



International Journal of Technology, Policy and Management

ISSN online: 1741-5292 - ISSN print: 1468-4322

<https://www.inderscience.com/ijtpm>

Supporting digital key workers: addressing the challenges faced by content moderators during and after the COVID-19 pandemic

Elena Martellozzo, Paul Bleakley, Paula Bradbury, Ruth Spence, Jeffrey DeMarco

DOI: [10.1504/IJTPM.2024.10058414](https://doi.org/10.1504/IJTPM.2024.10058414)

Article History:

Received:	11 April 2023
Last revised:	06 July 2023
Accepted:	06 July 2023
Published online:	05 April 2024

Supporting digital key workers: addressing the challenges faced by content moderators during and after the COVID-19 pandemic

Elena Martellozzo*

Department of Criminology and Sociology,
Middlesex University,
London, NW4 4BT, UK

Email: e.martellozzo@mdx.ac.uk

*Corresponding author

Paul Bleakley

Henry C. Lee College of Criminal Justice and Forensic Sciences,
University of New Haven,
West Haven, 06516, USA

Email: pbleakley@newhaven.edu

Paula Bradbury, Ruth Spence
and Jeffrey DeMarco

Middlesex University,
London, NW4 4BT, UK

Email: p.bradbury@mdx.ac.uk

Email: r.spence@mdx.ac.uk

Email: j.demarco@mdx.ac.uk

Abstract: Content moderators (CMs) are private or state sector professionals responsible for inspecting user-generated content and applying policies to decide whether contentious content should remain online or be removed. CMs were required to review an unprecedented increase in illegal content during the COVID-19 pandemic. The existing evidence, although mostly anecdotal, suggests that frequent exposure to such content affects CM's well-being, mental health, and quality of life, as well as their work competence, satisfaction, and productivity. This study utilises ($N = 9$) in-depth qualitative interviews with CMs, which explored the impact of this work in the exceptional and challenging period of the COVID-19 pandemic. The aim of this paper is to critically explore what lessons can be learned from CM's experiences at work and how they coped with the difficulties created by COVID-19. This research will in turn assist organisations in identifying what is needed to improve conditions and mental well-being in the workplace.

Keywords: content moderators; social media; tech industry; labour conditions; digital first-responders; COVID-19; CSAM; lockdown; work-from-home; mental well-being.

Reference to this paper should be made as follows: Martellozzo, E., Bleakley, P., Bradbury, P., Spence, R. and DeMarco, J. (2024) ‘Supporting digital key workers: addressing the challenges faced by content moderators during and after the COVID-19 pandemic’, *Int. J. Technology, Policy and Management*, Vol. 24, No. 2, pp.212–228.

Biographical notes: Elena Martellozzo is an Associate Professor in Criminology at Middlesex University, London. She also serves as an Associate Director at the Centre for Abuse and Trauma Studies. Her research focuses on online harms, with an emphasis on the production of innovative practitioner-focused research.

Paul Bleakley is an Assistant Professor of Criminal Justice at the University of New Haven, Connecticut. He was appointed a University Research Scholar in 2023 in recognition of his research program examining the prevalence (and prevention) of cyber deviance in modern society.

Paula Bradbury is a Lecturer in Criminology and Policing at Middlesex University, London. She is a research fellow at the Centre for Abuse and Trauma Studies, with an expertise on policing in a digital age.

Ruth Spence is a senior research fellow at the Centre for Abuse and Trauma Studies, with a primary emphasis in psychological research.

Jeffrey DeMarco is an Associate Director at the Centre for Abuse and Trauma Studies, with a focus on forensic psychology and online criminality.

1 Introduction

COVID-19 is the first pandemic in 100 years to pervasively affect the entire world (Maison et al., 2021) and led to a global health crisis with major economic, social, and psychological consequences, forcing the worldwide community to adapt and respond quickly. There is a general consensus within the online child protection field that one outcome of the COVID-19 pandemic was a net increase in the production and sharing of child sexual abuse material (CSAM) (Parks et al., 2020). In the early phase of the pandemic, the WePROTECT Global Alliance noted that COVID-19 had already “exacerbated drivers of online child sexual exploitation, providing new opportunities for abusers ... [and] it is highly probable that numbers of online child sexual abuse cases will increase” (2020, p.2). This prediction was ultimately borne out as accurate by looking at the increase in reporting¹: Salter and Wong (2021) recorded a 122% increase in public reports of online CSAM to the Australian Centre to Counter Child Exploitation (ACCCE) between April and June 2020, with the American National Centre for Missing and Exploited Children (NCMEC) registering a 97.5% annual increase in online enticement reports from 2019 to 2020 (Salter and Wong, 2021). Similarly, in the UK there was a 50% increase in reports of CSAM from members of the public during lockdown to the Internet Watch Foundation

(IWF), a charity responsible for identifying and removing images and videos of child sexual abuse from the internet in the UK.

The increase over this period was most severe in March 2020, with 11,689 public reports of online CSAM material to the IWF by members of the public (IWF, 2020). Reports of online CSAM were increasing up to 50% per annum, even before the pandemic (Bursztein et al., 2019). However, in the pandemic period the reported rate reflects around double this ‘usual’ increase in reports. A report published by Europol in June 2020 reinforces the sharp increase in online CSAM activity coinciding with the introduction of global lockdowns around March 2020; police monitoring of dark web CSAM forums indicate a direct correlation between the lockdowns and prevalence of online CSAM, where members alternatively cited ‘benefits’ to the lockdowns, such as driving young people online where they could be targeted, and reported ‘challenges’ such as the lockdown’s impact on restricting their ability to abuse children in a physical context (Europol, 2020, p.12).

