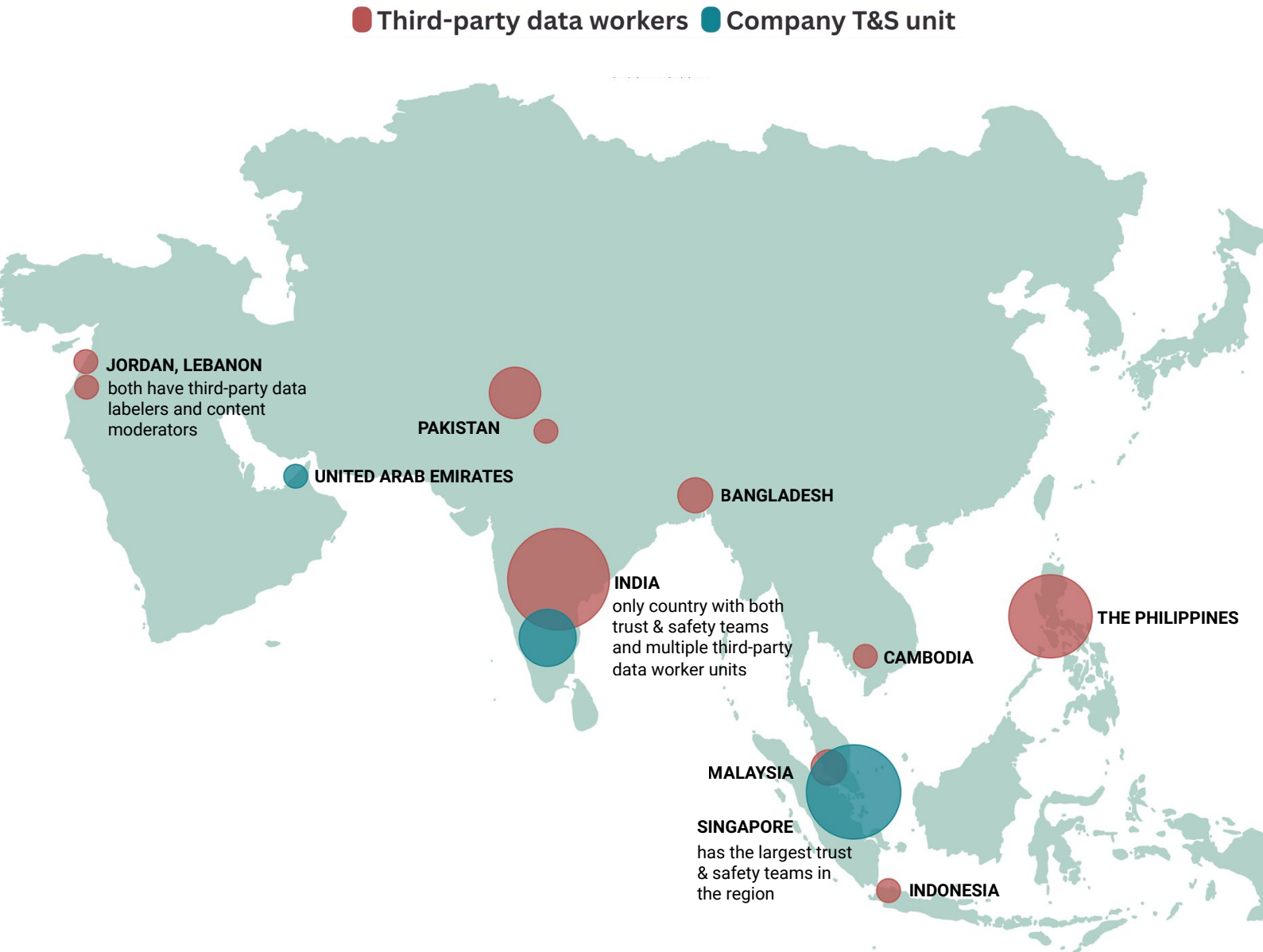
Size of circle represents staffing size in relative terms. "Data workers" mean both content moderators and data annotators for AI/ML models.

*Disclaimer*: This map is limited by available data and verification; it is not exhaustive.

# Content Moderation & Data Annotation Industries in Asia

## Vague Hiring Practices

It is commonplace for content moderators and data annotators to be hired for or assigned the roles without complete transparency as to their responsibilities and occupational risk.

Across jurisdictions with high prevalence of digital platform workers, companies can be seen keeping employment arrangements vague and informal, using terminologies like "partners", "agents", "executives" or others that are not recognized as employees under domestic law, stripping workers of baseline labor protections. *Indonesia*, a hub for logistical services platform labor, for instance, does not recognize gig work as work. Platforms use the term "partnership" to misclassify the employment arrangement, which falls outside the purview of Indonesia's primary labor code, *Law No. 13 of 2003 on Manpower* and is in fact regulated under *Law No. 20 of 2008 on Micro, Small, and Medium Enterprises*, which imposes starkly

## India

With a booming business process outsourcing industry employing over 1.1 million people, India is one of the world's largest hubs for outsourced content moderation. With moderators facing relatively low and stagnant wages, minimal benefits, limited opportunities for promotions and skill development, and poor management practices, the sector has a high attrition rate. These conditions are not incidental or isolated; they are deeply rooted in weak labor protections that fail to meet fundamental standards of occupational well-being and safety throughout the employment lifecycle, while enabling individualized contractual arrangements to dictate employment conditions. Job advertisements sometimes use vague or misleading titles—such as "system analyst," "website administrator," or "process executive"—to describe content moderation roles, thereby attracting unsuspecting candidates.

