Behind the Screen

Content Moderation in the Shadows of Social Media

SARAH T. ROBERTS

Yale UNIVERSITY PRESS
New Haven and London

Yale University Press books may be purchased in quantity for educational, business, or promotional use. For information, please e-mail sales.press@yale.edu (U.S. office) or sales@yaleup.co.uk (U.K. office).

Set in Minion type by IDS Infotech Ltd.
Printed in the United States of America.

ISBN 978-0-300-23588-3 (hardcover : alk. paper)
Library of Congress Control Number: 2018963972
A catalogue record for this book is available from the British Library.

This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

10 9 8 7 6 5 4 3 2 1

Contents

6
Digital Humanity

"You know, I've had conversations with people about my research and my interest in the topic," I told Max Breen, "and they say things . . . like, 'Well, why don't you just go get one of those jobs and document what it's like?' "

Max replied: "Don't put yourself through it."

Summer 2018, Los Angeles, California

Commercial content moderation, the human adjudication of online user-generated social media, is work that often occurs in secret and typically for relatively low wages. When I began my research in 2010, there was little language to describe professional moderation, locate where moderation happens in industrial contexts, or even identify who did the moderating. Understanding this phenomenon, giving voice to the workers who undertake it and situating it in its industrial context—among the firms that need these services and those that provide them—has become my life's work, in no small part because it has become clear to me that understanding moderation and the people who moderate is crucial for understanding the internet as we know it today. Over the past eight years, during which I have tracked these practices, met workers, and written about the moderators and the work they do in a variety of academic and public outlets, commercial content moderation as a fundamental facet of the commercial social internet has gained in prominence in the public consciousness and, increasingly, in the eye of legislators and regulators around the world.[1] In part, this is because academic researchers or investigators interested in the social impact of the internet have made this a concentrated area of study, often with little support and sometimes with active hostility from both industry and academia. Thankfully, this is now changing.

Key partners in bringing commercial content moderation to light have been the journalists covering Facebook, Google, and other firms, reporting on Silicon Valley, or responsible for similar beats related to social media. Some of them have published significant exposés that have had a deep impact: throughout 2014 I had lengthy conversations with the journalist Adrian Chen, who had covered some early commercial content moderation stories while at Gawker and was then writing a major feature for *Wired* magazine. His resulting article was the impetus for many people to consider, for the first time, the actions of social media companies in content gatekeeping, and the legions of commercial content moderation workers dispersed globally in places like the Philippines, where he traveled for the story. Chen's article continues to circulate today as a prime and early description in the popular press of commercial content moderation and its implications.[2]

Other reporters have followed suit, and the work they have done has been powerful: Olivia Solon and Jamie Grierson's "Facebook Files" for *The Guardian* in 2017, Julia Angwin's coverage of content moderation for ProPublica in 2017, Catherine

Buni and Soraya Chemaly's exposure on the "Secret Rules of the Internet" for *The Verge* in 2016, or the German case of content moderation reported by Till Krause and Hannes Grassegger in the *Süddeutsche Zeitung,* in 2016, are a few important examples.[3] The pressure from journalists has resulted in forcing responses from social media firms that has made it no longer possible for them to deny the existence of commercial content moderation workers as a mission-critical part of the social media production chain. Indeed, as scandal after scandal unfolded in online social spaces (murders streamed over Facebook video functions; the availability of disturbing content featuring or targeting children on YouTube; the eruption of concern over the existence of fake news on all platforms and its influence on electorates in various parts of the world), the moderators themselves were often invoked by the firms that employ them as the solution to the problems.

It was due to this pressure that we learned, for example, that Google planned to take its staffing to twenty thousand commercial content moderation workers dealing with a variety of its products and problems, and Facebook was acknowledging plans for half as many.[4] This stance represented a stark about-face from the past, when queries about commercial content moderation practiced at the platforms were met with silence or treated lightly by corporate spokespeople (such as Microsoft's "yucky job" quip to NPR's Rebecca Hersher in 2013). Now, it seemed, professional moderators were the front line of brand and user protection, although still, next to nothing was being said about the nature of the work and how the hiring would be undertaken. My sense was that the vast majority of those workers would be hired worldwide and that third-party contracting firms would likely be used to staff in-house and call center mods. After all, shoring up a workforce of this size so quickly would be difficult to impossible without using a global labor pool to provide the cultural and linguistic competencies needed for a worldwide audience at the bargain-rate prices that would appeal to the firms.