Regardless of whether COVID-19 was seen as an opportunity or barrier, the overall effect was to further drive CSAM producers and consumers into the online space, which is supported by data supplied by agencies like the ACCCE and NCMEC.

While entities like the ACCCE and NCMEC perform a critical role in combatting online CSAM, in practice it largely falls to online platforms, such as social media sites, to ‘police themselves’ as a product of the long-standing, historical preference for self-regulation in cyberspace (Cusumano et al., 2021). This role is typically performed by content moderators (CMs) employed by the private sector to respond to reports of inappropriate user-generated content (UGC) that breaches the terms and conditions of a particular website. When a CM determines that UGC does not meet the platform’s standards (or, often, is illegal), it is their duty to remove this content and, where appropriate, report it to the relevant authorities (Roberts, 2016). As Gray and Suri (2019) note, the role of the CM is purposefully designed to be invisible. Roberts (2014) interprets content moderation as one branch of a paradigm of control on the internet that goes “undetected by most, and thus help[s] constitute an illusion of volition and participation that is not reality-based” (Roberts, 2014, p.68). As private sector employees, not sworn law enforcement officers, the CMs role is not focused solely on child protection: rather, as Roberts (2014, p. 147) argues, they are “indispensable to the sites for which they labour ... [because they] guard against serious infractions contained in UGC that might do harm to a social media platform or company’s digital presence”. This is not to suggest that CMs do not also have a personal investment in child safety in addition to their corporate responsibilities. Bellanova and De Goede (2021) perceive the CM role as central to a “co-production between private expertise and public security” wherein decisions impacting on child protection are developed at an ‘intersection’ at which policing decisions are impacted by commercial considerations (and vice versa).

Using in-depth interviews, this paper focuses on how CMs have managed to continue to protect children during the challenging period of the COVID-19 pandemic, and how they have responded to the adjustments that their organisations put in place to support their work. Some general lessons can be learned from CM’s experiences working during the COVID-19 pandemic where, like many other

workers, they were forced to shift towards working remotely with likely effects on their mental well-being. The objective of this research is to help organisations reflect on what they need to do to improve mental well-being in the workplace and, in doing so, contribute to the development of a best practice model for content moderation, which is currently lacking.

2 Literature review

2.1 Changes in the work environment

The COVID-19 pandemic fundamentally changed the way that people interacted and fortified a reliance on the internet for social connection and communication. As countries around the world went into lockdown and social restrictions were put in place to prevent the spread of the virus, the proliferation and volume of CSAM also became an increasing concern (Europol, 2020; IWF, 2020). Just as the COVID-19 pandemic (and, importantly, related lockdowns) contributed to a rise in the *rate* of online CSAM proliferation, so too did it effect the capacity of CMs to identify, remove, and report it. While companies have, in the past, adopted a work from home (WFH) model for some CMs (see Roberts, 2014), the COVID-19 pandemic created a context in which working from an office environment was not feasible for most (Tubaro and Casilli, 2022). Dabrowska (2021, p.116) claims that WFH orders had direct consequences for content moderation, in that it led to a “limited number of [CMs] doing their job because of pandemic confinement, as CSAM viewing is not usually possible in remote work settings”.

It is also important to note that reporting CSAM is mandatory. Mandatory reporting refers to the legal obligation imposed on electronic service providers (ESPs) to report instances of CSAM discovered on their platforms to the appropriate authorities, such as the police or a designated mandated entity, like a hotline (Inhope.org). Generally, ESPs are not required to actively seek out CSAM but are obligated to report it once they become aware of its presence. Failure to fulfil their reporting obligations can result in various penalties imposed on ESPs. These penalties are in place to ensure compliance and hold providers accountable for their role in combatting CSAM.

Whilst artificial intelligence (AI) content moderation is about creating machine learning algorithms that can successfully detect inappropriate and illegal content, machines can still miss important nuances, such as hate speech and CSAM, due to its context dependent nature (Hill, 2022; Oliva et al., 2021). It is therefore important to acknowledge that current AI systems still heavily rely on human involvement. While there are ongoing discussions and debates about the capabilities and limitations of AI, there is a hidden aspect that demands attention. Gray and Suri (2019), in their book “Ghost Work”, shed light on a growing concern regarding the indispensable role played by an invisible army of human workers in enabling the seamless operation of services provided by tech giants like Google, Microsoft, Facebook and Twitter. These companies, known for their cutting-edge technologies, heavily depend on the expertise, judgement, and experiences of a vast workforce of human contributors. Behind the scenes, these individuals work tirelessly to ensure the smooth functioning of AI-driven services.

AI and machine learning can also lead to higher rates of false positives, which takes up human CMs time as they double-check the algorithms work, effectively making their work even more complicated (Duggan, 2023). Human intervention is essential in identifying and classifying Child Sexual Abuse Material (CSAM) due to the complexity and sensitivity of the content. Humans possess contextual understanding, cultural knowledge, and experience to accurately interpret and differentiate CSAM from legal depictions. They navigate legal and ethical considerations, stay updated on evolving trends, and contribute expertise to improve AI systems. Furthermore, human moderators handle the emotional impact, verify AI-generated results, and ensure the accuracy of identification, making their role crucial in protecting children and upholding standards (Tabi et al., 2023). Dabrowska also reinforces the importance of utilising human CMs, rather than a reliance on AI. Dabrowska refers to statements from Meta's global head of safety in 2021, who claimed that 90% of CSAM content referred to the company by NCEMC in October and November 2020 (the pandemic period) was "the same or visually similar to previously reported content" which, in a sense, suggests that the prevalence of 'unique' CSAM was not as high as reports would indicate (Davis, 2021). However, as Dabrowska notes, this "does not necessarily mean that CSAM rates on Facebook and other social media platforms have been low during the pandemic" and, instead, indicates a heavy reliance on AI technology that detects previously identified material, rather than utilising human moderators who are "critical in assessing previously unreported CSAM" (2021, p.116). With COVID-19 forcing a slowdown of content moderation – especially in the uncertain, early months of the pandemic – the essential human evaluation provided by CMs may have been limited, contributing to an incommensurate, rapid rise in online CSAM that occurred in the same timeframe.