While this obfuscation may partly stem from seemingly innocuous factors, such as varying skill requirements, the diversity of services offered by firms, or the limited use of "content moderator" as a formal occupational title in India, its impact is far from harmless. It not only obscures the nature of content moderation work but also prevents prospective employees from fully understanding the psychological risks involved.

different rights and obligations. More recent efforts to legislate on evolving labor issues, such as *Law No. 11 of 2020 on Job Creation*, have notably excluded digital platform workers from the conversation. The landscape in the *Philippines*, a significant market for content moderation, is simpler—the existing gamut of labor laws do not extend adequate protections to digital platform labor. In pleasant contrast, *Singapore's* new Platform Workers' law (Bill No. 26/2024) defines digital platform workers as a distinct legal category between employees and the self-employed.

## Opacity in Operations and Cascading Work Structures

Content moderators face complex and highly intermediated reporting structures. Generally, platforms do not employ moderators directly. Instead, they enter into agreements with multinational business process outsourcing firms like Sama or Teleperformance, and these contracts are then passed down to their local affiliates and subsidiaries. Opacity around contractual arrangements means content moderation operations could be further fragmented, passed from one intermediary to another, or outsourced to smaller entities— obfuscating lines of ownership, control, and accountability. At the base of this

chain are the final-tier outsourcing firms, which recruit moderators through fixed-term or rolling contracts that often exploit workers' economic vulnerability. These contracts share common identifiers: they mandate rotating assignments across multiple platforms and roles to further obscure employer responsibility, impose arbitration and jurisdictional clauses that shift legal recourse to inaccessible forums, and enforce strict non-disclosure agreements that prevent whistleblowing. In effect, a cascading system of outsourcing contracts undermines accountability for tech companies at the top for harms perpetuated further down by their subcontractors.

Respite comes in other forms, but elsewhere. The European Union (EU), through its Directive 2024/2831 on *Working Conditions in the Platform Economy*, provides that in every case where there is an element of direction and control by the platform over the worker, the contractual relationship between them is legally presumed to be an employment relationship, and the onus to prove that it is otherwise falls on the platform. An interim ruling in 2023 in a landmark lawsuit over unfair working conditions for content moderators in *Kenya* held Meta to be their primary employer, with the outsourcing firm Sama deemed merely an agent.

an agent. Claims by Meta that it does not have legal registration in the country and was never their employer, and are therefore neither liable for nor privy to the allegations of harm against moderators, were rejected by the court, setting a strong precedent for how these firms engage thousands of workers both regionally. In Asia, such institutional will is currently absent.

Data annotation work, as is explored in more detail in the sections below, is operationalized either through middle-income students or professionals looking to make extra money in their spare time, or through globally dispersed networks of refugees and low-income communities, intermediated by international agencies, non-profit agencies, freelancing platforms, and private contractors. Unlike outsourced content moderation, where workers may sometimes know which company/platform they serve, data annotation work is often so fragmented that workers are left with little or no insight into the end users of their labor or the technologies they help build. For instance, Syrian annotators in Lebanon cannot communicate directly with clients; all interaction is filtered through their employer, Humans in the Loop. As a result, if a client expresses dissatisfaction after a project is submitted, annotators are often required to redo tasks, without additional compensation. The network is so opaque and secretive that it is nearly impossible to determine, with any certainty, the intended use of the labor, or who ultimately benefits from it. With the growing integration of artificial intelligence in surveillance, predictive analytics, facial recognition, reconnaissance, target identification, autonomous decision-making, drones, and weapons, there is a real possibility that this labor is helping to build technologies that enable the very forms of violence and war crimes many of these workers have fled.

Data annotation platforms also operate on opaque algorithms where tasks are allocated based on an undisclosed meritocratic ranking system, and workers are constantly subjected to strict, inscrutable metrics. Algorithms can take on managerial roles and facilitate penalties and terminations without appropriate appeal mechanisms. Parallels can be drawn to traditional factory work, where capitalist corporations routinely deem crafts workers "disposable" and "replaceable"— only here, this disposability can be conveyed and facilitated by an algorithm rather than a real person. Algorithmic management and bias emerged as a possible avenue for intervention, but the ILO failed to reach consensus on whether it fell in its remit.

The EU Directive 2024/2831 as well as Uruguay's Law no. 20396 of February 2025 contains detailed provisions requiring transparency from platforms in using automated monitoring and decision-making algorithms, and advocating for them to be prohibited in certain use cases, such as employee termination.

role of a worker from *Honduras* was terminated in the midst of a severe mental health crisis. An investigation by Foxglove Legal, along with a joint inquiry by The Bureau of Investigative Journalism and *The Guardian* on these sectors in East Africa illustrate how platforms tend to leave markets with the slightest hint of pushback, leaving their

## Malaysia

Malaysia saw fleeting activity as a content moderation hub through Accenture, TikTok's moderation contractor for South Asia. Over 500 of these 800 moderators—including Pakistani moderators who had migrated to Kuala Lumpur in search of better opportunities—were abruptly laid off in October 2024 as the platform turned to artificial intelligence for its content moderation work. The landscape in Kenya, and now Ghana, is also fraught with such precarity.