In the past few years, the public's distrust and questioning of the impact of commercial social media platforms on all aspects of their lives have grown. This has come in the wake of many scandals like those I described earlier, but also in light of the unexpected political triumph of Donald Trump in the United States in 2016, and of the Brexit campaign in the United Kingdom in the same year. As scholars and analysts continue to pick apart the mechanisms by which these political turns were influenced by online disinformation campaigns (such as the Cambridge Analytica scandal of 2018, as one example), the public and, increasingly, legislators have begun to ask more sophisticated questions about how their social media ecosystem is made. In any such discussion, I argue, accounting for the internet's for-pay gatekeepers, commercial content moderators, must take place.

To that end, I have responded to numerous requests for media interviews, appeared as a guest on radio and television programs, and lectured extensively on commercial content moderation workers and their work. Far from the cloistered and lonely research of my early years, I have found a number of communities also committed to defining and drawing attention to these practices, sometimes for different ends, but all in overlapping ways. These include individuals involved in civil society advocacy; those concerned with internet freedom issues; people committed to freedom of expression and human rights, writ large; people concerned about worker well-being; legal scholars interested in internet jurisdictional, governance, privacy, and procedural issues; scholars who look at the future and the nature of work; researchers committed to the fostering of a healthier internet for all—the list goes on. I am indebted to these colleagues, whom I consider a community of practice for my own research and, indeed, the members of a nascent sort of content moderation studies.

Evidence of this coalescence can be found in many places, including in my convening of a first-of-its-kind open-to-the-public conference at UCLA in December 2017 that saw the participation of some one hundred people—academics, activists, students, journalists, and professional commercial content moderation workers among them—engaged in various ways with content moderation in its many forms. The participants and presenters at this conference, titled "All Things in Moderation," included the United Nations special rapporteur on the promotion and protection of the right to freedom of opinion and expression David Kaye; attorney Rebecca Roe, who represents a commercial content moderation worker suing his employer for disability after ten years on the job; a panel of journalists who cover the commercial content moderation beat in some way; and perhaps most powerfully, Roz Bowden and Rochelle LaPlante, two women who have worked in the past and currently work as commercial content moderators, respectively.[5]

Since then, several other conferences and events focusing on content moderation policy and practice have taken place, including the "Content Moderation at Scale" event at Santa Clara University in February 2018, one in Washington, D.C., in May 2018, and another in New York City in late 2018.[6] Numerous conversations about commercial content moderation are taking place at academic and tech conferences and congresses held throughout the world, with tangible policy outcomes just beginning to be the result. I anticipate much more to come.

Indeed, one industry insider who requested anonymity told me in the summer of 2017 that his firm considered commercial content moderation "a one-billion-dollar problem." I then knew we were only at the beginning of a long-term and major conversation.

Today, the stakes for social media firms are much higher than ever before, and online life can and does frequently have real-world offline impacts that are a matter of life and death. As just one bleak and disturbing example, the ongoing discrimination and violence toward the Rohingya minority of Burma (also known as Myanmar) has been stoked by online hate campaigns primarily conducted on Facebook. These online expressions of ethnic feuding have resulted in violent mob killings, leading many of the Rohingya to flee into exile in neighboring countries like Bangladesh, and may have been organized by the Myanmar government in a concerted effort of media manipulation to the end of consolidation of political power. Can commercial content moderation, undertaken on behalf of the world's largest social media firms, be an effective bulwark against such manipulation, exploitation, and propaganda with deadly outcomes, or is it simply part of a larger enterprise—a social media industry based on user-generated content at all costs—that ultimately facilitates it? As just one example, it may well be that the power and allure of using such channels unprecedented in their scope and ability to be exploited is simply too irresistible for state actors, among other parties eager to orchestrate and engage the platforms' capabilities to undisclosed propagandistic and other potentially even more nefarious ends.[7]

Facebook's head of Global Policy Management, Monika Bickert, was unusually candid about the challenges facing her company's platform when she spoke at the conference held at Santa Clara University in February 2018, as reported by Alexis

Madrigal in the *Atlantic*. Madrigal noted that "the content moderation challenge is different from the many competitive bouts and platform shifts that the company has proven able to overcome. It's not a primarily technical challenge that can be solved by throwing legions of engineers at the problem." Indeed, despite Facebook's vast wealth in both monetary and technical terms, Bickert acknowledged that many of the problems facing Facebook's content teams are unlikely to be quickly or easily solved in the near term. And as for artificial intelligence, Madrigal reported Bickert's unequivocal response. "That's a question we get asked a lot: When is AI going to save us all? We're a long way from that."[8]

What this means in the near- and even mid-term is that companies that require social media gatekeeping of user uploads will continue to turn to human beings to fulfill that need. This means that the global workforce of people who perform social media moderation as a professional, for-pay task as either part-time or full-time work will increase. And just as it has at Google and Facebook, the workforce will increase across all the sectors where commercial content moderation takes place, from the boutique in the cloud to the digital piecework microtasks on Amazon Mechanical Turk. Researchers, engineers, and social media firms themselves will assuredly continue to develop artificial intelligence tools to ease the burden of moderation and to contend with the volume of content uploaded. The widescale adoption of PhotoDNA, an automated tool that uses algorithms to find and remove child sexual exploitation material that has been recirculated on social media sites, across the social media industry portends the use of such automation that could be applied in other cases.[9] But such material must already be included in a database of identified bad content in order for it to be successfully removed by automated means.