Salter and Wong (2021, p.36) also highlighted deficiencies in commercial content moderation during the early COVID-19 period (March 2020-September 2020) as a barrier to effective CSAM response. Interviews conducted with investigations managers indicated 'significant frustration' with social media companies and a level of doubt that technology companies "had the capacity to detect even the basics" of CSAM on their platforms. Further, Salter and Wong reported that several CSAM analysts attributed much of the proliferation of online CSAM during the pandemic to perceived reductions in commercial content moderation. These interviewees reported a "delayed response of social media companies to take-down notifications and escalated complaints from [their] agency" and repeated reports from members of the public that "the platform is either not responsive or they get that sort of auto generated message saying ... because of COVID-19, their moderator team has decreased or there's not the same amount of staff and so it will take longer to get a response" (Salter and Wong, 2021, p.37). Despite these reports, a number of social media companies ultimately *did* transition CMs to a WFH environment which, though a response to criticism about cutbacks to CMs and moderation services, opened up a range of additional considerations and risks, particularly for CMs and their personal well-being.

2.2 Challenges for content moderators

Roberts (2014) reported noticeable challenges for CMs to disengage in a WFH context, even absent the additional emotional weight of working specifically on

CSAM. Steiger et al. (2021) also refer to the heightened risk of exhaustion and burnout for CMs, especially in a workplace setting which is largely unmonitored, and where disengagement is more difficult due to a lack of separation between work and personal life, as in the case of a WFH arrangement. Steiger et al. go on to observe that “developing a social support network may not be as simple to implement for content moderators ... [due to] industrial stratification and geographic distribution” making it challenging for CMs to create community within the industry (2021, p.9).

There is little doubt that this isolation and lack of social support becomes even more impactful in a WFH context, such as that which was forced on many workers as a result of the COVID-19 pandemic (Xiao et al., 2021), including those in the tech industry like CMs. As Steiger and colleagues acknowledge, there is little extant research on the effect that regular consumption of traumatic content has on CMs in general, let alone in a WFH context and, even less, during a global pandemic. This represents a clear gap in the literature that can be addressed by direct engagement with commercial CMs and lateral professionals working in the online CSAM field, which this research presents.

3 Research method

In total, nine ($N=9$) CMs were interviewed as part of this research. Participants came from different countries including: the UK, Ireland, Canada, Germany, the Netherlands, Colombia, Mexico, and India. All participants were recruited through professional networks and via social media platforms such as LinkedIn and Twitter and worked for internal and external (i.e., outsourced) moderation companies. The interviews were conducted online in English, lasting from 47 min to 78 min and focused on ‘external stressors’ such as moderating CSAM during the COVID-19 pandemic.

The interviews were semi-structured with participants asked about their role, how the material affected them, how they coped, other occupational stressors, and how the job was perceived by others. This paper focuses on data relating to how the CMs were affected by COVID-19. Relevant questions asked included: *How did COVID-19 impact on your work? What were the positive effects of the pandemic? What were the negative effects of the pandemic? What lessons can be learned?*

Participants were offered a choice of taking part in the interview by telephone or through the use of the online video conferencing program, Zoom. This choice enabled participants to decide which method they were most comfortable with and maintain anonymity (e.g., by completing the interview by telephone or switching the camera off during the interview). Informed consent was gained prior to the commencement of the interview and monitored throughout. Participants were told their answers were confidential and voluntary, they did not have to answer questions they were not comfortable with, and they could take breaks or stop the interview at any time. A debrief with signposting to support services was available for participants if needed. Recordings were professionally transcribed.

The transcripts were thematically analysed using Dedoose, a qualitative analytic software. The key themes were identified through familiarisation with the transcripts, which allowed ‘parent’ and ‘child’ codes to be identified. Two researchers spent time coding the transcripts independently, as a form of inter-rater reliability, with each

researcher periodically checking the others' coding for agreement and consistency. This process encouraged theme development and revision involving the wider research team.

3.1 Ethics

Ethical approval was granted for the study by the university's psychology research ethics committee, with anonymity and confidentiality guaranteed, and all data secured in line with Data Protection and UK GDPR regulations.

4 Results

Through a process of thematic analysis, one parent theme (impacts of the pandemic on content moderation) and five child themes were developed. The themes identified are outlined in Table 1 and subsequently discussed in this section.

Table 1 Themes identified through thematic analysis

<i>Impacts of the pandemic on content moderation</i>
Increase in CSAM
Switching off from work
Efficient way of working
Separating home life from work life
Increased isolation

4.1 Effects of the pandemic on content moderation

4.1.1 Increase in CSAM

CMs expressed their concerns with the increase in creation and distribution of online CSAM, which proved challenging both for individual CMs and the companies they worked for. The influx of CSAM and associated reports placed an increasing strain on the workforce, who felt unable to keep up with the increased demands of responding to, classifying, and removing large volumes of content, and in some organisations, proactively searching for CSAM. This led to feelings of burnout, wherein CMs felt that no matter how much work they completed they were unable to meet the demands of their role:

“Content-wise, it was hectic, I think is the best way to put it. Suddenly, we were receiving so many cases or we were seeing so much more content on the platform that we had to take down proactively. We had no other option than to finally hire more people and, even then, we're simply not enough.”
(CM 1)

Even when companies increased capacity and employed new staff, CMs continued to report feeling overwhelmed and exhausted. It is likely that these experiences led to a burnout workforce that was at heightened risk of experiencing psychological, physiological, cognitive, and relational impacts of the work.