### Precarity of Jobs

Employment in both sectors—content moderation and data annotation—is notoriously precarious. Global sociopolitical tides, domestic regulatory pressure, or internal platform policy changes, can all trigger termination from the role. In *India*, it is near industry practice for data annotators using freelance platforms to be terminated at any time and without cause, notice, or explanation—a power platforms exploit regularly. Content moderators, for instance, are routinely fired for not meeting their targets—th

workforce behind, unemployed and with few transferable skills.

Data annotators commonly identify as freelancers or self-employed. Freelance status means they are denied job security and basic protections, creating a precarious environment that takes a significant toll on their mental health—especially when a single mistake can cost them their only livelihood. Job status and conditions are especially precarious in the Philippines, with a market saturated with thousands of sub-contractors of American-based data annotation firms. International

labor instruments, spanning over 200 ILO Conventions, did not extend protections and standards to self-employed workers until deliberations on governing digital platform labor began in 2023. Additionally, there is also uncertainty about upskilling or when labor conditions for these roles may improve. Increasing profits for digital labor platforms may not always translate into better conditions for their workers. For instance, Meta is set to invest heavily in Scale AI, an AI annotation firm housed in the United States with subsidiaries around the world—but experts speculate the workforce, largely located in the Global Majority, are likely to not profit from this deal in any way.

## Regulatory Arbitrage as a Potential Labor Issue

Regulatory arbitrage, in its simplest form, is the tendency of corporations to exit jurisdictions they are facing regulatory pressures in to enter more lenient policy environments. Additionally, companies appear to set up bases in jurisdictions with friendlier sectoral regulation. Tech corporations outsourcing operations from *India* and *the Philippines*—two countries with matured business process outsourcing markets and ready availability of cheap labor, infrastructure, and logistics—rather than setting up regional offices in smaller markets like Bangladesh, Pakistan or Malaysia, points to another precarity: platforms choose jurisdictions primarily based on favorable economic and political climates, and even the smallest change in tide may drive them out.

In Eastern Africa, for instance, the mobility of moderation operations has become a strategic asset for tech companies, enabling them to shift operations across borders to evade accountability. In Kenya, Meta and its content moderation vendor Sama both responded by engaging in regulatory and geographical arbitrage. Sama discontinued its moderation services citing a challenging economic climate and new business needs, while Meta shut down its Nairobi operations and relocated to Accra, Ghana. For tech firms, it is often operationally and commercially more expedient to move to jurisdictions with permissive market conditions, weaker regulatory oversight, and a mobile, precarious workforce. At every level, the playbook appears to be engineered not only to deliver labor at scale, but to also ensure an arm's distance between platforms and the workers who power them—diffusing responsibility, deflecting blame, and deterring systemic accountability.

## Commodification of Socioeconomic Crises

Though crowdsourced data work is interspersed across jurisdictions by virtue of being remote, the Global Majority, once again, carries its weight. However, on a closer read, concentrated data annotation markets often coincide with one of two political climates: labor commonly comes from economically marginalized communities, or from zones of humanitarian crises. India and the Philippines are significant but severely underpaid IT markets, with an oversupply of labor often settling into any job they can find. Data annotation work is somewhat popular among Indian women—small-town, middle class women often turn to data annotation and other forms of clickwork as a means to enter the workforce, contribute economically to the household, or gain validation of the family. Similar to content moderation operations, intermediary companies like Appen, CrowdFlower, Microwork, Sama, and Upwork manage invisible workforces behind data annotation. But the sector runs rife with labor issues: typically bad pay, long hours, and no social protections.

Hundreds of millions more from economically marginalized groups across the Global Majority make up a vast, precarious labor pool increasingly drawn into digital microwork under exploitative conditions. In Venezuela, for instance, the economic crisis pulverized by triple-digit hyperinflation, food shortages, and a collapsing currency has forced thousands, many previously middle-class and well-educated, to turn to data work for the autonomous-vehicle industry. Intermediary firms, some having up to 75% of their workforce from the crisis-struck country, often pay them less than a dollar an hour.

Fatma, a Syrian refugee living in a camp in Bulgaria, recounted working part-time for a data-labelling company, tasked with sifting through and categorizing images —"separating the trees from the bushes, and cars from people, roads, and buildings." Her colleague, Diana, was engaged in labelling images of people based on race, age, and gender. Across refugee camps in Kenya's Dadaab camps, Lebanon's Shatila camp, and Jordan's Zaatari and Azraq camps—some of the largest globally—such annotation work has transformed informal digital labor into a managed industry for automation. Fatma, Diana, and others at refugee camps around the world are a part of what commentators have described as refugee-industrial complex, a system of economic, political, and institutional interests that benefit from and/or sustain the ongoing management,

containment, and control of refugee populations—often under the pretext of humanitarianism or security. Often framed as promoting economic self-sufficiency for refugees, digital microwork has instead commodified displacement and immiseration. Globally, the refugee population is likely exceeding 120 million, representing 1 in every 69 people on earth, making them a vulnerable source of low-cost, low-skill workforce with little bargaining powers or protections for tedious, repetitive, and labor-intensive tasks.

