https://www.thetimes.co.uk/my-articles/my-life-as-a-social-media-moderator-zb5sgqwtm

# My life as a social media moderator

Rachel Holdsworth spent eight years as a moderator sifting the net's ocean of vile posts. She tells Stephen Bleach how a ceaseless diet of racism, sexism and violence has changed her

You'd like Rachel Holdsworth. She's easy to get along with: intelligent, articulate, straightforward. Perhaps a little guarded, though. As we talk at a cafe in Clerkenwell, central London, she's chatty enough, but keeps a certain distance; her occasional smiles are brief and cautious. But then, considering what she's been doing for the past eight years, it's a wonder she smiles at all.

"I couldn't handle it any more. I couldn't take the drip, drip, drip banality of hatred that I dealt with every day. It's affected my ability to trust, because I know there are people in the world who think like this. I've seen things I can never unsee, I've read things I can't forget: I don't know who these people are but I know they're out there."

Rachel was a moderator — one of the invisible army that sifts through the billions of online posts, videos and comments that fill the internet every week. Working for social media platforms, news outlets and corporations, their job is to assess and, if necessary, delete material that may breach the organisation's rules. It is a business that involves engaging daily with hundreds of instances of bullying, racism, sexism, pornography and graphic violence.

It is also a business that is booming. Facebook alone has trebled its number of moderators from 4,500 to about 15,000 in the past 18 months. YouTube, Google, Twitter and Instagram each employ thousands of moderators worldwide. News outlets such as the BBC, The New York Times and, indeed, The Times and The Sunday Times have their own moderating teams; and large firms with a social media presence, such as BT and PlayStation, employ moderators to police their own platforms and Facebook pages.

Such has been the explosive growth of social media in recent years, moderation has struggled to keep up. By 2016, 400 hours of video were being uploaded to YouTube every minute; Twitter users were sending 21m tweets an hour; and Facebook had been used by more than a billion people in a single day.

The sheer volume of material made it difficult to track harmful content. As questions were asked in Congress and parliament about a series of scandals — the use of Facebook and Twitter by a St Petersburg-based "troll farm" to attempt to influence the US election; of YouTube by Isis propagandists; of Instagram by those who fed 14-year-old Molly Russell images that glamorised suicide — the platforms realised they had to police their content more effectively.

Hence the boom in the moderator business. While artificial intelligence can help to flag up problematic content, this is a job that ultimately has to be done by humans — lots of them. And amid all the demands that social media platforms clean up their act, one question isn't being asked: what does this do to the people doing the cleaning?

“This stuff is in my head. I’m never going to be able to get rid of the knowledge of so much hate,” says Rachel.

“I’m never going to forget that I had to delete 50 comments saying, and I quote, ‘Kill it with fire’, under a sweet, inoffensive video of a trans woman talking about her life. All of us who have worked as moderators are stuck with the general, incessant horror of it. It will last us for ever.”

Rachel is 41. She specialised in moderating news and current affairs. Originally from Yorkshire and now living in London, she is not, I’d say, an oversensitive person, but last summer she decided she could take no more and quit the moderating business, concerned at what the conversations she had to immerse herself in were doing to her. “I couldn’t bear it. It’s so dark. Just look at the posts under any news story that touches on race, gender or sexual orientation and you’ll see it.”

(I did this after meeting Rachel, searching Facebook and YouTube for references to the Isis bride “Shamima Begum”. Within five minutes I had read “Lethal injection . . . now”; “If you see her . . . beat her”; “This was brought by Jews. They were banned from Britain for 600 years for a reason”; “Muslims got one thing right tho, don’t allow women to vote, they’re just not mentally capable”; and, under a photo of a man being beheaded, “If they come back we will be waiting.” This is the stuff that hasn’t yet been removed by moderators — or, more worryingly, has been judged acceptable.)

Rachel started moderating in 2010 as a way of earning extra cash — the web start-up that was her day job paid peanuts. Like many moderators, she worked from home. She put in between 10 and 25 hours a week, being paid £8.50 an hour by an agency that provided moderation services for some of the world’s biggest media brands. We can’t specify the agency or its clients — all household names in the UK — because, as is standard practice in the moderation game, Rachel signed a non-disclosure agreement as part of her contract. The industry argues this is necessary to protect its staff and customers’ data; it has the side effect of making it difficult to expose the companies’ real moderation practices.

Rachel would meet up periodically with her fellow moderators. “They were in their twenties to forties, mostly female — a lot of new mums do it to earn some extra cash. Almost all white, straight, cis gendered — I don’t think you could do this job if you were black or gay. It would be too upsetting.”

At first, the job was fine. “It felt like we were doing something important, performing a useful service. We were able to take down the worst of the abuse and the extremist stuff, and there wasn’t too much of it.

“Three or four years ago it started to change. The extreme posts started to become the norm. It became impossible to keep on top of it. And the overall tone got nastier. People in comment threads got much quicker to scream at each other, to refuse to listen, to immediately assume bad faith, to insult. They would say they wanted someone to get raped or to die.”

Along with the rise of hate, more and more video links were posted on comment threads. Moderators were expected to check each of these videos and take down the link if it transgressed policies. It involved watching things that Rachel can’t forget.

“There was lots of 9/11 conspiracy theory stuff, and videos on Pizzagate [a false story about a Democratic Party paedophile ring in Washington, which led one user to open fire on the ring’s notional headquarters in a pizza restaurant]. That sort of thing didn’t bother me too much, apart from making me sad that some people believed them.