4.1.2 *Switching off from work*

Switching off from content moderation is challenging, however switching off when WFH is even harder: their homes became their workspaces. This meant losing the emotional separation from work-related thoughts, without performing those routine activities such as driving home or having a coffee with a colleague, which proved challenging for many CMs (Bolisani et al., 2020).

Employees that are unable to decompress from work-related stress may feel tense, frustrated, and ultimately burnout (Royal Society for Public Health, 2021). Some CMs had to deal not only with an increase in their workload due to the increased production of images, but also with governmental orders to self-isolate, forcing them to abandon in-person human encounters. In some cases, this meant facing difficulties in separating their professional and personal life. CMs felt overwhelmed by the amount of content they were required to process and felt trapped – both physically (e.g., in their homes), and mentally as a result of being unable to switch off from their work or using their usual coping mechanisms.

4.1.3 *Efficient way of working*

Alongside reviewing, classifying, and removing CSAM, CMs had to adapt rapidly to changed working conditions, which included working from home and/or an empty office. Through the interviews, it emerged that these changes in working conditions had both negative and positive effects on the way CMs operate. However, this experience depended on factors like the resources CMs had available at home (e.g., access to a private home office or having to work from a bedroom), and whether they had any dependents to look after whilst working, or to protect from viewing the content, adding heightened stress.

Some positive outcomes from the pandemic emerged from interviews. With increasing numbers of employees working at home or using their home as a working base for few days each week, there were some professional advantages such as flexibility and agility; increased productivity; financial benefits; and convenience. Some CMs reported that WFH was a net positive, as they found it to be a more efficient way of working. These CMs reported that WFH encouraged autonomy, allowed them to balance their workload, create their own routine, design their optimal work environment, and minimise distractions:

“I’d say that working from home I am way more effective than working from an office. I have my own set-up, my routine and everything. I love working from home better than from an office so I would say that this was, definitely, a benefit.” (CM 1)

CMs discussed how WFH meant that they could allocate protected time to view, identify, and classify their CSAM allotted quota without distractions, therefore planning and utilising their time more efficiently.

4.2 *Disentangling home and work life*

WFH required CMs to view CSAM in their own homes, which permeated their ‘safe spaces’ which had previously been free of the traumatic images and videos viewed in the physical workplace. It also meant that they had no distinguishable separation

between work and home, making it difficult to remove thoughts of CSAM and other elements of their work from their minds.

Prior to the pandemic, CMs had time to decompress and process their thoughts on their commute home from work. The physical distance between the office and their home encouraged a sense of separation and detachment from work and fostered a better work–life balance. CMs who worked at home during the pandemic, especially those who did not have a separate room to use as an office, reported that maintaining a work–life balance was difficult. For CMs who lived with their family, protecting them from viewing CSAM became an important consideration, and having to constantly shield these individuals from seeing their screens increased the sense of secrecy surrounding the role:

“What was hard for me was to go to work because I had to go to the office, not every day but three days a week and then go home and especially during the lockdowns, you’d just go to work and then go home and do nothing else. Yes, I went on walks but that was it. I had not any recreation time that I could create myself in a way that I needed and that was very hard.” (CM 2)

In addition, lockdowns and other restrictions on social behaviours made it increasingly difficult for CMs to access their usual coping behaviours. This was particularly difficult for individuals who used external coping mechanisms such as socialising or attending the gym since these were not accessible and could not provide an outlet for negative emotions. However, some CMs found positive ways of managing the traumatic work. They coped by compartmentalising the work, dividing their day by viewing CSAM in the morning and completing other work in the afternoon. This meant that they could end the day focusing on non-CSAM related work.

However, not all CMs had access to the same resources (e.g., a separate, safe room to view CSAM), and not all CMs lived alone and/or had no distractions. External and/or personal factors, such as the presence of children at home, significantly altered how CMs views WFH conditions.

4.2.1 *Increased isolation*

Content moderators raised concerns about the negative impact of working from home (WFH) on their ability to access essential emotional support required to effectively carry out their duties. The transition from working in a physical office environment to remote work disrupted their ability to tap into the emotional, physical, and mental resources necessary to navigate the challenging nature of their job.

In a traditional office setting, content moderators often had the advantage of direct, face-to-face interactions with their colleagues and supervisors. These interpersonal connections provided valuable emotional support, enabling them to share experiences or seek advice. The physical presence of supportive colleagues offered a level of reassurance and solidarity, helping content moderators to cope with the emotionally taxing nature of their work.

However, with the shift to WFH arrangements, content moderators have been deprived of these critical support systems. The physical separation from their peers has resulted in feelings of isolation and limited opportunities for informal discussions and debriefing sessions. The absence of immediate access to colleagues who

understand the unique challenges of their role has made it more difficult for content moderators to process and cope with the disturbing content they regularly encounter.

Furthermore, in an office setting, there may have been access to wellness programs, designated relaxation areas, or resources for stress management. Working remotely, however, content moderators may find it challenging to establish clear boundaries between work and personal life, leading to increased stress and difficulty in managing their mental and physical health:

“I found I couldn’t talk to anyone. I couldn’t tell anyone, but I don’t know what I could have told people anyway. Do you know, when I tried to tell people in the past about what I did, they didn’t want to hear about it. I don’t blame them. It’s like telling someone about a horror movie. They don’t want to hear about horror films. The only unfortunate thing if you watch horror films, it has effect on you detrimentally when you’re on your own.” (CM6)

Similar difficulties existed those CMs who continued to go into the office to work. However, this was physically isolating, as they were a large minority during a time in which communities were confined to their homes:

“There was this one time where I just asked a colleague to talk to me for a minute, and it was in the middle of COVID, so there was no one at the office besides us, and we could just sit somewhere else and talk about it, and I just cried.” (CM3)

In these cases, the content moderation companies that CMs worked for recognised the risks associated with doing this job from home such as isolation, lack of separation from work and home, increased stress, and the risk of other individuals being exposed to CSAM. As a result, they required their employees to work from an office environment, under exemptions to lockdown catering for key workers.