Though the right to work for refugees is established in the 1951 Refugee Convention and the International Covenant on Economic, Social and Cultural Rights, many host countries limit this right in practice. Even if digital platform work is legally protected and well legislated in the host country, discriminatory practices, exclusion from certain sectors, and policies restricting mobility, as noted among data annotators above, will obstruct refugees from seeking protection under domestic laws. In States that do not abide by international instruments, refugees have little to bank on. The Global Compact on Refugees calls for the enhancement of refugee resilience and self-reliance, in part through digital upskilling, but cannot similarly be enforced in practice unless through

domestic political will. Additionally, throwing another wrench in realization of international standards, platforms have actively lobbied against regulation and formalization of platform work in domestic contexts on the grounds that the "reclassification of platform labor as employment would make it more difficult, and in some cases impossible, for refugees to access this type of work."

### Cruel Work Hours and Paltry Pay Synonymous with these Sectors

Data annotators often work through crowdsourcing platforms or outsourcing vendors for as little as a few cents per task, with monthly earnings that fall well below local living wages.

Amazon Mechanical Turk workers in Venezuela have reported being paid in a form of scrip rather than legal tender—compensated with Amazon gift cards instead of cash due to U.S. sanctions and the country's banking isolation, constituting a violation of the right to fair and freely usable remuneration. The payment structure in Ghana, too, is uniquely cruel. Content moderators reportedly earn as little as US$80 a month—well below the average cost of living in Accra—supplemented by bonuses that, even at their highest, still fall short of covering basic expenses. While OpenAI, one of its clients, reportedly paid Sama US$12.50 an hour

for content annotation, the annotators received a take-home wage of around US$1.32 to US$2 per hour after taxes for reviewing between 150 and 250 passages of text of often disturbing content in a nine-hour shift. These conditions reflect a profound power asymmetry: most platforms are foreign, operate beyond the reach of local labor

not designed to benefit inmates, psychologically taxing data work—such as repeatedly tagging different meanings of the word "apple"—offers no real prospects for its workers. Contemporary understandings of slavery are industry-agnostic and define being compelled to work long hours under pressure with little to no pay as a component of modern

## Lebanon

University-educated Syrian refugees in Lebanon are constrained by restrictive labor policies, confining them to sectors such as agriculture, construction, and cleaning. Data annotation, while one of the few accessible income sources in the country, pays them anywhere between €0.014 to €0.04 per item—an unsustainable wage disproportionate to the labor involved. Efforts to supplement their earnings through global freelancing platforms such as Appen or Upwork are often thwarted by legal barriers. Refugees in Lebanon are prohibited from opening bank accounts or using financial services like PayPal and Payoneer, rendering them unable to receive international payments and pushing them further into economic marginalization.

laws, and classify workers as freelancers to avoid obligations like minimum wage, social protections, and the right to unionize. Consequently, many data annotators endure long hours for poverty-level wages, well below international standards for fair and decent work. Researchers have likened such "microwork" to the prison-industrial complex, arguing that just as physically grueling labor in prisons is

slavery.

International instruments stop at macro-level lawmaking, but crucial tenets of digital platform labor which distinguish it from traditional employment can be microscopic and highly specific. Present standards do not address obscure and opaque remuneration models, irregular payments to workers, or commissions

and fees paid by workers out of pocket—all common industry practices. Noted also in the ILO's assessment of normative gaps, "hours of work" are not defined in existing instruments and a numerical cap on hours has not been substantially discussed, all while exhausting workloads is a common feature of content moderation and data

among stakeholders—and in case of employers, strong opposition—to plug them. At the minimum, ILO's Protection of Wages Convention, 1949 sets out that wages shall be paid in legal tender at regular intervals, making operations in Venezuela highly irregular, even under a diluted international law regime.

## Pakistan

Amid incessant platform bans and a fragile geopolitical relationship with authorities, TikTok had hired up to 500 "vulgarity moderators" in Pakistan by the end of 2021, tripling its investment in local-language content moderation from 2020 to soothe government anxieties. Between April and June 2024, the platform removed 30 million videos for violating community guidelines. Behind these colossal numbers—a small team of overworked, underpaid, and distressed content moderators, who exercise extreme care in reviewing and removing high volumes of content, for even the smallest misstep can attract government scrutiny.

annotation industries. In context of fair pay or working conditions, the time data annotators spend learning new work, taking mandatory training courses, or simply waiting for the platform algorithm to allocate them new tasks is virtually unaccounted for. Since platform workers are at the disposal of the platform doing all these activities, they are in working time under the unitary model, and entitled to be appropriately paid. Even though recent ILO deliberations acknowledge these gaps, there is weak appetite

### Psychological Distress Comes with No Institutional Support

Content moderators are routinely exposed to highly traumatic content, without adequate mental health support or hazard compensation, leading many to develop severe mental health challenges and suicidal ideation. Training and courses platforms offer to new content moderators rarely prepare them for the psychological impact of content exposure or the systemic stressors embedded within the regimented moderation

process they must navigate. The work is highly standardized, tightly scripted, algorithmically monitored, time-tracked, and metric-driven. Mistakes—such as failing to identify harmful content, or incorrectly flagging allowable content—can lead to corporate penalties, and this pressure is passed directly to workers. An abundance of cheap labor and high unemployment also mean that content moderators are easily disposable: a failure to meet productivity benchmarks or complete assigned tasks can result in warnings, demotion, reassignment, or termination. The result is a mechanized and high-stress environment that leaves little room for mental respite or psychological safety, resulting in the proletarianization of white-collar workers.