Further, identifying what constitutes child sexual exploitation material may be a relatively straightforward, albeit grim, process, but the issue becomes a thorny and complex problem when applied to other types of automated moderation and removal. Consider the eGlyph project, based on the same technology of hashing algorithms as PhotoDNA, but targeting "terroristic" content instead.[10] Who will program these tools to seek out terroristic material? By whose definition? And how will users ever know when they are being deployed and on what terms? To be sure, an algorithm is much less likely to be responsive to meaningful or routine oversight, much less to leak to the media or be interviewed by academic researchers, than a human moderator. Indeed, this may be an incentive not lost on the firms and policy makers that seek to replace human screeners with artificial intelligence–based computational tools. To be sure, without the willingness of human moderators to talk to me, in violation of nondisclosure agreements and employment contracts in almost every case, this book could not have been written.

Ultimately, even the significant advancements in artificial intelligence automation fall short of current demand and remain primarily aspirational. One obvious solution might seem to be to limit the amount of user-generated content being solicited by social media platforms and others, but this is an option that no one ever seems to seriously consider, a point noted by Dartmouth computer scientist and PhotoDNA creator Hany Farid in his paper on the tool's uptake.[11] The content is just too valuable a commodity to the platforms; it is the bait that lures users in and keeps them coming back for updates to scroll, new pictures or videos to view, new posts to read, and new advertisements to be served.

So commercial content moderation will therefore continue to fall to legions of human beings worldwide. They will be asked to make increasingly sophisticated decisions and will often be asked to do so under challenging productivity metrics that demand speed, accuracy, and resiliency of spirit, even when confronted by some of humanity's worst expressions of itself. As Max Breen said so poignantly and concisely of the work he and tens of thousands of others do: "It's permanently damaging." The effects of that damage can be even more powerful when workers report an inability to sufficiently separate the responsibilities of their jobs from their time off the clock, whether it was sufficiently divorcing their sense of protecting users from seeing or experiencing harm, or the phenomenon of something disturbing from their workday invading their psyche when at home.

At this time, there are no publicly available short-term or longitudinal studies concerning the effects of commercial content moderation on the workers who do it. It is possible that some companies that rely on these practices have done internal psychological assessments or have otherwise tracked their employees' mental and physical health and well-being, but if they have been completed, the studies are assuredly highly guarded. Without this information, the development of effective wellness and resilience plans for commercial content moderation workers across the industry by qualified mental health and other professionals is difficult. That said, some members of the tech and social media industries have banded together into the benignly named "Technology Coalition," ostensibly to focus on these issues.[12] The firms listed as members of the Coalition as of late 2017 were Adobe, Apple, Dropbox, Facebook, GoDaddy, Google, Kik, LinkedIn, Microsoft, Oath, PayPal, Snap, Twitter, and Yahoo. As I wrote for *Techdirt* in early 2018:

> Several major industry leaders have come together to form the self-funded "Technology Coalition,"

> whose major project relates to fighting child sexual exploitation online. In addition to this key work, they have produced the "Employee Resilience Guidebook," now in a second version, intended to support workers who are exposed to child sexual exploitation material. It includes information on mandatory reporting and legal obligations (mostly US-focused) around the encountering of said material, but also provides important information about how to support employees who can be reasonably expected to contend emotionally with the impact of their exposure. Key to the recommendations is beginning the process of building a resilient employee at the point of hiring. It also draws heavily from information from the National Center for Missing and Exploited Children (NCMEC), whose expertise in this area is built upon years of working with and supporting law enforcement personnel and their own staff.[13]

The *Employee Resilience Guidebook* is a start toward the formation of industry-wide best practices, but in its current implementation it focuses narrowly on the specifics of child sexual exploitation material and does not appear to be intended to serve the broader needs of a generalist commercial content moderation worker and the range of material for which he or she may need support.[14]

Unlike members of law enforcement, who can invoke and lean on their own professional identities and social capital for much-needed support from their peers, family members, and communities, moderators often lack this layer of social structure and indeed are often unable to discuss the nature of their work due to nondisclosure agreements, professional isolation, and stigma around the work they do. The relative geographic diffusion and industrial stratification of commercial content moderation work can also make it difficult for workers to find community with one another, outside of their immediate local teams, and no contracting or subcontracting firm is represented in the current makeup of the Technology Coalition, yet legions of commercial content moderation workers are employed through these channels.