Although the ability to leave the house and go to work may initially seem like a positive aspect for some individuals, it is important to consider the challenges faced by CMs, who experienced similar difficulties despite the change in environment. One significant drawback was the lack of external social opportunities that could serve as distractions or provide a means of decompression. Without these outlets, CMs were deprived of the chance to utilise adaptive coping mechanisms, including engaging in sports or accessing creative outlets, which are known to help alleviate stress and promote well-being.

Moreover, some companies required their employees to return to work without adequately addressing their specific needs or providing necessary support. This oversight further compounded the challenges faced by CMs and hindered their ability to perform their jobs effectively. Without the assistance and resources, they required, CMs may have struggled to navigate their roles and fulfil their responsibilities to the best of their abilities. Overall, the absence of external social opportunities, coupled with the lack of support from companies, resulted in a difficult work environment for CMs.

While the psychological, physiological, cognitive, and relational difficulties faced by CMs existed before the pandemic, COVID-19 amplified these difficulties by removing most options for engaging in accessing mainstream coping strategies. Despite these challenges, some CMs not permitted to WFH felt that this decision was a useful one, as it allowed them to carry out their work effectively, whilst benefitting from support and interaction with their colleagues:

“Well, we aren’t able to work at home because of the material. We always have to be with someone else, and I thought that was good. It gave me a reason to go outside and not be stuck at home. I guess I found it quite comfortable knowing that we were the only ones at the office, because the rest of the other team in our company could work from home, so it was just my team, and I felt comfortable with that.” (CM 3)

Those CM saw that working from the office was beneficial because they were able to keep in contact with colleagues, struggled less and, in some cases, CMs noted that they found the experience less anxiety provoking because they did not have to protect others from seeing the content they worked with, whether colleagues or family members.

5 Discussion

The global pandemic compelled workers worldwide to transition to remote work arrangements, which had both positive and negative implications. On one hand, there was an observed increase in productivity resulting from this shift. On the other hand, the significant drawback was the prevalent sense of social isolation experienced by many individuals (Kim and Jung, 2021). However, content moderators faced distinct challenges during this time. Not only did they grapple with the effects of social isolation like their counterparts, but they also encountered a unique circumstance. The volume of work they were tasked with handling saw a simultaneous increase (Steiger et al., 2021). This created an additional burden for content moderators who were already navigating the complexities of remote work. The COVID-19 pandemic affected the work of content moderators (CMs) by increasing the volume of work they were required to handle (Steiger et al., 2021).

As the pandemic shifted many activities online, there was a surge in user-generated content, leading to a greater demand for moderation and an increased workload for CMs to ensure the safety and integrity of online platforms. CMs reported feeling unable to ‘keep up’ with the increased volume of CSAM and, despite companies hiring new staff to absorb some of the workload, reported feelings of being overwhelmed, exhausted, and frustrated. These conditions can result in burnout, wherein overwhelmed employees experience symptoms of physical, mental, and emotional exhaustion (Kreiner et al., 2006). While employee mental health and well-being should be a major consideration for management and employers, this is not the only negative outcome burnout produces: on a pragmatic level, it also effects the employees’ ability to perform their duties and, in the case of CMs, adequately respond to CSAM (Parasuraman and Greenhaus, 2002).

Furthermore, work-from-home practices were also adopted more extensively in the content moderation (CM) field as a result of the COVID-19 pandemic. However, it is important to carefully consider the requirements and implications of sustaining such working conditions in the long term, for content moderators’ well-being. CMs who participated in WFH suggested that symptoms of burnout were exacerbated by challenges in separating work and home life. They reported that maintaining a work-life balance was more difficult in the WFH context, largely due to a perceived inability to ‘switch off’ at the end of the workday because their physical workspace overlapped with their safe, home environment. Once again, these anecdotal reports

are supported by the existing research on burnout, which notes that successful management of work-home interfaces is key in addressing employee risk, particularly in high-intensity professions like content moderation (Pratt and Rosa, 2003).

Another product of the shift to WFH was a perceived lack of formal and informal support accessible to CMs. CMs reported not being able to decompress from their work on an emotional level. Notably, social isolation was also cited as a negative factor by those CMs who continued working from an office environment. In these situations, there were fewer other CMs present in the workspace, in order to comply with social distancing regulations. This meant there were less people for CMs to turn to for informal support in the workplace who shared the same experiences and could potentially relate to the multiple stressors of content moderation.

Our findings suggest that to a large degree, CMs found that they were responsible for maintaining their own well-being during the pandemic, with even those who were brought back into the office having little infrastructure put in place to support them. Therefore, for the tech industry to effectively sustain WFH practices in content moderation, several factors need to be addressed on Well-being and Support. They must prioritise the well-being of content moderators by providing mental health resources, regular check-ins whether conducted remotely or in person. Furthermore, fostering a sense of community and connection through virtual platforms can help mitigate the social isolation that may arise from remote work.