One troubling trend emerging as a result among content moderators in the region is the near-total absence of effective psychological support. A report by *The Guardian* documents how Indian moderators are regularly exposed to

**India and the Philippines**

One moderator in Hyderabad, India reported psychological trauma after reviewing stabbing videos frame by frame. He notes that while some psychological services are available, these are often perfunctory, functioning more as compliance tools than genuine therapeutic services. Other moderators spoke of being socialized into desensitization—expected to normalize exposure to pedophilia, bestiality, and extreme violence. As one worker explained, "the whole world's trash is being dumped in India," while another compared the experience to entering a "torture chamber each day." The psychological toll sometimes manifests in behavioral shifts: one moderator recounted how a child abuse video left him acting in ways he could not explain, despite insisting, "I am not a bad person."

Moderators in Manila, Philippines described a workplace where nightmares, paranoia and obsessive ruminations were common consequences of the job. Several described seeing colleagues suffer mental breakdowns at their desks. One of them said he attempted suicide as a result of the trauma. Psychological distress in content moderators is also known to manifest somatically, damaging their physical health. Sleep disturbances—such as nightmares, often connected to the violence they had witnessed during work, and general difficulty falling or staying asleep due to anxiety and exhaustion—are common symptoms of the job.

harrowing content with minimal support. Genpact, an outsourcing firm serving Facebook in India, greets freshly recruited content moderators with shockingly graphic content off the bat—prompting concerningly high attrition rates. The US branch of the firm had established protections for its content moderators by 2019—workers could access round-the-clock on-site psychological counselors, claim health benefits, and temporarily blur graphic images before reviewing them so they can control how they engage with disturbing content. But Indian moderators were left to fend for themselves.

ILO Conventions 155 and 187, namely the Occupational Safety and Health Convention, 1981 and the Promotional Framework for Occupational Safety and Health Convention, 2006, together create a framework of protections for both physical and mental health of workers, including employee obligations on creating safe and healthy working environments. However, there is a persistent emphasis on both instruments to establish national mechanisms and programs to achieve these goals. The extent to which these frameworks are implemented by governments, enforced by regulators or put into practice by employers, though remains either unclear or demonstratedly low. In WHO's latest assessment of Member States' preparedness on this front, only 35% reported having a national program for work-related mental health promotion and prevention.

The absence of effective legal interventions further exacerbates already precarious working conditions, for instance in India, one of the most overworked countries in the world. Existing legislations, such as the Mental Healthcare Act, 2017, nominally affirm the right to mental well-being but places no actionable duty on employers to implement preventive safeguards. Meanwhile, the new labor codes, designed with a factory-floor understanding of workplace hazards, remain silent on work-related mental health. As such, the legal infrastructure is not only outdated but structurally incapable of addressing psychological injury in digital labor environments, with workers absorbing psychological injury as a condition of employment, often without adequate support, recognition, or compensation. Some countries are moving towards adoption of first principles: for instance, under the Vietnamese labor code, employers are prohibited from taking disciplinary action against workers suffering from a mental illness. However, clear mechanisms for workers to hold their employers accountable for injuries to mental health remain absent—especially in context of digital platform work.

### Sociolegal Barriers Impede Mobilizing

As of April 2025, Meta and its new outsourcing partner, Teleperformance, are facing a fresh wave of lawsuits in Ghana over psychological trauma, substance abuse, and suicidal ideation directly resulting from the nature of the counterparts in India, the Philippines, and other content moderation hubs elsewhere performing similar tasks for even lower pay and under comparable—if not harsher—conditions. This is, in part, because content moderators in India, the Philippines, and other jurisdictions face significant obstacles in pursuing redress. In India, these barriers are both

### The Philippines

In the Philippines, in addition to similar sociolegal barriers as India, workers are frequently harassed and stifled through red-tagging—an authoritarian practice of labelling an action as communist, subversive, or terrorist—effectively deterring unionization, resistance, and collective worker action. Targeted harassment, threats, and violence, which often escalates during collective bargaining negotiations, can stretch over several years and take an immense toll on workers and their families. Digital platform workers have attempted to call upon their employers to intervene and extend protections against red-tagging, but since many worker collectives refrain from calling themselves "unions" out of fear, they lose the legal right to bargain directly with corporations. Present webs of policies not only fail to protect platform workers from harms, but also fall short of ensuring them social security—of the employment descriptors used for platform workers (terms like freelancer, independent contractor, part-time worker), only those considered self-employed, employee or entrepreneur are covered by various social security and welfare laws.

work. In 2020, Meta agreed to a US$52 million settlement with more than 10,000 US-based content moderators who suffered psychological harm as a result of their work. Notably, this compensation excluded their structural and practical; they include low rates of unionization, the absence of class-action mechanisms, inadequate whistleblower protections, limited legal awareness, restrictive arbitration clauses, high litigation costs, and

pervasive social stigma surrounding mental health. Constant threats of job loss, reputational harm, and reduced employability—particularly in an oversaturated labor market—further discourage collective action.