In another industry move toward harm reduction in professional moderation work, YouTube CEO Susan Wojcicki announced at the South by Southwest (SXSW) festival in 2018 that her platform's content moderators would be limited to four hours per day of disturbing or potentially harmful material going forward.[15] But without a concomitant reduction in the amount of content YouTube receives—currently some four hundred hours per minute, per day—it was unclear how YouTube would avoid needing to double its workforce of screeners in order to contend with the volume. It was also unclear whether the four-hour figure was arbitrary or based on knowledge about thresholds of what its workers could reasonably bear without suffering burnout or other ill effects of their jobs—and what, if anything, would be done for the workers already exposed to much more than that.

Left to its own devices, it seems unlikely that the social media industry will decide to self-regulate to the benefit of the commercial content moderators it relies on but, until very recently, had not even admitted it employed. Using a variety of distancing tactics, both geographic and organizational, to put space between the platforms and the workers, social media firms have been largely immune to being held responsible for damage that some workers allege was caused by their time spent screening user-generated content. The way responsibility for user-generated content is treated in U.S. law is also a factor with regard to thinking about social media platforms being held to account for the content they circulate and ask their moderators to screen. In the United States, Section 230 of the Communications Decency Act of 1996 is the doctrine that holds social media companies largely immune from liability for material they disseminate on their networks, platforms, properties, and sites.[16] This does not mean that social media platforms do not have a vested interest in controlling the content that appears on their branded sites, apps, and platforms; on the contrary, that desire for control is the reason commercial content moderation exists. But it is largely not a legal standard to which the companies have, until recently, moderated. By many accounts, the immunity provided by Section 230 is what allowed the internet industry to flourish and grow into the economic and social force that it is today. Proponents of this argument contend that it gives social media platforms discretion to allow or disallow such content based on norms that they decided, meaning that they could set the rules for the user uploads they solicited and then distributed. But when Section 230 was enacted, the notion of four hundred hours of video content per minute per day being uploaded to the internet as a whole, much less to just one commercial site, was largely beyond the reach of most people's imagination, if not simply just bandwidth and computational power.

As times and stakes have changed, Section 230 does not seem as unassailable as it has in the past. And as social media firms have taken their business global, they now find themselves responsible to governments and legal regimes not just in the United States but worldwide, many of which are demanding that their jurisdictions and sovereignties be acknowledged and their local laws obeyed. As such, some of the greatest challenges to the primacy of Section 230 and the legal immunity of social media firms for the content they solicit, monetize, and disseminate are coming not from the United States but from the European Union as a whole, as well as some of its individual member states. The recent German "network enforcement law," abbreviated as NetzDG in German, is one of the most stringent examples. It demands that any social media platform with two million users or more operating in Germany must adhere to German law regarding hate speech and content and must remove any questionable content within twenty-four hours of a complaint being made or risk massive fines of up to 50 million euros. In Germany, this restriction is particularly focused on Nazi glorification, imagery, or other similar material, but it also focuses on "hate speech" as constituted under German law more broadly. Love or hate Germany's demand, one thing is clear: social media firms have responded by hiring more commercial content moderators, such as contractors at the Berlin call center Arvato, to bear the burden.[17]

Beyond questions of liability for content, there are questions, too, of other types of legal liability, such as harm to employees. In a landmark case, two Microsoft employees from Washington filed suit in that state's court in December 2016, alleging permanent disability and post-traumatic stress disorder, or PTSD, due to the material—much of it child sexual exploitation content—they were required to view and remove on behalf of the company.[18] Unlike many of the situations described in this book, one factor that makes this case unique is that the two plaintiffs, Henry Soto and Greg Blauert, were full-time direct employees of Microsoft, not contractors, assuredly related to the fact that Microsoft famously lost a lawsuit almost twenty years ago brought by so-called contractors who were, they successfully argued, de facto full-time permanent employees being denied benefits.[19]

Because many employee-employer disputes are now settled out of court through arbitration proceedings that are typically subject to secrecy clauses, it is unknown if other professional moderators have alleged similar harm and reached a settlement or agreement in their favor. For this reason, this case, which continues to make its way through the Washington state civil court system as of this writing, will be an important one to watch, as precedent is key to successful litigation in the U.S. legal system. As but one interested observer, I have been shocked that Microsoft has not opted to quickly and quietly compensate these workers in order to end the litigation process and the public's ability to follow along. Indeed, other former commercial content moderators have followed suit and are pursuing legal means to address what they allege are the harms caused by doing their jobs. The latest instance has been the filing of a lawsuit in California against Facebook by former in-house contractor Selena Scola in September 2018. Unlike the lawsuit against Microsoft, this one has been filed as a class-action suit.[20]