While WFH has become more prevalent, it is important not to overlook the nature of content moderation work, which often involves exposure to disturbing and graphic material, presenting unique challenges. Research has found that people outside of the industry do not want to know about the role and moderators do not want to burden others with their experiences (Spence et al., 2023), a phenomenon which has been found in law enforcement who work in child exploitation and terrorism (Brady, 2017). Therefore, working remotely may intensify the emotional toll on moderators, as they lack the immediate support of their colleagues, as well as the physical separation from distressing content that a dedicated office environment can provide. There are concerns that prolonged remote work in content moderation could exacerbate issues such as isolation, burnout, and difficulty in separating work and personal life. The absence of in-person collaboration and immediate support systems could also hinder effective teamwork and decision-making processes (Spence et al, 2023).

This reinforces earlier research on law enforcement internet child exploitation (ICE) investigators, which found a perceived need to shield friends, family, and even other colleagues not directly dealing with CSAM from content (Burns et al., 2008; Brady, 2017; Reeve, 2020). This meant ICE investigators demonstrated a greater reliance on each other to provide informal social supports. The fact that social isolation was identified as problematic by both the WFH contingent and those who continued to work from an office emphasises the importance of informal social networks for well-being in this industry. This is also supported within the literature, with research indicating that social support is tied to effective psychological adjustment and enhanced resilience (Bonnano et al., 2007; Layne et al., 2007).

Although the responses recorded in this study highlight some potential barriers to WFH policies when applied to content moderation, CMs also identified a range of

potential opportunities that the COVID-19 enforced shift in workplace conditions revealed. CMs indicated that WFH provided them with more flexibility in managing their workday, presenting greater opportunities to avoid burnout. CMs reported being able to structure their workflow to avoid ending their day engaged with CSAM, providing a buffer zone between the work and home that prevented recollection of this content becoming intrusive in non-work hours. Building on the theme ‘efficient way of working’, some CMs highlighted the benefits that WFH had for engaging in intense periods of concentration. Working with CSAM material requires a great deal of concentration, and with workplace distractions from colleagues, it is possible that this enforced isolation allowed sustained engagement with daily tasks.

The level of concentration that WFH permits allows for shorter-duration, higher-intensity focus; this, in turn, allows a CM to not spend their entire day examining CSAM content, with the knock-on effect being greater ability to decompress at the end of the day with lower-intensity activities. Individual level coping strategies like this may be beneficial to both CMs well-being and job performance; however, it is important to note that not all companies that provide content moderation services are able to provide the same flexibility to give CMs higher degrees of control over their daily output, which is perceived to allow for more manageable target quotas and professional expectations (Newton, 2019). To attain the benefits of this WFH strategy, it would be necessary to review the expectations for CMs to determine whether they can reasonably accommodate more flexible work patterns and, if not, if there is potential to revise these expectations to allow for more individualised workload management or not.

As the industry reflects on these important findings, it is crucial to continue to engage in increased discussion about the implications and long-term effects of remote content moderation. Research and studies analysing the impact of WFH on content moderators’ well-being, job performance, and the quality of moderation outcomes can help inform best practices and shape future policies in the field.

6 Study limitations

One of the hallmarks of high-quality research is to transparently identify some of its limitations. While the greatest strength of conducting qualitative interviews is the depth and richness it provides, it is ultimately an interpretivist approach embedded in thematic analysis and can lend itself to researcher bias.

To counteract this risk, two researchers were involved in the process of coding the transcripts, and these codes were reviewed by the entire research team to ensure that they were consistent. The act of conducting a small number of individual interviews is also provides a more limited perspective of an experience, which may not be shared by others from the same organisation (Alsaawi, 2014) and cannot be generalisable. By expanding the recruitment of participants to a global network of CMs, however, the research was strengthened by being able to not just explore localised practices in content moderation during the COVID-19 pandemic, but rather a more global response to a similarly global event.

7 Conclusions

Across the world, more than 100,000 people are employed to moderate online content, including CSAM, with many of these CMs working in developing nations (Spence et al.'s 2022). This research has revealed that CMs who work with CSAM find this role demanding and has reinforced Steiger et al.'s (2021) conclusions that if the working conditions and support CMs receive are not appropriate, they may experience heightened levels of burn out, secondary trauma, and mental health difficulties. At present, it is not possible to fully automate effective content moderation (AI) for the foreseeable future (Cambridge Consultants, 2019) and, thus, we must rely on the expertise of human CMs to review highly disturbing content and ensure online safety.

Whilst the current Covid-19 seems to have abated in many countries, there are lessons which can be applied to further breakouts, or to working practice more generally if hybrid working practice continues. This research attests to the importance of CMs working in an appropriate environment where they are provided with professional support, clear communication, and education about the risks that content moderation involve. In this way, CMs can make informed decisions about their job choice and working practice. This research has also highlighted that moderating content in a WFH context might not be suited to everyone's personality or ability. Generally, some employees might prefer the routine and structure that working in an office environment provides them (Bolisani et al, 2020). Others prefer personal interaction with colleagues and find face-to-face guidance with their manager extremely beneficial in helping them complete tasks and achieve their goals (Nibusbusinessinfo, 2022). However, for those CMs that may be fully WFH or who have adopted a hybrid model, it is essential that they are still supported by in-person team to help them separate work from home life (Royal Society for Public Health, 2021) and create some clear boundaries.

Indeed, as CMs have shown, having boundaries in trauma work is essential, as it helps preserve emotional energy and can help mental well-being. The COVID-19 pandemic has been a moment of turmoil for society. Countries across the globe implemented a range of public health and social measures, including movement restrictions, and organisations were forced to improvise and effectively experiment with practical interventions in response to the unique conditions demanded by this epidemiological phenomenon. Although the turmoil of the pandemic is not completely over, it has provided us with the unique opportunity to critically analyse the lessons learnt from the past two and half years.

