

BETTER IN MODERATION: OCCUPATIONAL RISKS AND HAZARDS TO THE MENTAL HEALTH OF FILIPINO CONTENT MODERATORS*

*Danielle D. Abuel***

ABSTRACT

A documentary film entitled “The Cleaners” went viral on the Internet in 2018. The film followed the work of Filipino social media content moderators tasked with the harrowing job of “cleaning” the internet, which often required exposure to violent, sexually explicit, and other disturbing content, to determine what is deemed appropriate for their sites. Presumably, this grabbed the attention of the public because of the secretive nature of the work.

Owing to the large demand and the ever-growing population of social media sites, the workers in content moderation are more likely to be outsourced from third parties to save on costs. The Philippines, through Business Process Outsourcing (“BPO”), is a major provider of said workers.

This paper aims to identify the mental health-related problems faced by workers, especially concerning the occupational safety hazards that accompany the mentally burdensome nature of their work. It explores the policies in support of workers exposed to these hazards, as found in various laws and issuances, and the methods of implementing these policies. An overview of our existing body of labor law suggests a lack of space reserved for outsourced Filipino content moderators. Hence, this paper also hopes to discuss other kinds of recourse which are unavailable, but necessary.

* Cite as Danielle D. Abuel, *Better in Moderation: Occupational Risks and Hazards to the Mental Health of Filipino Content Moderators*, 98 PHIL. L.J. 148, [page cited] (2024).

An earlier version of this paper was awarded the Perfecto V. Fernandez Prize for Best Paper in Labor Law (2024), University of the Philippines College of Law.

** J.D. (2024) & B.A. in Creative Writing (2018), University of the Philippines.

I. INTRODUCTION	149
II. WHO ARE THE CLEANERS?.....	150
III. OVERVIEW OF PHILIPPINE LABOR POLICY	156
IV. LEGAL FRAMEWORK ON OCCUPATIONAL HEALTH AND SAFETY	160
A. Establishing the Occupational Safety and Health Center (OSHC).....	160
B. OSH for Call Center Workers, Under D.C. No. 1-08	163
C. Strengthening OSH Standards through Republic Act No. 11058.....	165
D. Workplace Mental Health Protections Under the Mental Health Act.....	167
E. National Policy Framework on Promotion of Healthy Workplace	172
F. Content Moderation as a Novel Issue in Philippine Labor Law	174
G. Protection of Content Moderators’ OSH in Criminal Law.....	176
V. COMPARATIVE ANALYSIS: OSH IN DIFFERENT CONTEXTS.....	178
A. India	178
B. Australia.....	182
VI. PENDING PHILIPPINE LEGISLATION AND RECOMMENDATIONS.....	187
VII. CONCLUSION.....	191

I. INTRODUCTION

The average Filipino spends four to five hours a day on social media. This is a step above, almost doubling, the global average of two to three hours. An estimated 83% of Filipinos are internet users according to a Department of Information and Communication Technology (DICT) official.¹

In its report, the DICT found that many users, particularly the younger demographic, rarely use the web or search engines to learn helpful information, as more of them tend to consult social media when faced with personal queries, making it a vital tool not only for entertainment or communication but also for information.² Social media sites thus pose a great challenge – and consequently, a great burden: how to effectively manage and

¹ Jose Rodel Clapano, *DICT: 83% of Pinoys are internet users, but...*, PHILSTAR.COM, June 4, 2023, at <https://www.philstar.com/headlines/2023/06/04/2271289/dict-83-pinoys-are-internet-users-but>.

² *Id.*

monitor the heavy volume of content posted on their platforms every day, every hour. It is no surprise that many of these platforms still face the issue of regulation, as they attempt to find a balance between maintaining a safe space and censoring their users outright.³

To maintain a pleasant user experience, social media companies need to protect their platforms, particularly from graphic or violent material, which would be unsafe for viewing by the average person. Normally, platforms make use of automated technologies to flag down such content, but the technology is still far from perfect and cannot always capture the context of each specific post. A human reviewer would be needed to screen the content, leaving to them the review of more complex questions, such as what would be considered “permissible” violent content.⁴ This ultimately gave rise to the industry of content moderation.

This Note aims to analyze the current body of Philippine labor law to understand whether it legally recognizes, or whether it should recognize, the occupational hazards of the “cleaning” work of content moderation in contrast to BPO work. Part II introduces who are the “cleaners” and their work; Part III lays down the overview of Philippine labor policy and addresses the question of how our current body of law protects Filipino workers in content moderation; Part IV examines the framework of Occupational Safety and Health (OSH) in the context of the content moderation industry and how the laws address or attempt to mitigate the hazards involved in the work; Part V provides a comparative analysis of OSH in different contexts; Part VI explores pending legislation and possible recommendations; and Part VII concludes the Note with the submission that content moderation should be considered as a hazardous work process, whose employees should be entitled to protection and compensation.

II. WHO ARE THE CLEANERS?

Content moderation, also known as “cleaning the internet,” is a process wherein a moderator looks at photo, video, or text content flagged

³ See also Esade Business & Law School, *Should Social Media Platforms Be Regulated?*, FORBES.COM, Feb. 10, 2020, at <https://www.forbes.com/sites/esade/2020/02/10/should-social-media-platforms-be-regulated>.

⁴ Kalev Leetaru, *Why Don't Social Media Companies Stop Violent Imagery?*, FORBES.COM, Apr. 19, 2017, at <https://www.forbes.com/sites/kalevleetaru/2017/04/19/why-dont-social-media-companies-stop-violent-imagery>.

down by users as inappropriate for whatever reason and determines whether or not it should stay up on the site, depending on its specific policies.⁵

The field of content moderation gained further notoriety in 2018 with the release of the documentary *The Cleaners*.⁶ The film gave viewers a look into the life of Filipino professionals who work as content moderators, and the motivations behind their work. Yet despite the secretive and closed-off nature of the work as portrayed in the film, one notable takeaway from *The Cleaners* was the extent of the demand for content moderators.

As depicted in the film, despite content moderation being a relatively new line of work, there is no denying its demand, given the ever-expanding myriad of content on social media.

Meta Platforms, or Meta, the company that owns Facebook and Instagram, stated in its Transparency Report⁷ that in the second quarter of 2024, Facebook took action on 14.9 million pieces of violent and graphic content, while Instagram took action on 10.3 million posts for the same violation. Violent and graphic content is only one among many of the violations provided for by Meta in their platforms' policies. For adult nudity and sexual activity, Meta reported taking action on 32.2 million posts in the same quarter.⁸

Meanwhile, in the same quarter, TikTok reported to having manually removed about 178 million total videos for content violations, with 31% or about 55 million videos removed due to sensitive and mature themes.⁹

With all the inappropriate content running rampant on the most popular social media platforms, *who* is in charge of cleaning?

⁵ Thomas Stackpole, *Content Moderation Is Terrible by Design*, HARV. BUS. REV., Nov. 9, 2022, at <https://hbr.org/2022/11/content-moderation-is-terrible-by-design>.

⁶ THE CLEANERS (Gebrueder Beetz Filmproduktion 2018).

⁷ Meta, *Community Standards Enforcement Report: Violent and Graphic Content*, META TRANSPARENCY CENTER, at <https://transparency.meta.com/reports/community-standards-enforcement/graphic-violence/facebook/>.

⁸ Meta, *Community Standards Enforcement Report: Adult Nudity and Sexual Activity*, META TRANSPARENCY CTR., at <https://transparency.meta.com/reports/community-standards-enforcement/adult-nudity-and-sexual-activity/facebook/>.

⁹ TikTok, *Community Guidelines Enforcement Report: April – June 2024*, TIKTOK TRANSPARENCY CTR., Sept. 26, 2024, at <https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2024-9>.

A lot of social media platforms, especially during their infancy, did most of their “cleaning” in-house—only outsourcing help from digital firms and only in specific areas needed. But as their user base grew, so did their demand. These platforms inevitably had to resort to outsourcing their workers to third-party contracting companies, or through business process outsourcing (“BPO”), as both an efficiency and a cost-saving measure. Often, workers in BPO come from countries that provide cheap labor and typically work in large-scale, industrialized offices specifically geared to carry out a certain kind of work.¹⁰

Globally, the top provider of BPO services is India, owing primarily to its expertise in information technology (“IT”) and software development.¹¹ In 2022, over 7.4% of India’s GDP came from the IT-BPM (Information Technology and Business Process Management) sector, underscoring the world’s reliance on services provided by Indian BPO employees.¹²

As a provider, the Philippines falls very closely behind India. According to the Information Technology and Business Process Association of the Philippines (“IBPAP”), the BPO industry alone recorded 29.1 billion dollars’ worth of revenue in the first half of 2022, with over 1.44 million Filipino workers employed full-time.¹³ The International Labour Organization (ILO) found that the accelerated growth of BPO in the Philippines was driven by a host of factors, most significantly the low labor costs coupled with the educated and highly skilled workforce.¹⁴ The soft skills, proficiency in English, and cultural affinity to North American markets¹⁵ in particular, play a crucial role in this growth, given that several BPO companies in the Philippines offer customer support services.

¹⁰ Stackpole, *supra* note 5.

¹¹ Tricia Pacete, *BPO situationer: PH poised to capture a bigger slice of the BPO global market share*, COLLIERIES, Sept. 20, 2022, at <https://www.colliers.com/en-ph/news/bpo-situationer-philippines-poised-capture-bigger-slice-global-market-share>.

¹² Ernestas Naprys, *Obsessed with chips, US overlooks its dependence on India*, CYBERNEWS, Jul. 13, 2023, at <https://cybernews.com/tech/india-outsourcing-software-chips>.

¹³ Pacete, *supra* note 11.

¹⁴ Lorenza Errighi, Sameer Khatiwadda, & Charles Bodwell, *Business process outsourcing in the Philippines: Challenges for decent work* 9 (ILO Asia-Pacific Working Paper Series, Dec. 2016) at https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---sro-bangkok/documents/publication/wcms_538193.pdf.

¹⁵ Pacete, *supra* note 11.

While BPO more commonly pertains to call and contact centers, it can refer to a variety of other skilled work as well, such as search engine optimization (SEO) and copywriting, IT and software development, and of course, content moderation.

Most people have a general idea of what call centers and IT support centers look like and how they operate, given their prevalence. But not all are familiar with the work of content moderation. Hence, the question: what does the average workday of a Filipino content moderator look like?