The opportunity for commercial content moderators themselves to organize and advocate for improvements in working conditions has not yet coalesced. The difficulty of such a campaign is, of course, heightened by the global dispersal of the workers and the varied legal, cultural, and socioeconomic norms of their contexts. In the United States, 2018 is not a high point for organized labor across the board, with the ascendency of the anti-worker Donald Trump to the presidency and the control of Congress by anti-union Republicans (and not a few hostile Democrats). One cannot also reasonably expect that Facebook, Caleris, Upwork/oDesk, or any other company involved in commercial content moderation would willingly ease the way for workers to organize and push back against damaging work conditions and unfair labor arrangements; this, after all, is one of the purposes of outsourcing and using intermediaries to supply the labor pool in the first place, along with taking advantage of the lack of regulation and oversight that these arrangements also offer.

Nevertheless, the rich labor history in the United States and around the globe could certainly inform worker organizing, whether via traditional labor unions or sector-by-sector (as in BPOs), by geography, or by type of work. In the Philippines, the BPO Industry Employees Network, or BIEN Pilipinas, is one such group endeavoring to bring together employees in the BPO sector into an organized, strong community of workers advocating for better conditions and better pay; this advocacy would include the commercial content moderation workers who perform their labor from Manila whom we met earlier. In the United States, newer labor upstarts such as the Tech Workers Coalition are focusing on labor organizing in the tech hotbeds of Seattle and San Francisco, sites where content moderators work as professionals in the social media and tech industries in large numbers.[21] A continued challenge to any organization of commercial content moderators will be in identifying who and where in the world the workers are in the fractured strata of moderation worksites, and if organizers will be able to make their case before the firms employing commercial content moderators do enough to improve the conditions of the job, on the one hand, or move contracts elsewhere in the world to avoid such organizing and demands for worker well-being.

Other types of activism and civil society intervention, too, can play a role in improving the work lives and working conditions for professional moderators. Civil society activists and academics have pushed for transparency in social media user content removal, such as in onlinecensorship.org, a project developed and led by the Electronic Frontier Foundation's Jillian York and academic and advocate Sarah Myers West, designed to give users access to tools to document takedowns of their online content. Myers West published an important study on the outcomes of the project, noting that, in the absence of explanations, users develop folk theories about the reasons behind the takedown of their content.[22] Is this the outcome social media companies were hoping for when they cloistered commercial content moderation workers and practices in the shadows?

Just as civil society advocates, academics, and policy specialists have come together to put pressure on the social media industry to be more transparent about the nature of user content takedowns (most notably via the Santa Clara Principles, drafted at the conference there on content moderation in early 2018), so too could they publicly demand increased transparency for working conditions, rates of pay and benefits, and support offered to the moderators responsible for those takedowns.[23] Inspired by the efforts of workers, advocates, and academics who together have demanded accountability from companies like Amazon and Google and their pursuit of technology development in the service of war or of the police state, a similar movement to bring support and justice to the commercial content moderators of the world would be an obvious next step.[24] I look forward to working with interested parties on just such an endeavor, with the workers themselves, and their needs, at the fore. Ultimately, I am reminded of the words of Rochelle LaPlante, a professional content moderator on Amazon Mechanical Turk and who served as a plenary speaker at "All Things in Moderation" in 2017. When asked by an audience member what was the one thing she would like done to improve her quality of life as a moderator and those of others like her, she said simply, “Pay us.” Of course, LaPlante was calling for improved wages for the essential work that she does. But it was clear that her words had a deeper meaning: “Value us. Value the work we do on your behalf. Value our humanity.” Without visibility, this will be impossible. And if we wait for the tech and social media industries to take the lead, this visibility may never come. Given the list of problems and criticisms facing the major mainstream platforms and their firms, justice for commercial content moderators may figure low on the list of priorities.