Collective action may also be significantly decelerated due to a distinct cultural glorification of the IT industry, especially in India. Some moderators settle into perilous ecosystems deeming them their only ticket to a decent living, foregoing an array of occupational harms and injuries for a chance at survival. Others take pride in tech jobs, view the industry as exciting and promising of growth, and think their "dirty work" is necessary to "guard the world against harmful content on social media." A deep cultural veneration for an IT job can blind some to its ills, and generally dissuade resistance. While American workers view content moderation as a stepping stone to better IT opportunities, Filipino workers view the exploitative job as one of the best they can get. Moreover, non-disclosure agreements and strict organizational policies often prohibit moderators from discussing their work even with colleagues, let alone organizing. Amid unfavorable working conditions and weak labor protections, these factors generate a climate of fear, retaliation, blacklisting, and social ostracization

that effectively silences worker dissent and suppresses accountability.

Despite numerous barriers to resistance, worker unions have historically found avenues for collective action. In August 2024, for instance, thousands of motorcycle taxi gig workers in Indonesia circled the headquarters of ride-hailing apps Grab and GoJek wearing green jackets, protesting bleak working conditions and lack of legal protections. Now a regular sight in Jakarta, workers' protests of this nature signal towards an increasingly organized sector. Gig workers —usually a dispersed, app-dependent, and marginalized people of Indonesia—have found their voice in getting platforms and governments to recognize them and their concerns. Similarly, negotiations for central and state-level gig work platform regulation laws across India have almost exclusively been driven by gig worker collectives; 2 of these laws have been passed recently, and 7 await consultations and passage. India is additionally preparing, in deep consultation with worker unions, a Social Security Code which would extend expansive protections to digital platform workers.

For data annotators, collective action presents a more complex challenge. An entirely remote workforce where each worker performs atomized tasks alone in their homes can find unionization to be a logistical nightmare. Employee isolation,

however, is a deliberate feature of digital platform work. With other forms of platform work, like end-mile delivery logistics for instance, workers have opportunities to identify their own through uniforms and local clusters—a luxury difficult to emulate online. Managerial automation, additionally, poses a novel barrier to traditional organizing. Indian women engaged in crowdsourced data annotation work <u>strive to overcome most of these obstacles</u> through local Facebook and WhatsApp groups—navigating a faceless, nameless, non-human manager, though, remains an uphill battle.

# Pinning accountability for labor injustices

Despite well-documented evidence of harm, holding perpetrators accountable continues to be an uphill task. Platforms now claim to uphold certain standards to ensure occupational safety, health, and a fair working environment for content moderators. Teleperformance claims to maintain transparency with prospective content moderators about the nature of the work, including potential exposure to graphic material. These are communicated during recruitment interviews, detailed in employment contracts, and reinforced through training and psychological resilience testing. Meta, for its part, contends that it takes the well-being of content moderators seriously and that its contractual agreements with outsourcing partners include explicit provisions for counselling, training, and other forms of psychosocial support. In Turkey, TikTok content moderators working at Telus are made to use their designated "wellness breaks" for appointments with the in-house psychologist; TikTok moderators at Teleperformance in Colombia describe the support as inadequate.

Yet, even when taken at face value, these commitments amount to procedural assurances rather than evidence of substantive implementation, rendering these protections ineffective when applied in practice. In Indonesia, for instance, a patchwork of ancillary labor laws have the effect of imposing more responsibilities on workers than on platforms simply because of the vocabulary gig work platforms have chosen to describe their workers. *Law Permenhub 12/2019*, for instance, obligates the taxi worker, and not the platform, to fulfill the safety, security, comfort, affordability, and regularity of ride-hailing services. These platforms refer to themselves as tech companies rather than transportation companies, circumventing a vast web of transportation regulations—offloading them on marginalized taxi workers.

In early 2024, a Spanish court ruled that a former Facebook content moderator sought psychiatric treatment for damaged mental health owing to his work reviewing graphic content for Facebook—a case which

Facebook claimed it was "not a party to". Content moderators enjoy next to no agency and autonomy over the work they do and the work they do not; thus, the transparency over potential exposure to graphic material that platforms embed in contracts as part of worker well-being is rendered virtually meaningless if not accompanied by free choice. Even newer, more "worker-friendly" contracts retain the take-it-or-leave-it approach, shrouded in the illusion of choice and reform. For data annotators, being labelled as "freelancers" or "self-employed" further allows platforms to distance themselves from workers' realities. Refugee groups, especially undocumented migrants, using these platforms may not want to identify themselves for security reasons, let alone flag complaints and seek accountability from their employers.

Because of these structural bifurcations, tech companies continue to benefit from the Global Majority's labor and can selectively intervene in operational matters without assuming full accountability for the conditions under which it is performed. Central to these practices is the outsourcing model, which, combined with jurisdictional limitations, enables Western tech companies to evade accountability for harm and perpetuate unequal treatment in the Global

Majority workforce, mirroring repression and exploitation from a shared colonial past.

# Assessing the international labor governance regime

The dramatic growth of global supply chains since the aughts has created considerable challenges for effective labor governance. Significant labor governance mechanisms, spanning over 200 ILO Conventions, for instance, are centered around national labor relations and service domestic employers and labor. With the bulk of lawmaking, regulation and governance happening at the national level, outsourcing labor has already created difficulties in the achievement of workplace compliance. A boom in digital platform labor with obscure employment arrangements and active contractual obfuscation by transnational corporations—where new categories of workers engage in new forms of labor without even the bare minimum standards of protections or redress met—has further complicated matters for regulators.