Content moderators go to work in a call center-like office setup, which is typically a shared office without designated personal workspaces. They then log onto whatever software is used or required by the client, where they can access queues of materials. In certain cases, the company providing content moderation will make use of a software separate from that of the client company. Either way, a flagged piece of content is served to the content moderator through the relevant software, after which they will decide whether to delete the content based on their interpretation of the client company's policies. After closing one case, they get another – a process that repeats all through the day as they work through the never-ending queue.¹⁶

These softwares are highly productivity-based. Reportedly, some can automatically boot them out if they take a break that lasts more than eight minutes, thus discouraging long bathroom breaks or chatter in the halls.¹⁷

Stackpole also reports that due to the growing use of artificial intelligence (“AI”)—or more specifically, machine learning algorithms—a portion of content moderation involves cleaning up after the work of automation. Some content moderators may be tasked to annotate data sets of media that will be used to train the AI. Depending on the system being used by the company, they may also be made to take action on posts that have been automatically flagged by AI. Naturally, AI is yet unable to capture any social and cultural nuances that only humans can detect—hence, the necessity to double-check. This is another way that content moderators “clean” the internet.¹⁸

¹⁶ Stackpole, *supra* note 5.

¹⁷ Elizabeth Dwoskin, *Inside Facebook, the second-class workers who do the hardest job are waging a quiet battle*, WASH. POST, May 8, 2019, <https://www.washingtonpost.com/technology/2019/05/08/inside-facebook-second-class-workers-who-do-hardest-job-are-waging-quiet-battle>.

¹⁸ Stackpole, *supra* note 5.

On a typical day, content moderators are exposed to all kinds of reported media. This includes violent material, posts referencing or even outright displaying suicide, hate speech, and in worse cases, child exploitation.

By scouring through thousands of reported images, videos, and posts every single day, a content moderator is at risk of being exposed to disturbing content multiple times a day. Occasionally, they also have to deal with outdated interfaces, preventing them from reviewing flagged content efficiently. This could also mean being exposed to a piece of disturbing content longer than is necessary, which only exacerbates the stress of the job.¹⁹

The risks discussed above are enough to warrant protections to ensure the mental and psychological well-being of content moderators who are constantly exposed to disturbing content. However, the aggravation caused by the demands that are typically associated with BPO work, which places emphasis on metrics and productivity, must also be considered. After all, there is an overlap between BPO workers in call center work and those in content moderation.²⁰ This overlap extends to the very terms of the job, which includes having to pick up night shifts, with workers jumping from one call (or post) to another and relying on a strict guide to determine the course of action to be taken. In this regard, content moderators and BPO workers in general, face common problems and challenges in their everyday work.

One such challenge faced by both is the excessive focus on performance metrics. The BPO companies would constantly raise and push the quotas set by their clients at cheaper costs, greatly affecting targets for employees.²¹ However, call center work is nothing like content moderation. Psychologists agree that the mechanized routines of call center work are ill-suited to a job that should require breaks, psychological support, and time to

¹⁹ *Id.*

²⁰ Elizabeth Dwoskin, et al., *Content moderators at YouTube, Facebook and Twitter see the worst of the web—and suffer silently*, WASH. POST, Jul. 25, 2019, at <https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/>.

²¹ Michael Sainato, *'No other way to fight back': Philippines call center workers battle unfair quotas*, GUARDIAN, Nov. 21, 2018, at <https://www.theguardian.com/business/2018/nov/21/no-other-way-to-fight-back-philippines-call-center-workers-battle-unfair-quotas>.

process emotionally harrowing material.²² Although BPO companies in content moderation would deny having quotas, many workers claimed to have been pressured to meet targets for accuracy and to review as many posts as possible during a shift, which made it harder to take allotted breaks. The effect was, thus, still similar to having a quota.²³

Why is there pressure on the part of workers to meet metrics and quotas? For one, failure to satisfy them could result in termination. Companies are allowed to enforce this, from a purely legal standpoint. Even recently decided cases hold that management prerogative—the right of employers to create and enforce their own company standards within the bounds of the law—may be availed of so long as they are exercised in good faith for the advancement of the employer's interest.²⁴

For another, poor performance, while not necessarily resulting in termination, could still inhibit a worker's professional growth, and prevent them from career opportunities such as raises or promotions, or from receiving a good recommendation from their supervisor.

This poses another challenge altogether for BPO workers. Due to the extreme focus on productivity and metrics, BPO workers who are terminated for failing to cope, or who stagnate professionally due to their substandard performance, would often need to apply to other companies in the same or a similar industry, essentially compelling them to “start over.”²⁵ Thus, there is very little opportunity for professional growth for the workers, and in this sense, a lack of security of tenure.

KABATAAN Party-list Representative Raoul Manuel, in his Explanatory Note for House Bill No. 8189 (H.B. No. 8189) or “The Magna Carta for BPO Workers,” also enumerated several key issues faced by workers in the BPO-IT industry. Among them is the prevalence of health and safety risks and hazards, as BPO workers face many health problems rooted in the nature of their work.²⁶ This can only be exacerbated by the

²² Dwoskin et al., *supra* note 20.

²³ *Id.*

²⁴ *Telephilippines, Inc. v. Jacolbe*, G.R. No. 233999, 893 SCRA 210, 224, Feb. 18, 2019, *citing* *Buiser v. Leogardo*, G.R. No. 63316, 131 SCRA 151, July 31, 1984. *See also* *Babar v. IBEX Global Solutions (Philippines), Inc.*, G.R. No. 249889, Aug. 19, 2020; *Magno v. Sutherland Global Services Philippines, Inc.*, G.R. No. 235821, Apr. 23, 2018.

²⁵ *Sainato*, *supra* note 21.

²⁶ H. No. 8189, 19th Cong., 1st Sess., Explanatory Note (2023). *Magna Carta for BPO Workers*.

existing problems they already face as BPO workers, such as the productivity demand, and as mentioned, the lack of security of tenure.

As for content moderation, the work itself can have negative effects on the mental state of the moderators, due to the latter's heightened and constant exposure to violent, sexually explicit, and other disturbing or graphic content.²⁷ Content moderation, in particular, is thus likely to pose safety and health hazards for its workers since content moderators are charged with the responsibility to make social media sites a safe space for their users, this comes with the price of compromising their mental wellbeing. Many workers have reported developing trauma, depression, and even suicidal tendencies, from the content they are made to review every day.²⁸ In addition, health studies have observed a connection between the work carried out in content moderation with mental illness, in particular, the prevalence of Post-Traumatic Stress Disorder ("PTSD") among reviewers.²⁹

Despite this growing problem, "there has been little formal study on the impact of content moderators' routine exposure to such imagery."³⁰ There is little research on how the law can protect these workers, and to what extent.

III. OVERVIEW OF PHILIPPINE LABOR POLICY

The Philippines adopts a general policy on full protection to labor. This general policy is dissected into what is now known as the seven cardinal labor rights of workers, as provided under Article XIII, Section 3 of the 1987 Constitution:

[The State] shall guarantee the rights of all workers to self-organization, collective bargaining and negotiations, and peaceful concerted activities, including the right to strike in accordance with law. They shall be entitled to *security of tenure*, *humane conditions of work*, and a living wage. They shall also

²⁷ Anna Drootin, "Community Guidelines": *The Legal Implications of Workplace Conditions for Internet Content Moderators*, 90 FORDHAM L. REV. 1197, 1206–07 (2021).

²⁸ Dwoskin et al., *supra* note 20.

²⁹ Ysabel Gerrard, *The COVID-19 Mental Health Content Moderation Conundrum*, SOC. MEDIA SOC'Y 6(3) (2020), available at <https://journals.sagepub.com/doi/10.1177/2056305120948186>.

³⁰ Dwoskin et al., *supra* note 20.

participate in policy and decision-making processes affecting their rights and benefits as may be provided by law.³¹

It is important to establish first that Article XIII, Section 3 is not a self-executing provision. In the landmark case of *Manila Prince Hotel vs. Government Service Insurance System*,³² the Court held that a provision that lays down a general principle is usually not self-executing, but one that becomes operative without the aid of legislation, or that which supplies sufficient rule by means of which the right it grants may be enjoyed or protected, is self-executing. In case of doubt, the Constitution should be considered self-executing rather than non-self-executing.³³

Based on the doctrine, the aforesaid provision would indicate that it is not self-executing, being that it provides for a broad coverage of rights, therefore requiring the aid of legislators to maximize understanding and application of said rights. Moreover, the very wording of the provision, in particular the inclusion of the phrase, “as may be provided by law,” although attached to a specific right therein, suggests by analogy that the other rights similarly require the aid of legislation.

The case of *Agabon vs. National Labor Relations Commission*³⁴ would later support this. It clarified that Article XIII, Section 3 is not self-executing, based on the intent of the framers, and considering that it would be highly unrealistic to expect courts to be able to approximate on their own the extent of rights being protected without the aid of a law or regulation by the appropriate agency defining the parameters of these guaranteed rights. However, the Court in this case also opined *obiter dictum* that an examination of the provision is warranted to determine whether it is complete in itself as a definitive law, or if it needs future legislation for completion and enforcement.³⁵

Among the cardinal rights, it is the right to security of tenure and the right to humane working conditions which benefit content moderators the most, in light of the unique problems posed by their work.

Applied, and of course, within the bounds of existing laws to be later discussed, the right to humane working conditions could afford content

³¹ CONST. art. XIII, § 3. (Emphasis supplied.)

³² G.R. No. 122156, 267 SCRA 408, Feb. 3, 1997.

³³ *Id.* at 432.

³⁴ G.R. No. 158693, 442 SCRA 573, Nov. 17, 2004 (Tinga, *J.*, concurring).

³⁵ *Id.* at 668.

moderators remedy from any ill effects of their work on their mental health. This could come in the form of health benefits, more conducive work environments, or extended sick leaves, among others. Consequently, security of tenure prevents workers from being unjustly terminated or discriminated against on the basis of suffering the aforementioned ill effects. A content moderator, for example, who falls behind on performance metrics due to extreme mental exhaustion or psychological distress should be met with more lenience than the average worker and be entitled to any medical support they need. Such applications would, naturally, depend on the result of legislation.

Apart from the policy laid down in the Constitution, the Philippines also ratified the International Labour Standards on Occupational Safety and Health (“ILO Standards on OSH”), specifically ratifying the Promotional Framework for Occupational Safety and Health Convention³⁶ in 2019.

According to the ILO, the Convention aims to establish and implement coherent national policies on occupational safety and health (“OSH”) through dialogue between government, workers’ and employers’ organizations and to promote a national preventive safety and health culture.³⁷ The Convention mandates member-states to take active steps toward securing a safe and healthy working environment through a system of defined rights, responsibilities, and duties, where the principle of prevention is accorded the highest priority.³⁸ Of importance is the fact that the ILO also considers mental and behavioral disorders as occupational diseases, in particular:

2. Occupational diseases by target organ systems

* * *

2.4. Mental and behavioural disorders

2.4.1. Post-traumatic stress disorder

2.4.2. Other mental or behavioural disorders not mentioned in the preceding item where a direct link is established scientifically, or determined

³⁶ See ILO Promotional Framework for Occupational Safety and Health Convention [hereinafter “Promotional Framework for OSH”], Convention No. 187, June 15, 2006. Entry into force on February 20, 2009.