Support for digital key workers falls under the responsibility of the companies that employ them, as they have a responsibility to provide a safe and supportive work environment. This includes ensuring fair compensation, adequate training, and access to necessary resources and tools. Employers should also prioritise the well-being and mental health of their content moderators by offering counselling services, regular breaks, and supportive management practices. In accordance with Harold's argument (2022), prioritising the promotion of mental health and addressing mental illness is crucial for maximising productivity and creating a resilient workforce for the future. By closely examining the working conditions of frontline content moderators and their responses to changes in those conditions, we can strategically shape a workforce that is adaptable to evolving circumstances, while maintaining a strong

commitment to employee well-being within the content moderation sector and the broader tech industry.

Acknowledgements

The authors would like to acknowledge the work of Antonia Bifulco, Boglarka Meggyesfalvi, and Amy Harrison, who have all contributed in various ways to the overarching research project from which this paper derives.

Funding

This publication has been produced with financial support from the End Violence Fund. However, the opinions, findings, conclusions, and recommendations expressed herein are those of the authors, and do not necessarily reflect those of the End Violence Fund.

References

- Alsaawi, A. (2014) 'A critical review of qualitative interviews', *European Journal of Business and Social Sciences*, Vol. 3, No. 4, pp.149–156.
- Bellanova, R. and De Goede, M. (2021) 'Co-producing security: platform content moderation and European security integration', *Journal of Common Market Studies*, Vol. 60, No. 5, pp.1–19.
- Bolisani, E., Scarso, E., Ipsen, C., Kirchner, K. and Hansen, J.P. (2020) 'Working from home during COVID-19 pandemic: lessons learned and issues', *Management and Marketing. Challenges for the Knowledge Society*, Vol. 15, No. 1, pp.458–476.
- Bonnano, G.A., Galea, S., Bucciarelli, A. and Vlahov, D. (2007) 'What predicts psychological resilience after disaster? The role of demographics, resources, and life stress', *Journal of Consulting and Clinical Psychology*, Vol. 75, No. 5, pp.671–682.
- Brady, P.Q. (2017) 'Crimes against caring: exploring the risk of secondary traumatic stress, burnout, and compassion satisfaction among child exploitation investigators', *Journal of Police and Criminal Psychology*, Vol. 32, pp.305–318.
- Burns, C.M., Morley, J., Bradshaw, R. and Domene, J. (2008) 'The emotional impact on and coping strategies employed by police teams investigating internet child exploitation', *Traumatology*, Vol. 14, No. 2, pp.20–31.
- Bursztein, E., Clarke, E., DeLaune, M., Eliff, D.M., Hsu, N., Olson, L., Shehan, J., Thakur, M., Thomas, K. and Bright, T. (2019) 'Rethinking the detection of child sexual abuse imagery on the internet', [Paper Presentation]. *WWW'19: The World Wide Web Conference*, 13–17 May, San Francisco, CA, <https://doi.org/10.1145/3308558.3313482>
- Cambridge Consultants (2019) *Use of AI in Online Content Moderation*. Retrieved from https://www.ofcom.org.uk/data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf
- Cusumano, M.A., Gawer, A. and Yoffie, D.B. (2021) 'Can self-regulation save digital platforms?', *Industrial and Corporate Change*, Vol. 30, No. 5, pp.1259–1285.
- Dabrowska, M. (2021) 'The unclear picture of child sexual abuse material (CSAM) online volumes during the COVID-19 pandemic', *Białystok Legal Studies*, Vol. 26, No. 6, pp.109–125.
- Davis, A. (2021) *Preventing Child Exploitation on Our Apps*, Meta, Retrieved from: <https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps/>

- Duggan, N. (2023) *Advantages and Disadvantages of Artificial Intelligence*, Retrieved from Simplilearn website: <https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article#:~:text=No%20Creativity,can%20write%20Forbes%20earning%20reports>
- Europol (2020) *Exploiting Isolation: Offenders and Victims of Online Child Sexual Abuse during the COVID-19 Pandemic*, Retrieved from: https://www.europol.europa.eu/sites/default/files/documents/europol_covid_report-cse_jun2020v.3_0.pdf
- Gray, M.L. and Suri, S. (2019) *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass*, Eamon Dolan Books, Boston, MA.
- Harold, G. (2022) *Wellbeing at Work—Whose Job is to Fix It?* [Audio podcast], BennettInstitute for Public Policy and IAST, 30 October, <https://podcasts.apple.com/gb/podcast/wellbeing-at-work-whose-job-is-it-to-fix-it/id1597966909?i=1000584415369>
- Hill, K. (2022) ‘A dad took photos of his naked toddler for the doctor. Google flagged him as a criminal’, *New York Times*, Retrieved from: <https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html>
- Inhope (n.d.) Retrieved 6 July, from <https://www.inhope.org/>
- IWF (2020) *Definite Jump’ as Hotline Sees 50% Increase in Public Reports of Online Child Sexual Abuse during Lockdown*, Retrieved from: <https://www.iwf.org.uk/news-media/news/definite-jump-as-hotline-sees-50-increase-in-public-reports-of-online-child-sexual-abuse-during-lockdown/>
- Kim, H.H.S. and Jung, J.H. (2021) ‘Social isolation and psychological distress during the COVID-19 pandemic: a cross-national analysis’, *The Gerontologist*, Vol. 61, No. 1, pp.103–113.
- Kreiner, G.E., Hollensbe, E.C. and Sheep, M.L. (2006) ‘On the edge of identity: boundary dynamics at the interface of individual and organizational identities’, *Human Relations*, Vol.59, No. 10, pp.1315–1341.
- Layne, C.M., Warren, J.S., Watson, P.J. and Shalev, A.Y. (2007) ‘Risk, vulnerability, resistance, and resilience: toward an integrative conceptualization of posttraumatic adaptation’, in Friedman, M.J., Keane, T.M. and Resick, P.A. (Eds.): *Handbook of PTSD: Science and Practice*, The Guilford Press, New York, NY, pp.497–520.
- Maison, D., Jaworska, D., Adamczyk, D. and Affeltowicz, D. (2021) ‘The challenges arising from the COVID-19 pandemic and the way people deal with them. A qualitative longitudinal study’, *PloS One*, Vol. 16, No. 10, p.e0258133.
- Newton, C. (2019) *The Terror Queue*, The Verge, Retrieved from <https://www.theverge.com/2019/12/16/21021005/google-youtube-moderators-ptsd-accenture-violent-disturbing-content-interviews-video>
- Nibusinessinfo (2022) *Employees Working From Home. Advantages and Disadvantages of Employees Working at Home*, Retrieved from: <https://www.nibusinessinfo.co.uk/content/advantages-and-disadvantages-employees-working-home/>
- Oliva, T.D., Antonialli, D.M. and Gomes, A. (2021) ‘Fighting hate speech, silencing drag queens? Artificial intelligence in content moderation and risks to LGBTQ voices online’, *Sexuality and Culture*, Vol. 25, pp.700–732.
- Parasuraman, S. and Greenhaus, J.H. (2002) ‘Toward reducing some critical gaps in work-family research’, *Human Resource Management Review*, Vol. 12, pp.299–312.
- Parks, A., Sparre, C., Söderquist, E., Arver, S., Andersson, G., Kaldo, V., Görts-Öberg, K. and Rahm, C. (2020) ‘Illegal online sexual behavior during the COVID-19 pandemic: a call for action based on experiences from the ongoing prevent it research study’, *Archives of Sexual Behavior*, Vol. 49, No. 5, pp.1433–1435.
- Pratt, M.G. and Rosa, J.A. (2003) ‘Transforming work-family conflict into commitment in network marketing organizations’, *Academy of Management Journal*, Vol. 46, No. 4, pp.395–418.