In fact, the creators of the products, platforms, and protocols that are the context for commercial content moderation, and explain its existence, may not be those best equipped to solve the problems it has created for workers and users—the public—alike. And we must not cede our collective imagination to this sector alone. Indeed, there are other vital institutions that have been neglected while the public has sated its desire for information online. I am talking, of course, about libraries. In the course of my work as a university instructor, I have had the pleasure of teaching and learning from students preparing to be librarians and other kinds of information professionals at four universities in the United States and Canada. These are bright, highly educated people with an orientation toward public service and assisting people with their information needs. As the internet has ceded its space to more and more sites of corporatized control and to models of information sharing that are fundamentally driven by a profit motive before all other values, libraries have remained largely more transparent, more open, and more responsible to the public. Media scholar Shannon Mattern has remarked on the untapped potential of librarians and libraries to serve the public in the age of digital information. Says Mattern: "Online and off, we need to create and defend [these] vital spaces of information exchange, and we need to strengthen the local governments and institutions that shape the public use of those spaces. The future of American democracy depends on it. . . . And we cannot depend on tech companies to safeguard those information spaces."[25] Mattern proposes a return to libraries and the visible, tangible human expert intermediaries who work in them to help us navigate the challenging and overwhelming information environment online and off.

Meanwhile, the story of commercial content moderation is being used as a powerful entry point for endeavors that put the nature of the contemporary internet into larger question. One such project, a documentary film called *The Cleaners,* premiered at the Sundance Film Festival in Park City, Utah, in January 2018, directed by Moritz Riesewieck and Hans Block. I served as adviser for the film and participated on panels with the directors. The film covered a good deal of territory I had set out in my research and that is covered in this book as well, including a focus on the work lives of Filipino commercial content moderation workers, but expands outward to question the political impact and social cost the millions of decisions these workers make has in aggregate.

Despite knowing the terrain intimately, I was shocked by the impact the film had on me personally, moving me, in some moments, to tears. It was a good reminder of how important it is to continue to tell the incredibly complicated story of the practices of commercial content moderation and those who undertake it. It was also a reminder to me that our work continues to have an impact, as I watched audiences around the world affected by the film and the stories of the workers it featured. And it was a reminder of the fact that I have an obligation to the workers who have shared their stories with me over the years, and allowed me to piece together the silhouette of commercial content moderation from invisibility, reverse engineering from their own firsthand accounts by corroborating with other sources and uncovering practices otherwise unknown, were it not for their willingness to speak to me. After this book, the research of other academics, the investigative journalism of so many reporters, and artistic interventions like *The Cleaners,* the truth is that we can no longer claim that commercial content moderation is "hidden." But what has to change is the status it is afforded, and the conditions of labor for its workers.

My understanding of what goes on behind the social media screen, in the form of commercial content moderation, has impacted my own engagement with social media and with all the digital platforms that make up my life, in both work and leisure. It is my hope that, in unveiling the presence of these previously unknown intermediaries—whose commercial content moderation work is fundamental and indispensable to their employers but also for all of us as users of the platforms for which they toil—we might ask who else exists behind the scenes of the digital media landscape. I believe firmly that our own understanding of the breadth of human traces in this landscape is woefully limited, and that it therefore must be expanded if we are to truly be able to weigh the impact of our own escapism into digital platforms characterized by their fun, inviting, accessible, and always-on affordances yet that give little honest appraisal of their true costs.

It is considered bad form, generally speaking, to leave readers with a list of rhetorical questions, yet one important outcome for me, as a researcher, is to formulate these questions, such that they may guide my future research and be the many threads that I follow as I trace commercial content moderation and its workers around the globe, and in and out of the digital and physical spaces they inhabit. But here is another truth: we are all implicated—those of us, anyway, who log on to Facebook, upload to YouTube, comment on a news item, up-vote or down-vote a post. Our desire—our human desire—to feel engaged and connected has created the very niche that commercial content moderation workers fill, the need to which they respond. I often remember Josh Santos, who so acutely diagnosed the problem of suicidal ideation on MegaTech as one that was undoubtedly, incurably self-perpetuating. If they build it—the platforms, those empty vessels ready to fill up and rebrand and disseminate digitally around the globe—we will come. We will fill them—with our user-generated content, our preferences, our behaviors, our demographics, and our desires. We are often even more directly connected to commercial content moderation than that; after all, it is a user who flags a video that begins the commercial content moderation cycle of review at MegaTech. It is a user report on Facebook that sends a post or an image through the circuits described in Chapter 2. And it is user-generated content that is the subject of the reviews. Unless and until we unplug or otherwise force a renegotiation of our own relationship to the platforms, we, as users, are perhaps the most vital piece of all in the social media cycle of production as generators of the content and as its insatiable consumers. Commercial content moderation workers, who make the platforms bearable, tolerable, and fun, are our unseen partners in a relationship of symbiosis, the yin to our yang, balancing, curating, and working on making our online activity feel pleasurable, like leisure, or like something to which we want to return. Yet the up-votes, the flagging, the video sharing: our participation is an illusion of volition in an ever shrinking, ever compartmentalized series of enclosures, governed by community guidelines and Terms of Service, with human interventions hidden away by nondisclosure agreements, and human traces erased as soon as they appear, anomalous errors, in the system. This book, small in the scheme of human things though it is, hopes to bring those human traces to the fore and render them visible, not so far removed from the finger on the book scan.