³⁷ See ILO Occupational Safety and Health Convention [hereinafter “ILO OSH Convention”], Convention No. 155, June 22, 1981.

³⁸ Promotional Framework for OSH, art. I(d).

by methods appropriate to national conditions and practice, between the exposure to risk factors arising from work activities and the mental and behavioural disorder(s) contracted by the worker.³⁹

However, the provision is embodied in a Recommendation, a separate instrument that merely serves as non-binding guidelines. Thus, the Philippines can adopt the provisions of such Recommendation at its discretion. As the Convention itself pertains to general policy, the recognition of specific illnesses as occupational diseases would be left to the will of member-states, subject to general principles laid down in said Convention.

The Convention, likewise, cannot be said to be “self-executing” in the general sense. Even assuming that the Convention requires the recognition of specific occupational diseases, the Convention has yet to be concurred in by the Senate, which is a constitutional requirement for treaties.⁴⁰ Afterward, it would then have to be domesticated, wherein a State incorporates an international law and weaves it into its domestic legislation.⁴¹

In 2022, ILO amended its Fundamental Principles and Rights at Work by adding “a safe and healthy working environment” to its commitments.⁴² While this is an important change that could potentially aid the plight of Filipino content moderators, it also poses the same question of legislation, *id est*, how will our domestic lawmakers ultimately decide to integrate it into our body of law?

³⁹ List of Occupational Diseases Recommendation, ILO No. R194, Annex, June 20, 2002, *amended* Mar. 25, 2010.

⁴⁰ CONST. art. VII, § 21.

⁴¹ Rommel Casis, *Domesticating International Law: Resolving The Uncertainty And Incongruence*, 19 PHIL. Y.B. INT’L L. 128 (2020), at <https://law.upd.edu.ph/wp-content/uploads/2021/11/Domesticating-International-Law-Resolving-the-Uncertainty-and-Incongruence.pdf>, *citing* Anthony D’Amato, *The Coerciveness of International Law*, 52 GER. Y.B. INT’L L. 437, 443 (2009).

⁴² ILO Declaration on Fundamental Principles and Rights at Work, ILO WEBSITE, June 18, 1998, *amend.* 2022, *available at* <https://www.ilo.org/ilo-declaration-fundamental-principles-and-rights-work/about-declaration/text-declaration-and-its-follow>.

IV. LEGAL FRAMEWORK ON OCCUPATIONAL HEALTH AND SAFETY

A. Establishing the Occupational Safety and Health Center (OSHC)

Executive Order No. 307 (“E.O. No. 307”), promulgated by Former President Corazon Aquino in 1987, established the Occupational Safety and Health Center (OSHC) in the Employees' Compensation Commission (ECC), an agency attached to the Department of Labor and Employment (DOLE).⁴³

Presidential Decree No. 626 established the ECC to promote and develop a compensation program whereby employees and their dependents, in the event of work-connected disability or death, may promptly secure various benefits.⁴⁴ Meanwhile, the OSHC carried the mandate of assisting government agencies and institutions in the formulation of OSH policies and standards and issuing technical guidelines for the prevention of occupational diseases and accidents.⁴⁵

DOLE, through the ECC, issued the Implementing Rules and Regulations (“IRR”) for E.O. No. 307. The IRR defines occupational health and safety as follows:

Sec. 2. *Definition of Terms.* – The terms as used in these Rules shall have the following meanings:

* * *

9. Occupational Safety and Health (OSH) – is defined as the:
 - a) promotion and maintenance of the highest degree of physical, mental and social well-being of workers in all occupations;
 - b) prevention among its workers of any departures from health *caused by their working conditions*;
 - c) *protection among workers in their employment from risks usually from factors adverse to health*; and

⁴³ Exec. Order No. 307 (1987). Establishing the Occupational Safety and Health Center in the Employees' Compensation Commission.

⁴⁴ Pres. Dec. No. 626 (1972), *as amended*, art. 166.

⁴⁵ Exec. Order No. 307 (1987), § 2(f).

- d) placing and maintenance of the worker in an occupational environment adapted to his/her physiological ability.⁴⁶

“Worker” and “workplace” are defined in the same section as follows:

10. Worker – any member of the labor force, whether employed or unemployed, wage or non-wage
11. Workplace – refers to the office, premises or worksite where a worker is [temporarily] or habitually assigned. [...].⁴⁷

That the IRR does not distinguish between a regular and contractual employee is important to note, since BPO workers are not always regularized. Thus, even content moderators with a contractual status should be able to benefit from OSH standards, assuming that there are protections provided and applicable to them.

Gleaning solely from the definitions provided, content moderators should then be covered by the protections offered by E.O. No. 307 and its IRR. However, in the annex of the IRR, occupational hazards are limited to specific conditions only:

8. Occupational Hazards – refers to various environmental factors or stresses that can cause sickness, impaired health, or significant discomfort in workers and can be classified as *chemical, physical, biological or ergonomic*.⁴⁸

Likewise, in the Department of Labor and Employment Memorandum Circular No. 2, Series of 1998 (“DOLE M.C. No. 02-98”), or the Technical Guidelines for Classifying Hazardous and Non-Hazardous Establishments, Workplaces and Work Processes, none of the criteria for classifying hazardous establishments or workplaces contemplate an office set-up, instead focusing more on manual types of labor such as drilling, machining, and the like.⁴⁹ Neither do they contemplate the psychological hazards involved in content review.

⁴⁶ Exec. Order No. 307 Rules & Regs (1987), § 2 (10). (Emphasis supplied.)

⁴⁷ § 2 (11)–(12).

⁴⁸ Ann. I. Glossary of Terms, (8). (Emphasis supplied.)

⁴⁹ Dep’t of Lab. & Emp’t (DOLE) Mem. Circ. No. 2-98 (1998), §§ 3–4. Technical Guidelines for Classifying Hazardous and Non-Hazardous Establishments, etc.

Thus, whatever protections offered by E.O. No. 307 to content moderators would be very limited. The importance of assigning specific activities as risk or hazard cannot be further underscored. This is because of the Employees' Compensation Program ("ECP"), which is a government program designed to provide a compensation package to public and private sector employees and/or their dependents, as well as self-employed members of the Social Security System (SSS), in the event of work-related sickness, injury, or death.⁵⁰

Under the present ECP, for an occupational disease and the resulting disability or death to be compensable, all of the following conditions must be satisfied:

1. The employee's work and/or the working conditions must involve risk/s that caused the development of the illness;
2. The disease was contracted as a result of the employee's exposure to the described risks;
3. The disease was contracted within a period of exposure and under such other factors necessary to contract it; and
4. There was no deliberate act on the part of the employee to disregard the safety measures or ignore established warning or precaution.⁵¹

However, only those injuries, disability, or death resulting from occupational and work-related diseases listed under Annex "A" of the Amended Rules on Employees' Compensation are considered compensable.⁵²

If a content moderator, assuming the risk or hazard of their work was recognized, were to develop a mental or psychological condition, they would be compensable for lost income due to inability to work, and as applicable, medical benefits, rehabilitation services, carer's allowance, and

⁵⁰ *Employees' Compensation Program*, SSS WEBSITE, at <https://www.sss.gov.ph/employees-compensation-program>.

⁵¹ Employees' Comp. Comm'n (ECC) Amended Rules on Employees' Comp. [hereinafter "ECC Amended Rules"] (1987), ann. A, *amended by* ECC Board Res. No. 11-05-13 (2011). *See* SSS v. Cuento, G.R. No. 225827, July 28, 2021.

⁵² *ECP Handout*, ECC WEBSITE, at https://ecc.gov.ph/wp-content/uploads/2016/08/The_ECP.pdf.

even death benefits to the beneficiaries.⁵³ But as earlier stated, the risk or hazard of their work is yet to be recognized.

There is also the matter of prevention. By recognizing an activity as an occupational risk or hazard, the law would not only compel the employer to take extra precautions in the work they require their workers to carry out, but also allow the government to hold the employer liable, should it fail to take these precautions.

Given that E.O. No. 307 did not include mental or psychological illnesses or stresses in its classification, a content moderator would only be entitled to the same extent of protection as any office worker.

Nonetheless, it must be considered that E.O. No. 307 is an old promulgation, and hence suffers from the constraints of its time. As initially discussed, content moderation is a relatively new line of work. It is thus appropriate to look at how E.O. No. 307 has been expanded and strengthened to meet the demands of modernity.

B. OSH for Call Center Workers, Under D.C. No. 1-08

The DOLE issued Department Circular No. 1 series of 2008 (“D.C. No. 001-08”) to provide for policy guidelines governing the occupational safety and health of workers in the call center industry.

As defined under D.C. No. 001-08, a call or contact center is “a central customer service operation where agents – or customer care specialists or customer service representatives – handle business-related telephone calls and other IT-related activities on behalf of a client.”⁵⁴

While the definition itself provides for a catch-all with the phrase “other IT-related activities on behalf of a client,” which content moderators could potentially fall under, it is admitted that the Circular is not explicitly applicable to content moderators. Nonetheless, there is an observable overlap between content moderation and the call center industry. Workers sufficiently skilled in both often shift from one side to another. Both industries also involve similar office-type setups as well as similar work

⁵³ ECC Amended Rules, r. VII, § 1. Types of Benefits under Employees Compensation.

⁵⁴ DOLE Dep. Circ. No. 001-08 (2008), § 2 (a). Policy Guidelines Governing the Occupational Safety and Health of Workers in the Call Center Industry.

arrangements. It would thus not be unreasonable for content moderators to require a similar set of protections, if not more, given the particular concerns of the workforce as earlier discussed. In any case, a quick review of the Circular is warranted if only to ascertain what kind of rights and remedies are available to call and contact center workers, which may be beneficial to content moderators as well.

The Circular requires the employer to formulate and implement a suitable OSH program, which “refers to planned activities aimed to prevent, eliminate, reduce, or control occupational risks and hazards.”⁵⁵ Each establishment needs to draft an OSH policy consistent with the OSH Standards and other related OSH issuances, one of which is the Technical Guidelines on OSH for the Call Center Industry (“Technical Guidelines”).

The Technical Guidelines provide for specific action points to be taken by the employer, as well as key questions for the employee, to effectively assess the risk of work-related injury, illness, and sickness. It highlights the hazards posed by night work, which can result in sleep disorders, accidents, and even substance and alcohol abuse, among others.⁵⁶

Apart from night work, it also provides key questions for the assessment of psychosocial stressors at work, especially given the stressful work environment that call centers are known for. Some of these questions pertain to the enforcement of work quotas, the quality of training given to the workers, and whether there are systems in place to help the workers deal with violent and irate clients. At the end, it also asks, “Do you have security of tenure in your job?”⁵⁷ This is a particularly important question, as it indirectly asks if a worker may feel a need to endure any psychosocial stressors in fear of sanction, or even termination.