- Reeve, Z. (2020) 'Repeated and extensive exposure to online terrorist content: counter-terrorism internet referral unit perceived stresses and strategies', *Studies in Conflict and Terrorism*, Vol. 0, No. 0, pp.1–25.
- Roberts, S.T. (2014) *Behind the Screen: The Hidden Digital Labor of Commercial Content Moderation*. [Doctoral dissertation, University of Illinois at Urbana-Champaign] IDEALS, <https://www.ideals.illinois.edu/items/50512>
- Roberts, S.T. (2016) 'Commercial content moderators: digital laborers' dirty work', in Noble, S.U. and Tynes, B.M. (Eds.): *The Intersectional Internet: Race, Sex, Class, and Culture Online*. Peter Lang, New York, NY, pp.147–160.
- Royal Society for Public Health (2021) *Survey Reveals the Mental and Physical Health Impacts of Home Working during Covid-19*. Retrieved from: <https://www.rsph.org.uk/about-us/news/survey-reveals-the-mental-and-physical-health-impacts-of-home-working-during-covid-19.html>
- Salter, M. and Wong, W.K.T. (2021) *The Impact of COVID-19 on the Risk of Online Child Sexual Exploitation and the Implications for Child Protection and Policing* (Research Report). UNSW Sydney, <https://www.arts.unsw.edu.au/sites/default/files/documents/eSafety-OCSE-pandemic-report-salter-and-wong.pdf>.
- Spence, R., Harrison, A., Bradbury, P., Bleakley, P., Martellozzo, E. and DeMarco, J. (2023) 'Content moderators' *Strategies for Coping with the Stress of Moderating Content Online*. *Journal of Online Trust and Safety*, Vol. 1, No. 5, p.0.
- Spence, R., Martellozzo, E. and DeMarco, J. (2022) *Invisible Workers, Hidden Dangers*, Retrieved from <https://www.bps.org.uk/psychologist/invisible-workers-hidden-dangers>
- Steiger, M., Bharucha, T.J., Venkatagiri, S., Riedl, M.J. and Lease, M. (2021) 'The psychological well-being of content moderators: The emotional labor of commercial moderation and avenues for improving support', [Paper Presentation]. *CHI'21: CHI Conference on Human Factors in Computing Systems*, 8–13 May, Yokohama, Japan, <https://doi.org/10.1145/3411764.3445092>
- Tabi, C., Hewage, C., Bakhsh, S.T. and Ukwandu, E. (2023) 'Contemporary issues in child protection: police use of artificial intelligence for online child protection in the UK', *Digital Transformation in Policing: The Promise, Perils and Solutions*, Springer International Publishing Cham, pp.85–107.
- Tubaro, P. and Casilli, A.A. (2022) 'Who bears the burden of a pandemic? COVID-19 and the transfer of risk to digital platform workers', *American Behavioral Scientist*, <https://doi.org/10.1177/00027642211066027> v
- WePROTECT and Global Alliance (2020) *Intelligence Brief: Impact of COVID-19 on Online Child Sexual Exploitation*, Retrieved from: <https://alliancecpha.org/en/child-protection-onlinelibrary/weprotect-global-alliance-intelligence-brief-impact-covid-19-online>
- Xiao, Y., Becerik-Gerber, B., Lucas, G. and Roll, S.C. (2021) 'Impacts of working from home during COVID-19 pandemic on physical and mental well-being of office workstation users', *Journal of Occupational and Environmental Medicine*, Vol. 63, No. 3, p.181.

Note

¹Increases in reports do not necessarily correlate to an increase in content. However, with little available data for the relevant period using other metrics (e.g., CSAM hashnets), public reports of content provide the best measure of prevalence.