Presently, there are no international instruments that apply squarely to content-related digital labor. The United Nations' (UN) framing of digital platform economy has been restricted to discussions on digital inclusion, personal data protection, or socioeconomic marginalization. In

March 2023, the ILO decided to begin deliberations and standard-setting on decent work in the platform economy through 2025-26. Efforts to legislate on this end began with an assessment of normative gaps in international labor governance led by the ILO Governing Body, which examined legal lacunae in the application of international labor standards to digital platform labor as well as specific rights issues that may not be fully addressed in existing ILO frameworks.

Primarily, it found, standards thus far applied only to "employees" and not to self-employed workers—many forms of freelance digital labor, including crowdsourced work like data annotation, deems workers as self-employed. Other identified gaps included simply the realization that existing ILO standards did not adequately cover the cross-border nature of digital platform work, did not protect worker personal data, and failed to address certain wage and employment security concerns such as "hours of work". Subsequently in January 2024, the ILO published its report and corresponding framework on realizing decent work in the

platform economy. As a starting point, this report studied existing regulatory frameworks—of Member States and its own—and presented a baseline standard which may be collaboratively reviewed and revised by a range of stakeholders. Previously identified normative gaps were fixed to an extent—while self-employment was painstakingly accommodated for at every turn, more contentious themes of algorithmic governance were weaved in merely as open-ended questions. Refraining from prescriptive regulation at most turns, the report concluded with a questionnaire, calling upon stakeholders to opine on the minutiae of the ILO's new framework.

The ILO operates on a unique pluralist tripartite structure comprising equal representation from governments, national worker collectives, and employers. Responses from Member States, encapsulated in a February 2025 report, are rife with conflict between the three quarters. Though agonistic pluralism is seen, in most quarters, as a foundational catalyst of labor governance, fundamental misalignments from significant stakeholders can often decelerate rights-first lawmaking—allowing worker harms the time to firmly embed into supply chains. Policy decisions at the ILO are arrived at with consensus. Consensus, as the ILO defines, is characterized by the absence of any

objection. Objections are not uncommon in multi-stakeholder bodies—but have also been used as effective tools for delays. The 113th and 114th sessions of the International Labor Conference (slated June 2025 and 2026) will see attempts to mediate conflicts and arrive at a new standard on decent work in the platform economy. But absent normative force, these standards may not move the needle on present-day grievances in any appreciable way.

## EU Directive on Working Conditions in the Platform Economy goes beyond ILO standards to provide worker protections.

The ILO has adopted close to 200 Conventions since its conception, but ratification remains uneven—core labor standards may be more widely ratified, while technical standards have lower uptake. Over the last three decades, there has been a shift from conventions towards soft law instruments—such as declarations and recommendations—due to difficulties in enforcing binding norms in a globalized economy. Aligning with the ILO's transnational multi-stakeholderism may become useful to governments and corporations merely to improve the credibility of their mechanisms—like corporate codes of conduct, ESGs or supply chain standards—by linking

them to any of the 200 conventions. Generally, aesthetic value aside, businesses do not view ILO conventions as binding. At best, they may become discursive weapons in national collective bargaining. Enforcement is weak, relying on supervision and reputational pressure rather than sanctions. States largely support these normative works when aligned with national interests, but there is significant resistance due to perceived intrusions on sovereignty.

In contrast, efforts to legislate on digital platform labor see better reception in the EU. EU Directive 2024/2831 on Working Conditions in the Platform Economy goes beyond proposed ILO standards to provide additional worker protections. Provisions prohibiting many categories of personal worker data processing, extending equal protections to intermediary models, and others mandating regular labor inspections, access to dispute resolution mechanisms, and right of redress channels make the Directive a comparably cohesive charter of rights. However, the transnational nature and hypermobility of this work are in direct conflict with the instrument's limited geographical reach.