Finally, it recommends medical surveillance, an important element of an OSH program, urging employers of call centers to be aware of the common health problems of its workforce and implement intervention measures thereafter.

D.C. No. 1-08 along with the Technical Guidelines were a necessary update to E.O No. 307, which ultimately inured to the benefit of BPO

⁵⁵ § 2 (b).

⁵⁶ DEP’T OF LAB. & EMP’Y-OFF. SAFETY & HEALTH CTR. (DOLE-OSHC), TECHNICAL GUIDELINES ON SAFETY AND HEALTH FOR CONTACT CENTERS (2007).

⁵⁷ *Id.*

workers. However, while its application extends to those who perform other IT-related activities on behalf of a client, it is not squarely applicable to all of the concerns of content moderators. The fact that their work itself may cause or contribute to the development of mental or psychological conditions is not at all addressed or contemplated. But where there may not have been an urgent need for it at the time the guidelines were issued, there is certainly a need for it now.

As earlier mentioned, the work of a call center agent and a content moderator differ significantly in many aspects. A call center agent will not always need to deal with irate clients, but content moderators always have to anticipate viewing something that is at best, unpleasant, and at worst, disturbing, consisting of the reported and flagged posts which would then be their task to “clean up.”

C. Strengthening OSH Standards through Republic Act No. 11058

The next pertinent update to E.O. No. 307 came in 2018 when Republic Act No. 11058 (R.A. No. 11058) was signed into law.

R.A. No. 11058, as a matter of policy, expands the coverage of occupational safety, again recognizing the role of the State to protect every worker against injury, sickness, or death.⁵⁸ The law, however, does not expressly provide its own definition of OSH standards. Rather, it delegates the task of definition to the Secretary of Labor and Employment (SOLE).⁵⁹ Hence, barring any new definition by a later SOLE, the definition under the IRR of E.O. No. 307 applies.

The Act identifies the covered workplaces as follows:

- (c) *Covered workplaces* refer to establishments, projects, sites and all other places where work is being undertaken wherein the number of employees, nature of operations, and risk or hazard involved in the business, as

⁵⁸ Rep. Act No. 11058 (2018), § 1. An Act Strengthening Compliance with Occupational Safety and Health Standards and Providing Penalties for Violations Thereof.

⁵⁹ See § 3 (j). Occupational Safety and Health (OSH) standards refer to the Occupational Safety and Health Standards issued by the Secretary of Labor and Employment [...].

determined by the Secretary of Labor and Employment, require compliance with the provisions of this Act; [...]⁶⁰

Other provisions in the Act contain rules that require further determination of another office or authority, whether by the SOLE or the DOLE itself.⁶¹ It is therefore clear that R.A. No. 11058, while updating the old law, is still not intended to be the “end-all, be-all” of OSH standards. It only establishes the mechanisms for the provision and enforcement of the same.

For content moderators in particular, the protections they seek to avail would depend on the recognition by the SOLE of the risks and hazards involved in their work. Any enforcement sought based on R.A. No. 11058 would likewise have to depend on that determination. As of now, there is no DOLE issuance pertaining to content moderators or BPO workers in general.

DOLE Department Order No. 198, series of 2018 or the IRR of R.A. No. 11058, defines a “worker” as “any member of the labor force, regardless of employment status.”⁶² This echoes the definition provided in the IRR of E.O. No. 307.

The IRR of R.A. No. 11058 also contains a more comprehensive list of workplaces covered by OSH standards. These workplaces are classified into three categories of establishments: high-risk, medium-risk, and low-risk.

High-risk establishments, as explicitly listed, only pertain to those “wherein the presence of hazard or risk may affect not only the workers, but also persons outside the premises of the workplace.”⁶³ This includes work such as construction, mining, and deep-sea fishing, among others. Office jobs are, reasonably, not contemplated within this category. However, there is no such enumeration in the other two categories, meaning it would have to be determined on a case-to-case basis. The establishment is also responsible for determining its own level of classification, based on the

⁶⁰ § 3 (c).

⁶¹ *See e.g.* §§ 6, 9, 14, 15, 16, 17, 22.

⁶² R.A. No. 11058 Rules & Regs. (2018), § 3 (bb). DOLE Dep. Order No. 198-18.

⁶³ § 3 (h).

Hazards Identification and Risk Assessment and Control (HIRAC) conducted by the employer.⁶⁴

D. Workplace Mental Health Protections Under the Mental Health Act

In any event, whether low or medium risk, the OSH program to be formulated and implemented by the covered workplace must include, among others, mental health services in the workplace, in accordance with Republic Act No. 11036 or the Mental Health Act.⁶⁵

Enacted in 2018, the Mental Health Act contains key provisions promoting mental health awareness in the workplace and is geared, among its many goals, toward outlawing discrimination against workers who are afflicted with mental illness.⁶⁶ The burden falls on both the employer and DOLE, and the CSC in applicable cases, to enact measures in this regard:

Section 25. *Mental Health Promotion and Policies in the Workplace.* — Employers shall develop appropriate policies and programs on mental health issues, correct the stigma and discrimination associated with mental conditions, identify and provide support for individuals with mental health conditions to treatment and psychosocial support.⁶⁷

* * *

Section 35. *Duties and Responsibilities of the Department of Labor and Employment (DOLE) and the Civil Service Commission (CSC).* — The DOLE and CSC shall:

- (a) Develop guidelines and standards on appropriate and evidence-based mental health programs for the workplace as described in this Act; and
- (b) Develop policies that promote mental health in the workplace and address stigma and discrimination suffered by people with mental health conditions.⁶⁸

⁶⁴ DOLE Adv. No. 4 (2019). Guide For Compliance of Establishments to D.O. No. 198-18.

⁶⁵ See Rep. Act No. 11036 (2018), § 25. Mental Health Act.

⁶⁶ See §§ 2–3.

⁶⁷ § 25.

⁶⁸ § 35.

The Mental Health Act also provides recourse against discrimination, which is defined as such:

- (e) *Discrimination* refers to any distinction, exclusion or restriction which has the purpose or effect of nullifying the recognition, enjoyment or exercise, on an equal basis with others, of all human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field. It includes all forms of discrimination, including denial of reasonable accommodation. Special measure solely to protect the rights or secure the advancement of persons with decision-making impairment capacity shall not be deemed to be discriminatory;⁶⁹

Discrimination, as defined, is penalized under Section 44 of the Mental Health Act. A person who commits discrimination against a person with a mental health condition and is convicted, shall be punished by imprisonment of not less than six months but not more than two years, or a fine of not less than PHP 10,000 but not more than PHP 200,000, or both, at the discretion of the court. If discrimination is committed by a juridical person, the penalty is imposed on the directors, officers, employees, or other persons responsible for the offense.⁷⁰

Meanwhile, the Mental Health Act IRR expands Section 25 by requiring the employer to take a more proactive role in promoting the mental health of its employees by providing support not only to those with already existing conditions, but also to those at risk.⁷¹ It also expands Section 36 by tasking the DOLE to “issue appropriate guidelines in the development and implementation of policy and programs to promote mental health in the workplace in coordination with the DOH and consultation with mental health professionals and stakeholders” and assist the employers in “the development and promotion of mental health programs in the workplace, including access to appropriate mental health services.”⁷²

⁶⁹ § 4 (e).

⁷⁰ § 44.

⁷¹ Rep. Act No. 11036 Rules & Regs. [hereinafter “Mental Health Act IRR”] (2019), § 26.

⁷² § 36.

While the Act and its IRR provide express mandates on both the employer and the DOLE to provide mental health support to a worker, the particular courses of action, as well as any minimum conditions or requirements, will ultimately come from the guidelines issued by the DOLE. The Mental Health Act itself does not provide direct enforcement of its workplace mandate to enable a worker to readily benefit from it. Only discrimination, in addition to the other specific acts provided under Section 44, provides a direct cause of action, given that it is penalized under the Act. Hence, a review of the said guidelines is necessary to determine what employers are required by law to do in order to promote mental health in the workplace.

DOLE Department Order No. 208, series of 2020 (“D.O. No. 208-20”) contains the Guidelines for the Implementation of Mental Health Workplace Policies and Programs for the Private Sector. It directs all workplaces and establishments to formulate a Mental Health Workplace Policy and Program.⁷³ This shall be jointly prepared by management and workers' representatives and be made an integral part of the company's OSH policies and programs.

D.O. No. 208-20 provides for several responsibilities on the part of the employer, some of which may find relevance in the issue of content moderators. Pertinent responsibilities include, but are not limited to, the following:

- (1) To develop, implement, monitor and evaluate mental health workplace policies and programs;
- (2) To ensure that there are adequate resources to implement and sustain mental health workplace programs;
- (3) To ensure that they provide the necessary training to the OSH personnel and Human Resource Officers who will develop, implement and monitor the mental health workplace policies and programs;
- (4) To provide the necessary work accommodation when needed; and
- (5) To develop mechanisms for referral of workers at risk of developing or with a mental health condition for appropriate management.⁷⁴

⁷³ DOLE Dep’t Order No. 208 (2020), § III.A. Guidelines for the Implementation of Mental Health Workplace Policies and Programs for the Private Sector.

⁷⁴ § V.A.

The implementation and monitoring of the responsibilities listed shall also be the responsibility of the employer, through the OSH committee. A report shall subsequently be submitted by the committee to the DOLE Regional Office with jurisdiction over the company.⁷⁵ In effect, the implementation of Mental Health Act is on a self-reporting basis.

D.O. No. 208-20 calls on employers to provide mental health benefits. However, the provision of benefits to a worker with a mental health condition would be subject to the current health benefit packages under PhilHealth, ECC, or SSS, as applicable. Workers with conditions would also be entitled to all monetary and non-monetary statutory benefits, but without prejudice to the exercise of company policy, rules and regulations on compensation and other benefits.⁷⁶ For companies using private healthcare providers, the inclusion of mental health services is merely encouraged.⁷⁷ Employers are still given much leeway in determining what kind of health benefits and services they deem appropriate to provide their workers.

Similarly, the DOLE Guidelines require employers to provide a mechanism for access to counseling, but only through a referral system,⁷⁸ rather than a service made available to the worker whenever needed. Nor is it required to be provided by the employer free of charge.

The lack of inclusions surely poses problems in a workplace where it is more likely, compared to an ordinary office workplace, that the work itself is contributory to the condition or even the development thereof, thus requiring immediate support for the affected employee.