Is it possible for working conditions to improve for commercial content moderation workers? Until computational power and computer vision make exponential strides, it is a task that, for the foreseeable future, demands human intervention. Even then, human labor, particularly as it follows globalization circuits to sites of large and ever cheaper labor pools, will likely still be preferred. The series of decisions that come into play that a commercial content moderation worker must make about each piece of user-generated content that he or she deals with is sophisticated beyond the scope of any algorithm or filter. The cultural nuances and linguistic specificities only add to the challenge. That unparalleled supercomputer of the human brain, and its massive data banks of cultural knowledge and life experience coupled with its own sophisticated onboard meaning-making software of our minds, will still be preferable to any machine in terms of cost and capability. Social media platforms fueled by user-generated content also do not show any signs of disappearing; the proliferation of mobile computing devices and more people in the world having access to them suggests, indeed, the very opposite. Nor does it seem likely that human nature will change such that the jobs that Josh and Max and Melinda, Sofia and Drake and Rick, and people like them have done will just disappear. And so the need for commercial content moderation will continue. People willing to take on a job that provides little status, is often shrouded by contractual preclusions to even acknowledging its existence, exposes workers to abhorrent and disturbing aspects of humanity, and leads to almost assured burnout will still be needed. I am thankful to all of the commercial content moderation workers for the job they do, as I am grateful that it is not I who have to do it. Out of the shadows, out from behind the screen, and into the light.

23. Jan Maghinay Padios, "Listening Between the Lines: Culture, Difference, and Immaterial Labor in the Philippine Call Center Industry" (Ph.D. dissertation, New York University, 2012), 17, 18, http://search.proquest.com.proxy1.lib.uwo.ca/docview/1038821783/abstract/317F9F8317834DA8PQ/1?accountid=15115.

24. Padios, *A Nation on the Line*.

25. For "space of flows," see Manuel Castells, "The Space of Flows," ch. 6 in *The Rise of the Network Society,* 2nd ed., 407–59 (Oxford: Blackwell, 2000).

26. Harvey, *A Brief History of Neoliberalism*.

27. Cecilia Uy-Tioco, "Overseas Filipino Workers and Text Messaging: Reinventing Transnational Mothering," *Continuum* 21, no. 2 (2007): 253–65, https://doi.org/10.1080/10304310701269081.

6
Digital Humanity

1. Sarah T. Roberts, "Social Media's Silent Filter," *The Atlantic,* March 8, 2017, https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796.

2. Adrian Chen, "The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed," *Wired,* October 23, 2014, http://www.wired.com/2014/10/content-moderation.

3. Olivia Solon, "Underpaid and Overburdened: The Life of a Facebook Moderator," *The Guardian,* May 25, 2017, http://www.theguardian.com/news/2017/may/25/facebook-moderator-underpaid-overburdened-extreme-content; Jamie Grierson, " 'No Grey Areas': Experts Urge Facebook to Change Moderation Policies," *The Guardian,* May 22, 2017, http://www.theguardian.com/news/2017/may/22/no-grey-areas-experts-urge-facebook-to-change-moderation-policies; Nick Hopkins, "Facebook Moderators: A Quick Guide to Their Job and Its Challenges," *The Guardian,* May 21, 2017, http://www.theguardian.com/news/2017/may/21/facebook-moderators-quick-guide-job-challenges; Julia Angwin and Hannes Grassegger, "Facebook's Secret Censorship Rules Protect White Men . . .," *ProPublica,* June 28, 2017, https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms; Ariana Tobin, Madeleine Varner, and Julia Angwin, "Facebook's Uneven Enforcement of Hate Speech Rules . . .," *ProPublica,* December 28, 2017, https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes. Catherine Buni and Soraya Chemaly, "The Secret Rules of the Internet," *The Verge,* April 13, 2016, https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech; Till Krause and Hannes Grassegger, "Inside

Facebook," *Süddeutsche Zeitung,* December 15, 2016, http://international.sueddeutsche.de/post/154513473995/inside-facebook.

4. April Glaser, "Want a Terrible Job? Facebook and Google May Be Hiring," *Slate,* January 18, 2018, https://slate.com/technology/2018/01/facebook-and-google-are-building-an-army-of-content-moderators-for-2018.html.