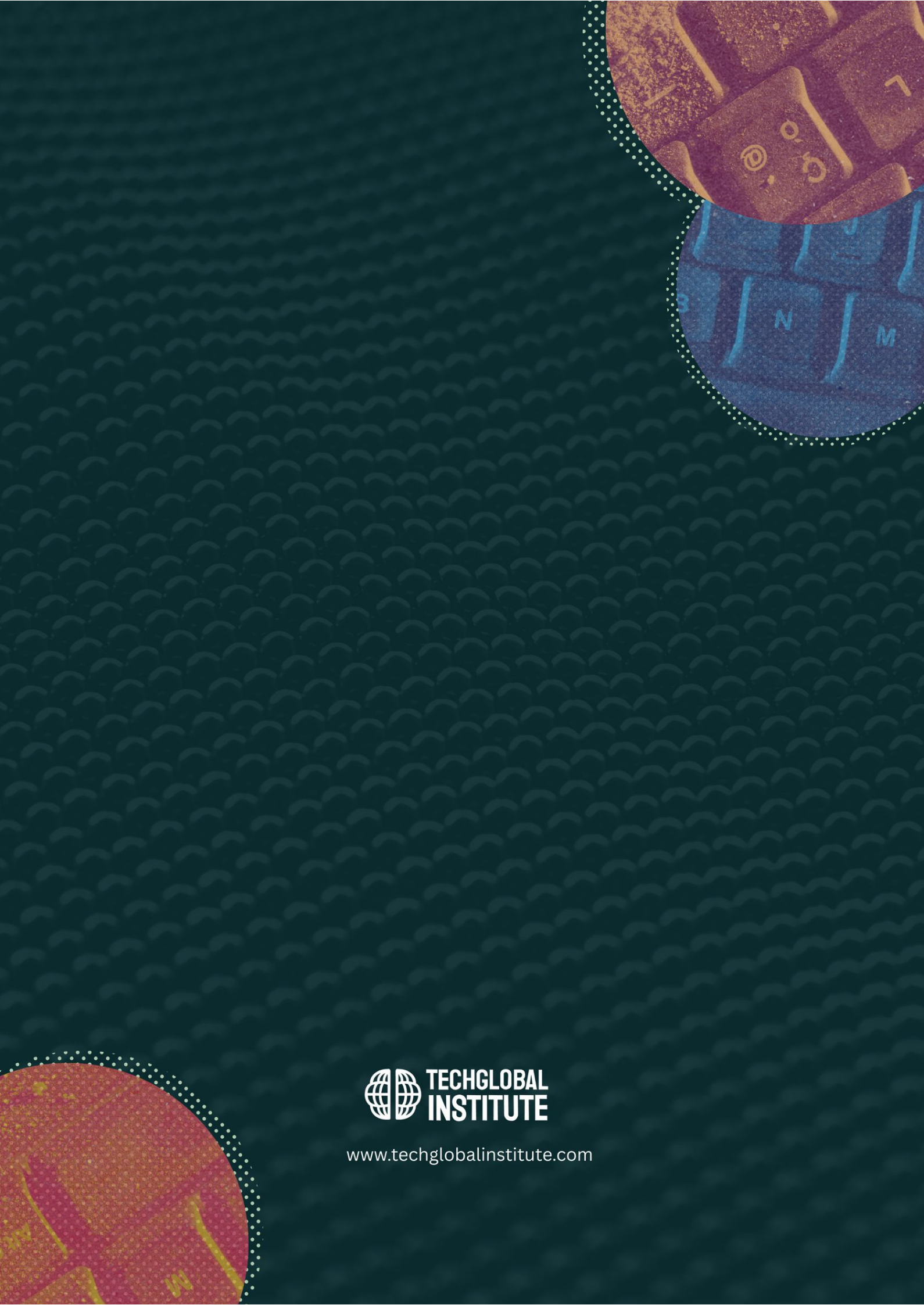A potential rights issue that has gone undiagnosed in formal international deliberations is surveillance and information asymmetry between workers and platforms. As outlined by Privacy International in their response to the ILO questionnaire, workers are facing an unprecedented amount of surveillance from the platforms they work for, who routinely capture and process worker data to make decisions that can affect them. Content moderators, for instance, have reported working in highly surveilled and policed environments. Data annotators, too, especially on crowdwork platforms, are deeply scrutinized and inexplicably terminated for the smallest of errors. Such invasive tools, taken together with operational secrecy and contractual opacity, redefine the relationship between workers and their employers and create power imbalances. While the ILO and EU standards stress on transparency, especially in context of algorithmic decision-making, and attempt data protection to an extent, worker surveillance is a thematic and normative gap regulators have not identified yet. Uruguay's new labor legislation, Law no. 20396 of February 2025, has approached this concern from a novel lens, where "digital reputation" is considered a private and portable capital of the platform worker. Workers have a right to the intangibility of their digital reputation, and can access all data collected about them by the employer or platform during the employment relationship and up to one year after its termination.

# Conclusion

An invisible content moderation and data annotation workforce, hailing largely from the Global Majority, powers platforms and complex AI systems worldwide. This labor is extracted by tech companies from cheap labor markets while remaining distant and largely unaccountable to the workers who make their business a profitable reality. While both industries create opportunities for dignified and accessible work in the jurisdictions they operate in, they are also marked by a wide spectrum of harms and precarity, warranting platform accountability and policy reform. There is a pressing need to not only point to these structures of power within labor landscapes, but also investigate and map the socioeconomic enablers in South and Southeast Asia that make these markets fertile ground for outsourced digital labor. As a starting point, this essay identifies and flags labor rights issues in global tech supply chains.

Both industries are marked by vague hiring practices and an opaque wall between workers, management, and platforms—and opacity continues in operations, expectations, performance metrics, as well as decisions related to termination. The industries are also thus notoriously precarious and termination often unappealable. Cruel work hours and low wages continue to be a labor issue without many domestic protections as some of this work may fall outside the purview of domestic labor law frameworks. Psychological distress and resultant impact on worker well-being has come to be a common feature of content moderator jobs, with little to no support from intermediaries or platforms. Companies extract labor from socioeconomically vulnerable zones or populations, and have the freedom to relocate their business to jurisdictions with more amenable regulation, leaving workers in a parlous state. Despite the numerous labor rights issues, workers may find it difficult to unionize or collectively bargain for their welfare due to cultural factors, remoteness of work, or skewed power dynamics with platforms. The present international labor law order also does not adequately protect this industry from labor harms—but with ongoing deliberations on digital platform work at the ILO, there may be a more comprehensive regulatory framework on the horizon.