As for discrimination, the Guidelines claim to protect a worker from termination on the basis of actual, perceived, or suspected mental health conditions. This is, according to the provisions on non-discriminatory policies and practices, “unless the condition progresses to such severity that it affects the worker’s safety or the safety of co-workers, and work performance and productivity, upon a certification issued by a competent public health authority with expertise on mental health.”⁷⁹

⁷⁵ § VII.

⁷⁶ § IV.E.2.

⁷⁷ § IV.E.3.

⁷⁸ § IV.F.1.

⁷⁹ § IV.C.1.e.

The Guidelines also provide that the mental health condition of a worker shall not interfere with their performance as an employee:

C. Social Policy

1. Non-discriminatory policies and practices

- a. There shall be no discrimination in any form against workers who are at risk of developing or who are found to have mental health condition. Workers shall not be discriminated against from hiring, promotion, and/or other benefits of employment because of their condition *provided, however, that such conditions shall not interfere with the employee's performance of their job* or unduly affect his own safety or that of his co-workers, clients and the general public.

* * *

- e. A worker shall not be terminated from work on the basis of actual, perceived or suspected mental health condition *unless the condition progresses to such severity that it affects his/her own safety or safety of co-workers and work performance and productivity* upon the certification issued by a competent public health authority with expertise on mental health.⁸⁰

Herein lies the problem. As earlier discussed, BPO companies can and often do dismiss their employees if they are unable to hit a certain quota or metric, which is typically the measure for productivity. This is standard practice in the industry. Thus, content moderators who begin to fall behind as a direct effect of the psychological distress brought on by their very work, may still face termination.

The law, gleaned from the above Guidelines, does not consider this to be discriminatory, even if the reason for the lack of productivity pertains to a mental health condition. On the contrary, exceptions and conditions are even provided specifically so that it cannot be considered a discriminatory practice.

D.O. No. 208-20 also provides that companies and workers shall agree on work arrangements that accommodate a worker's mental health

⁸⁰ § IV.C. 1.a., 1.e. (Emphasis supplied.)

condition, but this is likewise premised on the condition that the worker's performance shall not suffer.

Granted, the Guidelines state that the Mental Health Workplace Policy and Program of the company shall promote workers' well-being through "identification and management of work-related stress and stressors" and "effective management of changes in the work organization and the utilization of human resources systems (e.g., addressing burnout, review of workload)."⁸¹ This signals a level of responsibility on the part of the employer to alleviate the load of content moderators as needed. However, the cited examples above are included only as a recommendation and not legally mandated. Likely, it is considered more as an ethical responsibility rather than a legal one.

E. National Policy Framework on Promotion of Healthy Workplace

Under its mandate, the DOLE also issued the DOH-DOLE-CSC Joint Administrative Order No. 2023-0001 ("J.A.O. No. 23-01"), which establishes the National Policy Framework on the Promotion of Healthy Workplace, requiring various agencies and local government units ("LGUs") to promote a healthy workplace, through implementations particular to each office.⁸²

In this regard, a Healthy Workplace Framework is developed to ensure that health promotion measures are instituted in the workplace to promote and protect the workforce's health, safety, and well-being.

Under its specific guidelines, mental health, as well as occupational injuries, are both included as part of its priority areas:

1. Priority Areas for Health Promotion. All health promotion interventions in the workplace shall focus on the following priority areas or risk factors:

* * *

⁸¹ § IV.B.

⁸² DOH-DOLE-CSC Joint Admin. Order No. 1 (2023), § II.

- e. *Mental Health — Psychosocial and mental well-being among workers shall be increased and protected to reduce the burden of mental health disorders and incidence of suicide in the workplace;*

* * *

- g. *Violence and Injury Prevention — Safe and inclusive workplace environments shall be fostered to eliminate the various forms of violence and injuries, including interpersonal violence or gender-based violence, and occupational-related injuries.*⁸³

J.A.O. No. 23-01 mandates the convening of a National Technical Working Group on Healthy Workplace (“NTWG-HW”) to oversee the overall planning, implementation and enforcement, and monitoring and evaluation of the Healthy Workplace Framework.⁸⁴ The member-government agencies are tasked to coordinate with employers in implementing workplace policies that promote OSH, as well as to extend assistance wherever needed. However, the employers are merely encouraged to follow through with implementation as there appears to be no strict compliance mechanism, deviating only slightly from previous issuances and laws, nearly all of which rely on a self-reporting mechanism. In this instance, the burden to report primarily falls on NTWG-HW.

J.A.O. No. 23-01 provides for a Monitoring and Evaluation Plan as well, the criteria for which are likewise formulated by the NTWG-HW, though the Joint Administrative Order does not provide for what NTWG-HW or any of the departments involved are allowed to do in case of non-compliance. Neither does the source legislation, R.A. No. 11223, provide any recourse against the employer, should it fail to fully comply with any requirements or standards given by the NTWG-HW. Instead, the Joint Administrative Order adopts a strategy of positive reinforcement, giving NTWG-HW only the duty to establish a mechanism to recognize and award workplaces for compliance.⁸⁵ It also includes general provisions for monitoring and evaluation,⁸⁶ but does not explicitly provide for action steps that would compel, rather than merely encourage, companies in content moderation to protect their workers from the mental health hazards involved in the work.

⁸³ § VI.A.1. (Emphasis supplied.)

⁸⁴ § VI.B.1.a.

⁸⁵ § VI.B.5.

⁸⁶ § VI.B.4.

F. Content Moderation as a Novel Issue in Philippine Labor Law

In sum, due to the constantly evolving nature of labor, there are some lines of work which Philippine labor laws do not yet seem to know how exactly to protect.

E.O. No. 307, barring the fact that BPO-IT was not yet an industry at the time, did not recognize mental health as part of its OSH standards. Notwithstanding this fact, it established the baseline that OSH shall include prevention of workers' departure from health caused by their working conditions. It would hence be incumbent upon succeeding legislation to determine how this should apply to different industries, taking into consideration the evolving landscape. Unfortunately, in the case of content moderators, the law has yet to provide adequate protections.

R.A. No. 11058, the law strengthening OSH, does not contemplate mental distress or illness as injury, hazard, or sickness in its provisions. It only contemplates physical work, and in particular, work that is involved in industries requiring manual labor. The corresponding IRR, as discussed, also only requires the promotion of mental health to the extent of the mandate in R.A. No. 11036, or the Mental Health Act.

As for R.A. No. 11036, while taking large strides insofar as the mental health of a worker is recognized as something that must be protected, the law does not contain any procedure for when it is the work itself that puts a worker at risk of developing any mental illness. It also does not have a strong compliance mechanism and it relies on self-reporting on the part of the employer.

Finally, in J.A.O. No. 23-01, mental health is considered a priority area. However, it also lacks a strong compliance mechanism.

Why are protections important? What is the significance of an explicit provision protecting the mental health of a content moderator? This again relates to the issue of performance and productivity. Labor groups such as the BPO Industry Employees' Network ("BIEN") stress that performance-based assessment creates unrealistic metrics, which each

employee must meet to be a regular employee with improved benefits. The higher the standards, the less likely regularization will take place.⁸⁷

What could this mean for a content moderator? Given the nature of their job, this would mean higher numbers or more content reviewed, all without compromising accuracy. It would mean a heightened risk of exposure to more violent and graphic content. Moderators are expected to be able to withstand it to keep their jobs.

A study conducted by the ILO in 2016 also found that Filipino BPO workers reported high-stress work environments with detrimental impacts on health.⁸⁸ An older study in 2010 also described the average BPO work environment as “high-strain,” due to excessive and tedious workloads, tight rules and procedures, and electronic monitoring, where workers are granted little autonomy. Regular night work was also cited as one of the stress-inducing factors.⁸⁹

On this note, night work, which is extremely common in the BPO industry, poses another risk factor to the health of workers. The 2010 ILO study provides that about half of BPO workers suffer from sleep problems and insomnia as an effect of night work.⁹⁰ This certainly cannot help, and can do nothing but aggravate, the mental distress that content moderators are already prone to, due to the nature of their work.

Hence, it is crucial to have protections in the form of strictly mandated and implemented mental health policies and strategies in the workplace. To reiterate, and it cannot be stressed enough: the right to safe and humane work conditions is a cardinal labor right. Without the necessary protections, Filipino BPO workers, and specifically content moderators, are essentially playing a losing game. They are consistently made to expose themselves to psychologically distressing content, and to do so beyond their capacity, all to meet impossible metrics set by their clients. They are given nearly no mental health support. They may not even be entitled to any health benefits depending on their employment status. Then, if and when they

⁸⁷ *Pass the Pro-worker SOT Bill*, BPO INDUS. EMP. NETWORK, July 20, 2022, at <https://bienphilippines.wordpress.com/2022/07/20/pass-the-pro-worker-sot-bill>.

⁸⁸ Errighi et al., *supra* note 14.

⁸⁹ Tonyo Cruz, *BPO industry should improve work conditions, says ILO study*, ASIAN CORR, July 21, 2010, archived at <https://web.archive.org/web/20100725202901/http://us.asiancorrespondent.com/tonyo-cruz-blog/bpo-industry-should-improve-work-conditions-says-ilo-study>.

⁹⁰ *Id.*

suffer any mental health condition, whether on account of or exacerbated by the workload, they are afforded very few accommodations. Worse, they even face the possibility of termination, particularly because the law does not consider it a discriminatory practice to terminate employment after a worker fails to perform adequately due to a mental health condition.

The tragedy of such a setup cannot be ignored. This is especially within the context of a growing BPO workforce, wherein more and more workers will be left with no adequate protections. It runs contrary to the State policy on full protection of labor, and in general, the principles of equity and justice.

G. Protection of Content Moderators' OSH in Criminal Law

As content moderators are constantly exposed to graphic content, it raises the questions of what liabilities content moderators may incur due to their access of graphic content as part of their work, and what protections they may find in criminal law with respect to protecting their well-being.

Under Republic Act No. 11930 or the Anti-Online Sexual Abuse or Exploitation of Children Act and Anti-Child Sexual Abuse or Exploitation Materials Act (“Anti-OSAEC and CSAEM Act”),⁹¹ child sexual abuse or exploitation material or child sexual abuse material (“CSAEM/CSAM”) refers to “any representation, whether offline, or by, through or with the use of ICT, by means of visual, video, audio, written, or any combination thereof, [...] of a child engaged or involved in real or simulated sexual activities.”⁹²

These forms of child sexual abuse fall squarely under the type of graphic content consumed every day by content moderators as part of their profession. As such, it may be intuitive to assume that the law would have the potential to inadvertently penalize content moderators through their consumption of⁹³ and derivation of financial benefit from⁹⁴ the consumption of OSAEC content, despite said consumption being for the purposes of eradicating OSAEC content.