5. The website for *All Things in Moderation,* held at UCLA on December 6–7, 2017, includes links to the full schedule, guest posts by participants and others, and links to videos of some of the plenaries and keynotes. It is accessible at https://atm-ucla2017.net.

6. See the first COMO event's website at http://law.scu.edu/event/content-moderation-removal-at-scale.

7. Betsy Woodruff, "Exclusive: Facebook Silences Rohingya Reports of Ethnic Cleansing," *Daily Beast,* September 18, 2017, https://www.thedailybeast.com/exclusive-rohingya-activists-say-facebook-silences-them; Paul Mozur, "A Genocide Incited on Facebook, with Posts from Myanmar's Military," *New York Times,* October 18, 2018, https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

8. Alexis C. Madrigal, "Inside Facebook's Fast-Growing Content-Moderation Effort," *The Atlantic,* February 7, 2018, https://www.theatlantic.com/technology/archive/2018/02/what-facebook-told-insiders-about-how-it-moderates-posts/552632.

9. Hany Farid, "Reining in Online Abuses," *Technology & Innovation* 19, no. 3 (2018): 593–99, https://doi.org/10.21300/19.3.2018.593.

10. "How CEP's EGLYPH Technology Works," Counter Extremism Project, December 8, 2016, https://www.counterextremism.com/video/how-ceps-eglyph-technology-works.

11. Farid, "Reining in Online Abuses."

12. Technology Coalition, "The Technology Coalition—Fighting Child Sexual Exploitation Online," 2017, http://www.technologycoalition.org.

13. Sarah T. Roberts, "Commercial Content Moderation and Worker Wellness: Challenges & Opportunities," *Techdirt,* February 8, 2018, https://www.techdirt.com/articles/20180206/10435939168/commercial-content-moderation-worker-wellness-challenges-opportunities.shtml.

14. "Employee Resilience Guidebook for Handling Child Sexual Abuse Images," Technology Coalition, January 2015, http://www.technologycoalition.org/wp-content/uploads/2015/01/TechnologyCoalitionEmployeeResilienceGuidebookV2January2015.pdf.

15. Nick Statt, "YouTube Limits Moderators to Viewing Four Hours of Disturbing Content per Day," *The Verge,* March 13, 2018, https://www.theverge.com/2018/3/13/17117554/youtube-content-moderators-limit-four-hours-sxsw.

16. "CDA 230: Legislative History," Electronic Frontier Foundation, September 18, 2012, https://www.eff.org/issues/cda230/legislative-history.

17. Ben Knight, "Germany Implements New Internet Hate Speech Crackdown," *DW.COM,* January 1, 2018, http://www.dw.com/en/germany-implements-new-internet-hate-speech-crackdown/a-41991590.

18. Greg Hadley, "Forced to Watch Child Porn for Their Job, Microsoft Employees Developed PTSD, They Say," *McClatchy DC,* January 11, 2017, http://www.mcclatchydc.com/news/nation-world/national/article125953194.html.

19. Steven Greenhouse, "Temp Workers at Microsoft Win Lawsuit," *New York Times,* December 13, 2000, https://www.nytimes.com/2000/12/13/business/technology-temp-workers-at-microsoft-win-lawsuit.html.

20. Timothy B. Lee, "Ex-Facebook Moderator Sues Facebook over Exposure to Disturbing Images," *Ars Technica,* September 26, 2018, https://arstechnica.com/tech-policy/2018/09/ex-facebook-moderator-sues-facebook-over-exposure-to-disturbing-images. The text of the lawsuit as filed is available here: https://www.documentcloud.org/documents/4936519-09-21-18-Scolav-Facebook-Complaint.html.

21. "About BIEN," *BIEN Philippines* (blog), February 5, 2018, http://www.bienphilippines.com/about; and "Tech Workers Coalition," https://techworkerscoalition.org.

22. Sarah Myers West, "Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms," *New Media & Society,* May 8, 2018.

23. "Santa Clara Principles on Transparency and Accountability in Content Moderation," Santa Clara Principles, https://santaclaraprinciples.org/images/scp-og.png.

24. Scott Shane and Daisuke Wakabayashi, " 'The Business of War': Google Employees Protest Work for the Pentagon," *New York Times,* July 30, 2018, https://www.nytimes.com/2018/04/04/technology/google-letter-ceopentagon-project.html; Kate Conger, "Amazon Workers Protest Rekognition Face Recognition Contracts for Police," *Gizmodo* (blog), June 21, 2018, https://gizmodo.com/amazon-workers-demand-jeff-bezos-cancel-face-recognitio-1827037509.

25. Shannon Mattern, "Public In/Formation," *Places Journal,* November 15, 2016, https://doi.org/10.22269/161115.