“But some still bother me. You remember the coup in Turkey? I’m pretty sure I saw a video of a guy getting decapitated by a tank. It was shot at night, dark, but I’ve watched a lot of videos and that one was real.

“Then there were the Rohingya in Burma. I was moderating UK pages but whenever any major news organisation posted anything on the Rohingya, loads of Sri Lankan accounts would pile in and post propaganda, with YouTube links. They claimed the Rohingya set fire to their own villages and the genocide was faked.” (The UN later issued a report stating that Facebook had played a “determining role” in stirring up hatred against Rohingyas.)

“One video showed 10 or 15 people standing in a ditch, with a bulldozer behind them. I didn’t take it down because there was no violence shown. It wasn’t until a few days later I read a report that the army were making people stand in ditches and then executing them. I’d probably been looking at the precursor to a massacre. I felt awful. I still do.”

Shocking as many of the videos were, the sheer unpleasantness of the developing online conversation depressed Rachel more. “It’s the normalisation of nastiness that saps your faith in ordinary people, that affects your world view. It’s knowing that any story about something simple and nice is another excuse for people to be vile. A story about women’s football will be leapt on by misogynists, threatening to rape the players. A story about a charity project in India will be leapt on by people from Pakistan, saying Indians are rapists.

“I’ve worked on support group sites for people with illnesses or disabilities. Even there you’ll see cliques and bullying straight away. You’ll see people coming on to the site to take advantage of the vulnerable. Social media seems to enable the worst side of us. I couldn’t take it any more. I had to get out.”

Rachel’s experience is mirrored by moderators worldwide. Last week the technology news website The Verge published an investigation into working conditions at Cognizant, an Arizona-based agency that provides moderating services for Facebook. Employees spoke of having to watch videos of drug cartel murders, of men having sex with farm animals, of people being stabbed to death. That last one was part of the training process.

Cognizant’s moderators work in an office rather than from home. Pay is modest — $28,800 (£22,000) a year, compared with an average Facebook employee’s $240,000 (£181,000) — and their breaks and work rate are rigorously monitored by management, but at least there’s the chance for some everyday camaraderie. That can get out of hand: staff have been found having sex in stairwells and the breastfeeding room in what one employee called “trauma bonding”. Others smoke cannabis in their short break times: the work is easier to face when stoned.

Another coping mechanism was black humour, employees said. They would compete to send the most racist memes to their colleagues and share jokes about self-harm and suicide (most had watched videos of real suicides). One spoke of becoming worried about his mental health, saying:

“We were doing something that was darkening our soul.” The Verge alleged that a number of people had developed symptoms similar to post-traumatic stress disorder (PTSD) after leaving Cognizant.

Cognizant pointed out that it provides counsellors and a healthcare plan, and that its pay is 20% above the local minimum wage. Nevertheless, other Facebook subcontractors face the issue of stress: last year a moderator working with the company Pro Unlimited sued for what she alleged was work-related PTSD. The case is continuing.

Not all moderators are doing it for a living. Reddit, a news aggregator site, is moderated by volunteers. Users can message moderators and sometimes find out their usernames — which they then use to harass those who delete their posts. One moderator from London, Emily, told the news site Engadget: “I’ve had rape threats, I’ve had death threats. I’ve had people send images of animal abuse, child pornography. They want to disturb you because they’re angry. People are willing to dedicate their time to make you feel threatened for the hell of it.”

One of the messages sent to Emily read: “i want you to get cancer because i wanna see your own body killing you. i want you to get cancer, your mother, sister, brother, father, grandparents i want them all to get cancer . . .”

Another moderator, Robert, received the message: “i will f****** murder you in your London flat when you walk back from your work at [Robert’s real workplace]. you might want to take this shit seriously for once.”

Reddit said: “Harassment and persistent abuse toward moderators are not acceptable,” adding that it was working to tackle the problem.

Whether it’s coping with personal threats or just living with what Rachel calls “the torrent of hate”, being a moderator changes people. “Research indicates there can be a cumulative effect to being exposed to so much negative material,” says Lucy Bowes, associate professor of experimental psychology at Magdalen College, Oxford.

“Being exposed repeatedly to violence, or hateful messages, is related to increasing levels of poor mental health — depression, low self-esteem and anxiety in particular.

“Some may be more vulnerable than others, but even among people who appear to be ‘immune’, it is possible that being exposed to this sort of information may skew their perception of normality. There is a lack of research on the long-term harms of acting as a moderator, which really needs to be addressed.”

Rachel stirs her hot chocolate and stares out of the cafe window. Ordinary-looking people hurry past along Clerkenwell Road, under a grey London sky. She observes them for a moment.

“Trusting people is tricky,” she says. “I could never do online dating, for instance. If they’re colleagues or friends of friends, it’s OK, but strangers — no, I’m finding that’s really difficult. Because somebody’s doing it, somebody’s putting all this stuff online. And those people have friends and families and children like everyone else. So who are they?”

I ask her how we can fix this, how we can make the internet a civilised space again. She looks a little sad. “I honestly don’t know, short of shutting it all down. People need to remember that the

ones who are reading this stuff are human beings. What they say can hurt. But it’s hard to empathise through a keyboard.”

Rachel has not completely abandoned social media. She still uses Twitter, but these days most of her posts are from her ginger tomcat, Parkin. They concern chasing mice, snuggling by a radiator or upsetting the recycling bin. Parkin’s posts receive scores of “likes” and dozens of comments. All are positive, or playful, or funny. No one hopes Parkin will get cancer, or be raped, or murdered.

“Cat Twitter is the nicest, kindest corner of the internet,” says Rachel. And she looks out of the window again.