Interestingly, the Anti-OSAEC and CSAEM Act already takes this concern into account through its “Good Samaritan” provision:

⁹¹ Rep. Act No. 11930 [hereinafter “Anti-OSAEC and CSAEM Act”] (2022).

⁹² § 3(d)

⁹³ § 4(d), 4(g), 4(r), & 4(s).

⁹⁴ § 4(k).

Any person who has the responsibility of reporting cases under this Act, blocking an internet address, removing a website or domain, taking down of shared videos, pictures, or messages for the services provided by an internet intermediary, and providing information for the purpose of an investigation or prosecution of a case involving acts of OSAEC shall not be held civilly, criminally or administratively liable: *Provided*, That the action was:

- (1) done in good faith;
- (2) necessary to prevent access or dissemination of CSAEMs; and
- (3) reported within twenty-four (24) hours from the act of blocking an internet address, removing a website or domain, or taking down of shared video, picture or messages.⁹⁵

This provision, though it does not label content moderators by that particular name, clearly contemplates a broad exemption for content moderators from liability under the Anti-OSAEC and CSAEM Act. Consequently, the Anti-OSAEC and CSAEM Act seems to be the first and only law to date that deliberately and specifically recognizes the existence of content moderators since the repealed Anti-Child Pornography Act⁹⁶ makes no distinction between the consumption of child pornography of content moderators from that of actual perpetrators, unlike the Anti-OSAEC and CSAEM Act.

While child sexual abuse is certainly a common form of disturbing content which content moderators are exposed to regularly, it is not the only form. As shown in *The Cleaners*, content moderators also often have to watch or screen violent content such as beheading videos, torture videos, self-harm videos, death on video, and other forms of graphic content.

While content moderators have to sift through thousands of content per day, research has shown that even mild exposure to this kind of content correlates significantly with post-traumatic stress symptoms.⁹⁷ An anecdote in *The Cleaners* typify this—“[T]he job damages your brain - making you think that violence is normal. Making you think that killing others, bombing is

⁹⁵ § 7.

⁹⁶ Rep. Act No. 9775 [hereinafter “Anti-Child Pornography Act”] (2009).

⁹⁷ E. Alison Holman, Dana Rose Garfin, & Roxane Cohen Silver. *It matters what you see: Graphic media images of war and terror may amplify distress*. 121 PROCEEDINGS OF THE NAT’L ACAD. SCI. U.S.A. 29 (2024)

normal.” Another anecdote in the film even described a situation where one content moderator assigned to specialize in self-harm videos ended up taking their own life.⁹⁸

Interestingly, there are no Philippine laws that explicitly prohibit the uploading of graphic content. Laws like the Cybercrime Prevention Act of 2012,⁹⁹ and the Anti-Photo and Video Voyeurism Act of 2009¹⁰⁰ would seem to be the closest analogues, but both laws only expressly prohibit sexual content such as cybersex¹⁰¹ and electronic coverage of sexual acts without the participants’ consent.¹⁰²

Evidently, this proves to be another gap in the law when it comes to protecting the OSH of content moderators. While the law at least contemplates the concepts of victims and perpetrators when it comes to OSAEC and CSAEM, the unregulated proliferation of violent content is essentially deemed victimless and, at worst, perfectly permissible. When this happens, it once again comes at the expense of the content moderator as invisible victims, with no remedies to claim damages or seek justice against those who share violent content in the first place.

V. COMPARATIVE ANALYSIS: OSH IN DIFFERENT CONTEXTS

A. India

India, as the top provider of BPO workers worldwide, tells a similar story. Though India offers more of its labor in IT, it nonetheless finds a large amount of its BPO workers in content moderation, accounting for over 10% of the global workforce.¹⁰³

⁹⁸ THE CLEANERS, *supra* note 6.

⁹⁹ Rep. Act No. 10175 [hereinafter “Cybercrime Prevention Act”] (2012).

¹⁰⁰ Rep. Act No. 9995 [hereinafter “The Anti-Photo and Video Voyeurism Act”] (2009).

¹⁰¹ Cybercrime Prevention Act, § 4(c)(1).

¹⁰² Anti-Photo and Video Voyeurism Act, § 4.

¹⁰³ Sana Ahmad & Martin Krzywdzinski, *Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains*, in DIGITAL WORK IN THE PLANETARY MARKET 77, 80 (Mark Graham & Fabian Ferrari ed., 2022).

Like the Philippines, India has a constitutionally-enshrined guarantee of just and humane conditions of work.¹⁰⁴ However, this is where the similarities begin to fade.

The first major deviation is that India is not a signatory to the Promotional Framework for Occupational Safety and Health Convention of the ILO. The only Conventions ratified by India that pertain to workplace safety include the Workmen's Compensation (Occupational Diseases) Convention. This was ratified in 1927, and the revised version of which ratified in 1964. Both versions of the said Convention contemplate only physical occupational hazards and diseases.¹⁰⁵ The other pertinent Convention was ratified in 2008, namely, the Prevention of Major Industrial Accidents Convention. But likewise, this only contemplates physical hazards and accidents.¹⁰⁶ Evidently, content moderators in India will find no recourse here.

Neither do the domestic laws of India provide any remedies for content moderators. Presently, India has four major labor laws: The Code on Wages,¹⁰⁷ The Code on Social Security,¹⁰⁸ The Industrial Relations Code,¹⁰⁹ and The Occupational Safety, Health and Working Condition Code,¹¹⁰ the last of which is comparable to E.O. No. 307 and R.A. No. 11058 under Philippine law.

The Occupational Safety, Health and Working Conditions Code of 2020 provides for labor protections against hazards, with “hazardous” defined as “involving danger or potential danger.”¹¹¹ While it does not explicitly refer to physical hazards, its surrounding provisions pertain to physical situations, which may mean that only physical hazards are covered by the Code.

This is affirmed further by the fact that no reference is made to mental or psychological injury, hazard, or health in the entirety of the law, whether express or implied. Furthermore, most of the provisions in the Code

¹⁰⁴ INDIA CONST. art. 4, § 42.

¹⁰⁵ ILO Workmen's Compensation (Occupational Diseases) Convention (Revised), Convention No. 121 (1952).

¹⁰⁶ Prevention of Major Industrial Accidents Convention, ILO (1993).

¹⁰⁷ Act No. 29 of 2019 (2019) (India). The Code on Wages.

¹⁰⁸ Act No. 36 of 2020 (2020) (India). The Code on Social Security.

¹⁰⁹ Act No. 32 of 2020 (2020) (India). The Industrial Relations Code.

¹¹⁰ Act No. 37 of 2020 (2020) (India). The Occupational Safety, Health and Working Conditions Code.

¹¹¹ Ch. 1, § 2, cl. (z).

contemplate physical work or manual labor, with a few exceptions including working journalists and workers in audio-visual production.

The Code mandates employers to “ensure that [the] workplace is free from hazards which cause or are likely to cause injury or occupational disease to the employees,”¹¹² and “provide and maintain, as far as is reasonably practicable, a working environment that is safe and without risk to the health of the employees.”¹¹³ While again, the provisions do not expressly exclude non-physical injury or risk, the surrounding provisions are instructive in ascertaining the coverage of the cited OSH standards.

The Code also gives the Central Government the power to provide OSH standards for workplaces which it considers necessary on the report of the authority designated by such Government for such purpose.¹¹⁴ Thus, despite the implied exclusion of the occupational hazards of content moderation, the law offers some leeway for a declaration of its inclusion by the Central Government as such, akin to the determinations of DOLE and SOLE in the case of the Philippines.

Both Philippine and Indian laws face the issue of classification. To recall, neither are office jobs considered a “high risk” workplace, nor is content moderation legally considered a hazardous work process under the protection of E.O. No. 307 and R.A. No. 11058, our domestic OSH laws.

Under Philippine OSH laws, compliance with the R.A. No. 11036 is included in the implementation of its mandate for workplace safety. India also has a corresponding mental health law, namely, The Mental Healthcare Act of 2017.¹¹⁵ This law identified the State’s role in ensuring that mental healthcare is available and affordable, as well as reducing the stigma attached to mental illness.¹¹⁶ Most significant in the law is its decriminalization of suicide, which was formerly penalized under their laws. Instead, persons who attempt suicide will now have government-provided access to mental health-related care, treatment and rehabilitation.¹¹⁷

¹¹² Ch. 3, § 6, cl. 1(a).

¹¹³ Ch. 3, § 6, cl. 1(d).

¹¹⁴ Ch. 4, §18, cl. 2(b).

¹¹⁵ Act No. 10 of 2017 (2017) (India). The Mental Healthcare Act.

¹¹⁶ Ch. 6, § 30, cl. (b).

¹¹⁷ Ch. 16, §115, cl. 1.

Mental illness is still widely stigmatized in India, making the Act a very important and progressive piece of legislation in overall public health. However, it does not address the very specific problems of mental health in the workplace. Unlike its Philippine counterpart, the Act made no reference to mental health and safety in the workplace.

Public health research conducted in 2022 revealed that few mental health interventions in India were accompanied by comprehensive needs assessment, impact evaluation, or workplace policy initiatives. Instead, interventions focused on the curative aspect, such as counseling services. This effectively limited the scope of mental health promotion activities and strategies in the workplace.¹¹⁸ This is problematic due to the varying environments across industries, requiring the assessment and employment of different mental health promotion strategies, rather than a “one-size-fits-all” approach. For another, it is not responsive to the specific issue that, given the nature of content moderation work, it is the work itself that can, and is more likely to, pose a risk to workers’ mental health and safety.

The 2016 ILO study identified key labor issues in the BPO industry of India. Among them are high attrition rates, limited women’s participation, and limited workers’ representation in general.¹¹⁹

The high attrition rate fluctuated between 30 and 40% and was attributed to various factors, including the strong result orientation, continuous monitoring, night-shift work characteristic of BPO jobs, work-related stress, and work-life imbalances.”¹²⁰ In particular, Indian content moderators face the same problem of work-caused psychological distress, aggravated by the subpar working conditions and highly unrealistic quotas provided by management. PTSD and clinical depression were also commonly reported among workers, many of whom are young people.¹²¹

Another study by Ahmad and Krzywdzinski found that Indian content moderators’ dissatisfaction with the work was due to bad management practices, long working hours, and a lack of growth

¹¹⁸ Apurvakumar Pandya et al., *Workplace Mental Health Interventions in India: A Rapid Systematic Scoping Review*, 10 FRONT. PUBLIC HEALTH (2022), at <https://www.frontiersin.org/articles/10.3389/fpubh.2022.800880/full>.

¹¹⁹ Errighi et al., *supra* note 14.

¹²⁰ *Id.*

¹²¹ Prasad Banerjee, *Inside the secretive world of India’s social media content moderators*, MINT, Mar. 18, 2020, at <https://www.livemint.com/news/india/inside-the-world-of-india-s-content-mods-11584543074609.html>.

opportunities.¹²² They are often not fully informed about the nature of their tasks upon application and are later discouraged by management from talking about their work, citing “trade secrets” and a need to protect the client’s brand.¹²³ As Ahmad and Krzywdzinski noted, they are then made to “work with very stressful content, in a strictly standardized workflow, and under enormous time pressure.”¹²⁴

All of these echo the issues faced by Filipino BPO workers and content moderators. With no protections available under the current OSH law, nor sufficient mental healthcare coverage, Indian content moderators are exposed to risk by their very work, with no accessible legal recourse should they fall ill from it. Thus, Indian content moderators may find themselves in a similar predicament to their Filipino counterparts.

B. Australia

The story could not be more different, however, in a place not far down south. As far as the Asia-Pacific region is concerned, Australia is often a good point of comparison when it comes to issues involving labor. According to the 2024 Labour Rights Index of the Centre for Labour Research, Australia ranked first in the region (i.e. Southeast Asia and Oceania) in providing decent work for employees; taking into account the basic indicators of fair wages, decent working hours, employment security, family responsibilities, maternity at work, safe work, social security, fair treatment, child and forced labor, and freedom of association.¹²⁵ It should come to no surprise then that when it comes to OSH policies, Australia has developed significant workplace health and safety legislative and regulatory action addressing risks around mental health conditions in the workplace.¹²⁶

The national OSH policy of Australia is much owed to the formation of Safe Work Australia (“SWA”), an independent government agency created by law to develop national policy relating to work health and safety and workers’ compensation. Since Australia has a federal system of government, the SWA itself does not have the jurisdiction to regulate or

¹²² Ahmad & Krzywdzinski, *supra* note 103, at 88.

¹²³ *Id.* at 89.

¹²⁴ *Id.* at 91.

¹²⁵ CTR. FOR LAB. RESEARCH, LABOUR RIGHTS INDEX 2024 34–41 (2024).

¹²⁶ Crowley Woodford et al., *Mental health in the workplace: is it a safety issue?*, PEOPLE MGMT., Feb. 8, 2024, at <https://www.peoplemanagement.co.uk/article/1860731/mental-health-workplace-safety-issue>.

enforce OSH¹²⁷ laws or administer workers' compensation arrangements.¹²⁸ But the Australian Commonwealth, state and territory governments agreed under the Intergovernmental Agreement for Regulatory and Operational Reform in Occupational Health and Safety to harmonize their OSH laws and to establish appropriate governance arrangements to support this cooperation.¹²⁹ Hence, the advisories provided by SWA as to how the codified OSH laws should be implemented are at least persuasive, and at most, binding.

The SWA is responsible for the development of a uniform set of laws that works to harmonize the varying OSH laws across jurisdictions. The Model Work Health and Safety Act, amended in 2023, is the primary legislation on OSH. According to this law, health equates to both “physical and psychological health.”¹³⁰ The scope of its regulation also includes “the prescribing of standards relating to the use of or exposure to any physical, biological, chemical or psychological hazard.”¹³¹ At the outset, mental or psychological health is contemplated as an answer to the question of workplace safety.

The Model Work Health and Safety Regulation expands the definitions, with a section to define specific psychosocial hazard and risk.

Psychosocial hazard is defined therein as “a hazard that (a) arises from, or relates to (i) the design or management of work; or (ii) a work environment; or (iii) plant at a workplace; or (iv) workplace interactions or behaviours; and (b) may cause psychological harm, whether or not the hazard may also cause physical harm.”¹³² Meanwhile, psychosocial risk is “a risk to the health or safety of a worker or other person from a psychosocial hazard.”¹³³

The law thus provides for the duty of a person conducting a business or undertaking, or the employer, to manage such psychosocial risks. In

¹²⁷ Australian laws refer to OSH as “Work Health and Safety” or “WHS.” However, it will be referred to as OSH in this paper for purposes of consistency.

¹²⁸ *Who we are and what we do*, SAFE WORK AUSTRALIA, at <https://www.safeworkaustralia.gov.au/about-us/who-we-are-and-what-we-do>.

¹²⁹ Inter-Governmental Agreement for Regulatory and Operational Reform in Occupational Health and Safety (2008) (Cth) (Austl.).

¹³⁰ Model Work Health and Safety Act (2023), pt. 4, div. 3, sub-div. 1 (Cth) (Austl.).

¹³¹ Sch. 3, item 5.

¹³² Work Health and Safety Regulation (2011), ch. 3, pt. 3.2, div. 11 sub-div. 1, item 55A (Cth) (Austl.).

¹³³ Item 55B.

implementation, the employer must take into account certain factors, some of which include the duration, frequency or severity of exposure to psychosocial hazard, how the psychosocial hazards may interact or combine, and the design of work, including job demands and tasks.¹³⁴

SWA considers, for instance, traumatic events or materials a psychosocial hazard. It mandates employers to eliminate psychosocial risks, or where not reasonably practicable, minimize them so far as is reasonably practicable.¹³⁵

There are numerous ways through which this could be implemented. In fact, Australia already has prevention and treatment guidelines in place when it comes to professions which are heavily at risk insofar as psychosocial hazards and the potential of resulting disorders such as PTSD, Complex PTSD, and Acute Stress Disorder (ASD). The said guideline recommendations are those of Phoenix Australia,¹³⁶ a nonprofit organization which calls itself “Australia’s National Centre of Excellence in Posttraumatic Mental Health.” The guidelines were approved by Australia’s National Health and Medical Research Council last 2021,¹³⁷ in accordance with the Council’s power under Australia’s National and Medical Research Council Act of 1992.¹³⁸

An interesting aspect of the guidelines laid out by Phoenix is that the guidelines are segmented based on different types of trauma, including special populations of professions that are deemed at risk for psychosocial hazards. Chapter 9, for instance, has two different profession-specific guidelines: one for emergency services personnel, and one for military and ex-military personnel.

While content moderators specifically aren’t contemplated in Phoenix’s guidelines, the specific tailoring of OSH guidelines to different professions, and the identification of at-risk professions at all are good

¹³⁴ SAFE WORK AUSTRALIA, MODEL CODE OF PRACTICE FOR MANAGING PSYCHOSOCIAL HAZARDS AT WORK 6–7 (2022), at <https://www.safeworkaustralia.gov.au/doc/model-code-practice-managing-psychosocial-hazards-work>.

¹³⁵ *Id.* at 6.

¹³⁶ PHOENIX AUSTRALIA, AUSTRALIAN PTSD GUIDELINES, at <https://www.phoenixaustralia.org/australian-guidelines-for-ptsd/>.

¹³⁷ *Guidelines for Guidelines*, AUSTL. NAT’L HEALTH & MED. RES. COUNCIL, at <https://www.nhmrc.gov.au/guidelinesforguidelines/nhmrc-approved-guidelines>.

¹³⁸ Nat’l Health and Med. Research Council Act (1992), § 14(A) (Cth.) (Austl.).

foundational baselines to be built upon. As to what guidelines could be tailored specifically to content moderators, some proposals could include allowing the removal of users after a single serious breach instead of making the moderator review all their content and risk further exposure. It could also mean increasing breaks, recovery time and support if employees are exposed to traumatic events or materials or creating a safe space for workers to report traumatic or distressing events.¹³⁹ Other ideas could be the reduction of quotas, or softer margins of error for or failsafe protocols in the event of failures in moderation.

In addition, content moderator-specific guidelines could take inspiration from the existing trauma guidelines Phoenix developed for military personnel and emergency services personnel. Insofar as emergency personnel, one interesting suggestion for treatment offered by Phoenix that may be of use to content moderators is that during the treatment stage, the trauma patient is “not given a sick leave, but rather, kept in a meaningful role in the organization but without exposure to traumatic events.”¹⁴⁰ This treatment suggestion could apply to the case of content moderators who, like emergency personnel, are thrust into continuous work with little reprieve. Meanwhile, the guidelines for military and ex-military personnel also have fascinating perspectives on the root of trauma for those individuals, as many of them undergo a loss of identity due to their participation in acts of violence. This guideline could be useful when dealing with content moderators who feel complicit in the graphic nature of the content being moderated.

Ultimately, however, since every worker is different, there is no one standard on how OSH should be promoted in the workplace. Hence, the importance of consultation—the Model Code of Practice repeatedly stresses that employers should regularly consult their employees as to what policies directly benefit them and are most responsive to the concerns of their work.¹⁴¹ After all, they may have ideas on improving work design and minimizing the risks of psychological harm.

SWA is also tasked by the Intergovernmental Agreement to develop compliance measures. According to SWA, OSH regulators promote and encourage OSH compliance through a range of methods, which may vary

¹³⁹ SAFE WORK AUSTRALIA, *supra* note 134, at 45–46.

¹⁴⁰ PHOENIX AUSTRALIA, *Specific Populations and Trauma Types—Emergency services personnel 4*, in AUSTRALIAN PTSD GUIDELINES.

¹⁴¹ SAFE WORK AUSTRALIA, *supra* note 134, at 8–11.

per jurisdiction. In general, there are regulators who are tasked to inspect workplaces, advise on and enforce laws, and who may also issue sanctions, including giving infringement notices and commencing prosecutions, as applicable.¹⁴²

There are many ways through which an employer may be held liable for non-compliance with OSH standards. The regulator, an inspector authorized by the regulator, the Department of Public Prosecution, or if no prosecution is brought after six (6) months but not later than 12 months after the incident, the regulator at the request of an aggrieved employee may file commence criminal proceedings against employers for breaching OSH offensive provisions.¹⁴³ Under the Model Work Health and Safety Act, offenses are categorized into three (3) categories:

- (1) Category 1, in which a duty holder, without reasonable excuse, engages in conduct that recklessly exposes a person to a risk of death or serious injury or illness. This is the most serious category.
- (2) Category 2, in which a duty holder fails to comply with a health and safety duty that exposes a person to risk of death or serious injury or illness.
- (3) Category 3, in which a duty holder fails to comply with a health and safety duty.¹⁴⁴

The penalties for the corresponding categories also vary, but some involve imprisonment if found guilty, and a range of AUD 78,000 to 11,561,000 worth of fines¹⁴⁵ (around PHP 3,000,000 to 446,000,000),¹⁴⁶ depending on the offense committed and whether an individual or a body corporate committed it. Suffice to say, in any case, the penalties for even the smallest infractions may be consequential enough for a company to exert their best efforts to comply with the OSH standards imposed.

¹⁴² *How regulators enforce WHS laws?*, SAFE WORK AUSTRALIA, at <https://www.safeworkaustralia.gov.au/how-regulators-enforce-whs-laws>.

¹⁴³ *Id.*

¹⁴⁴ *Id.*

¹⁴⁵ *Penalties under the WHS laws*, SAFE WORK AUSTRALIA, at <https://www.safeworkaustralia.gov.au/law-and-regulation/legislation/penalties-under-whs-laws>.

¹⁴⁶ Based on conversion rates dated October 7, 2024 (1 AUD = 38.2974 PHP) provided by the Bangko Sentral ng Pilipinas. Rounded off for brevity.

Australia provides compensation for both physical and psychological injury and disease. According to SWA, an employee may be entitled to workers' compensation for a psychological injury such as PTSD, anxiety and depression, if (1) the worker meets the scheme's definition of worker or deemed worker, and (2) the work is a significant contributing factor to the psychological injury.¹⁴⁷ Employees affected by psychological injuries are also entitled to more time off work, recognizing that psychological injuries tend to be more complex than physical ones.¹⁴⁸

The importance of the described hazards being recognized under OSH laws cannot be denied, as it is recognition that would create a level of accountability on the part of employers. It would give workers solid ground to stand on when asserting their rights in the workplace. However, as this comparison between jurisdictions would reveal, a strong and comprehensive mental health policy framework will also be instrumental in providing workers with the support that they need. Finally, there must be an enforcement system that compels employers to create a safer and more supportive environment for employees.

The laws of the Philippines and India alike could look to the OSH laws of Australia and adopt some of their policies, considering the distinct circumstances of their own States, and understanding how they may most be effectively integrated into already existing laws. By considering mental and psychological hazards in the codified OSH standards, coupled with comprehensive mental healthcare coverage and strong compliance mechanisms, Filipino content moderators may be adequately protected from the risks and hazards of their work.

VI. PENDING PHILIPPINE LEGISLATION AND RECOMMENDATIONS

As of October 2024, there are no laws in the Philippines specifically geared towards content moderation work. Even the pending bills and proposals in Congress aim to address concerns of BPO workers in general: additional benefits, such as mandatory hazard pay for mental health check-

¹⁴⁷ *Workers' compensation for psychological injuries*, SAFE WORK AUSTR., at <https://www.safeworkaustralia.gov.au/workers-compensation/workers-compensation-psychological-injuries>.

¹⁴⁸ *Id.*

ups, among many other benefits, and compliance with international standards on occupational safety.¹⁴⁹

Earlier discussed is H.B. No. 8189 or the Magna Carta for BPO Workers, which seeks to add protection to the right of security of tenure of BPO workers, and entitle workers to medical benefits upon entry into the BPO company and not merely upon regularization, to highlight a few.¹⁵⁰

Notably, the Magna Carta seeks the entitlement of BPO workers to medical and health benefits, apart from Philhealth, such as the requirement of having a resident psychologist onsite available 24/7. The bill enumerates some examples of stressors such as “high pressure to meet performance targets” and “exposure to obscene and violent content, especially for content moderators.”¹⁵¹ Should the Bill be passed with this provision, it would be the first time content moderators are expressly recognized and protected in a national law. This could be further strengthened if their work can be classified as occupational risk or hazard, to afford workers full benefits under the law, such as the ECP injury and death benefits.

Under the “Better Working Conditions” section, the Bill also requires the employer to allow employee representation in the OSH committee through election, as stipulated by the DOLE.¹⁵² It makes no distinction as to whether the representation would be required through a union. In any case, this would provide a platform for workers to voice out workload-related concerns.

The Bill also requires each BPO establishment to address the OSH concerns in BPO workplaces and worksites, in accordance with OSH standards and other related issuances. It addresses directly the issue of self-reporting by mandating that OSH standards be enforced and inspected by the concerned and authorized departments, prohibiting any acts of self-regulated OSH standards inspection. In addition, the company shall also institutionalize OSH committees with genuine and sufficient representation

¹⁴⁹ H. No. 8189, 19th Cong., 1st Sess. (2023); H. No. 8733; 19th Cong., 2nd Sess. (2023).

¹⁵⁰ H. No. 8189, 19th Cong., 1st Sess. (2023). This is the bill filed by Representatives Raoul Manuel, France Castro, and Arlene Brosas.

¹⁵¹ H. No. 8189, 19th Cong., 1st Sess. § 25(c) (2023).

¹⁵² § 23(c).

from the employee sector.¹⁵³ Finally, the Bill includes a penal provision, subjecting a violator to fines or imprisonment.¹⁵⁴

Though not as comprehensive as the proposed Magna Carta, House Bill No. 8733 (“H.B. No. 8733”), filed by Representative Luis Raymund Villafuerte, Jr., also seeks to protect the labor rights of BPO workers by requiring that the minimum provisions of the OSH standards established by DOLE to meet the ILO recommendations.¹⁵⁵ This could be highly beneficial given the earlier cited provision which officially considered mental and behavioral disorders as occupational diseases, especially where the worker is exposed to risk factors arising from work activities,¹⁵⁶ as in the case of a content moderator.

On the part of the Senate, Senate Bill No. 1817 of the 19th Congress, filed by Senator Raffy Tulfo, seeks to promote mental health in the workplace by providing for mental health leave with full pay,¹⁵⁷ with DOLE as the implementing agency.¹⁵⁸

Several other Senate Bills, such as Senate Bill No. 920, filed by Senator Sonny Angara, and Senate Bill No. 2062, filed by Senator Mark Villar, seek to improve our laws on mental health, and expand the coverage of PhilHealth benefit packages to include better coverage for mental health, noting that its current coverage is very limited, and not sufficient to support a patient suffering from a mental condition.¹⁵⁹

Labor group BIEN consistently calls for the passage of bills protecting their rights, most particularly the Magna Carta for BPO Workers. Recently, the Senate concurred with the Philippines’ ratification of the ILO Convention No. C190,¹⁶⁰ which recognizes the right of everyone to a workplace free from violence and harassment.¹⁶¹ This was considered by

¹⁵³ § 26.

¹⁵⁴ § 29.

¹⁵⁵ H. No. 8733, 19th Cong., 2nd Sess., § 29 (2023).

¹⁵⁶ See List of Occupational Diseases Recommendation, ILO No. R194, Annex, June 20, 2002, *amended* Mar. 25, 2010.

¹⁵⁷ S. No. 1817, 19th Cong., 1st Sess., §8 (2023).

¹⁵⁸ § 9.

¹⁵⁹ S. No. 920, 19th Cong., 1st Sess. (2022); S. No. 2062; 19th Cong., 2nd Sess. (2023).

¹⁶⁰ ILO Violence and Harassment Convention, Convention No. 190, June 21, 2019.

¹⁶¹ *Senate concurs with ILO C190 ratification*, Press Release, SENATE OF THE PHIL. WEBSITE, Dec. 11, 2023, at https://legacy.senate.gov.ph/press_release/2023/1211_prib4.asp.

BIEN as a victory for BPO workers, many of whom have shared stories of verbal and psychological violence and harassment that often come from abusive managers and team leaders, and in other cases, from co-workers.¹⁶²

Despite the lack of specialized laws and regulations, there has been some progress in exploring the compensability of work-related mental health conditions, and in developing potential policy for the same. In 2024, the ECC convened with the Institute for Labor Studies (“ILS”), where it was reported that despite the policies already released by DOLE, additional training for mental health awareness in the workplace is still being formulated by the OSH sector, giving emphasis to the sanctions attached thereto.

As to the inclusion of mental health disorders as a compensable contingency under the ECP, the ILS recommended the establishment of parameters or determinants of work-related mental health conditions. The ECP would include catering to the mental health conditions of workers who suffer from PTSD, particularly if the condition was especially triggered by work-related sickness or disability.¹⁶³ Should these parameters be adequately determined, it could be a huge step for content moderators to finally receive the support that has long been needed and demanded.

This is neither to say that an amendment or addition to the law will automatically solve all of the problems presented. As earlier shown, it is not enough to merely have the mandates in place. The big issue to be solved by policymakers and department heads relates to how said mandates can be enforced effectively.

¹⁶² *C190 Ratified! A Big Win for Workers!*, BPO INDUS. EMP. NETWORK, Dec. 11, 2023, at <https://bienphilippines.wordpress.com/2023/12/11/c190-ratified-a-big-win-for-workers>.

¹⁶³ E. De Guzman, *ECC and ILS to explore compensability of work-related mental health conditions*, ECC WEBSITE, Sep. 18, 2024, at <https://ecc.gov.ph/ecc-and-ils-to-explore-compensability-of-work-related-mental-health-conditions>.

VII. CONCLUSION

Based on the overview of the existing body of labor law, there is a lack of space for content moderators to be adequately protected. Given the particularity and severity of their grievances, this space must be created by further legislation that is rooted in research and consultation, and an overall consideration of the present context. The Philippines, similar to India—the two countries leading the global BPO industry—struggle in protecting BPO workers and meeting the OSH standards established under existing ILO conventions, compared to Australia.

There is also an overall need to strengthen existing laws on regularization, benefits, and security of tenure, to allow Filipino BPO workers to maximize the remedies they would receive under the law. More importantly, the classification of content moderation as a hazardous work process, classified under potential mental or psychological health risk, may be a concrete and specific way to offer this protection.

In addition to more specific laws, stronger compliance and enforcement mechanisms are also necessary to ensure not only that companies are providing the care that their workers need, but also that the government agencies involved are fulfilling their mandate. The State should thus take a more proactive role in monitoring compliance, and in likewise providing adequate support and meaningful incentives for non-compliant companies.

Policy should also be geared toward mitigating the risk of mental illness as a result of the work. A comprehensive mental health benefits coverage, including the provision of free psychological services onsite, mental health leaves, or compensation for workplace psychological injuries, would certainly ease the financial burden on workers in content moderation, and allow them to receive the support they need. However, any contributory factors, such as excessive workload, forced night work, outdated systems, or anything more likely to cause a “high-strain” work environment, should also be eliminated.

In sum, further legislation, followed by stricter implementation, is required to ensure that “cleaning” social media sites is done with as minimal risk and hazard as possible to those most exposed to the perils of graphic and disturbing content. This is a non-negotiable goal. After all, content moderation work is nothing short of essential to the survival of any social media platform. To ensure a safe experience for the billions of people who

use them, many of whom are Filipinos themselves, someone has to do the difficult but necessary work of cleaning.

Cleaners should never be overlooked. While steps are being taken to expand and refine labor law in order to address the new and specific needs of BPO workers at large, content moderators are still waiting to receive good news. But it is clear that what is good news for them is good news for all. In empowering our content moderation BPO workforce, the State would not only carry out its own policy of affording full protection to labor, but also ensure public safety.

- o0o -