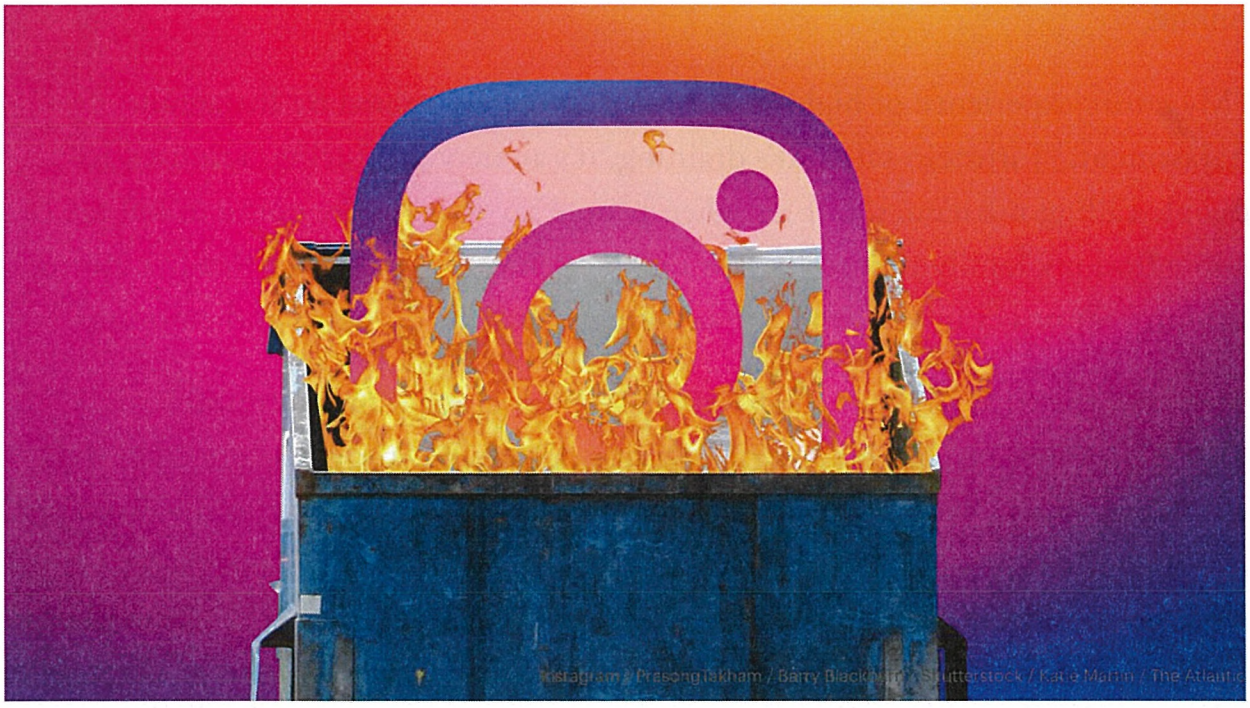
EXHIBIT 150

The Atlantic

Instagram Has a Massive Harassment Problem

The platform has cast itself as the internet's kindest place. But users argue harassment is rampant, and employees say efforts to stem it aren't funded well or prioritized.



Story by Taylor Lorenz | OCTOBER 15, 2018 | TECHNOLOGY

Like *The Atlantic*? Subscribe to The Atlantic Daily, our free weekday email newsletter.

Enter your email

Sign up

depobook.com
EXHIBIT
META-
Castaneda-3

WHEN Brandon Farbstein first joined Instagram in 2014, he was 14 and optimistic. Farbstein was born with a rare form of dwarfism, and he wanted to use the photo-sharing site to educate people about his condition—to, as he told me, "show people a glimpse into my life and inspire people."

Soon enough, though, the hateful messages started coming: death threats, expletive-laden comments about his appearance, worse. A meme page put his face on Hitler's body. Multiple accounts popped up with the explicit purpose of taunting him. His house was swatted. When he does a live video, the insults float onscreen, fast and furious. "It's been hard to keep my composure," Farbstein told me. After trolls started posting pictures of him in the hallways at his high school, he started to fear for his safety. Eventually, he left and finished high school online.

"My entire experience of high school was completely ruined by Instagram harassment," Farbstein said. "It's draining, it's anxiety producing. I'm used to people calling me names, but it's when people say that they're going to kill me or come find my family that really gets me in a sense of pure terror. Really nothing can prevent or get in the way of that taking over your thoughts and emotions."

Farbstein has tried to make the harassment stop. He said he's filed numerous reports through Instagram's internal reporting tool, but the company takes days to address them, if it does at all. Most of the time he simply deletes the messages and comments himself. "The reporting system is almost like it's not there sometimes," Farbstein said. "You want it to end, but you also know that nothing is going to happen if it takes months and months for your report to go through. It produces more fear and anxiety ... than whatever's posted."

The harassment, he said, has escalated sharply over the past year. "Instagram is the No. 1 platform that I experience hate on," he said.



He's not alone. Despite a long-standing and well-crafted reputation for being the nicest place on the internet, to many of its users—a large number of whom are very young—Instagram doesn't feel very kind at all. To some, it's getting worse. In interviews, 22 users described painful, sustained, sometimes terrifying abuse on the platform—abuse they say Instagram has repeatedly failed to stem.

For years, Instagram has traded on its reputation as a place for positive, aspirational content—for shopping, connecting with friends, and following interests. Even as other major social platforms, such as Twitter, Facebook, Reddit, and YouTube, have been forced through massive, public reckonings with harassment, Instagram has emerged largely unscathed. Often it's heralded as a good example for the rest of the social web.

In 2016, Instagram's founder and then-CEO, Kevin Systrom, touted his plans to turn it into what *Wired* editor in chief Nicholas Thompson described as "a kind of social media utopia: the nicest darn place online" in a long and rosy feature about the effort. A few months later, after Systrom took to the company's blog to announce new anti-harassment measures that included auto-banning inappropriate comments, *HuffPost* lauded the company for "tackling hate speech the way Twitter should have done."

Instagram doubled down on its warm and fuzzy image last fall with a multipronged "Commitment to Kindness," which included various product tweaks as well as a #KindComments campaign, in which celebrities like Jessica Alba encouraged users to leave nice feedback on one another's photos. "I've seen how other companies have misstepped in managing communities," Systrom told

Variety at the time. He didn't specifically name Twitter, Reddit, or Facebook (Instagram's parent company), but his implication was clear: *We're different, and we're better.*



Shutterstock / Thanh Do / The Atlantic

"Our goal is to be the safest platform online," Karina Newton, Instagram's head of public policy, told me in September when I asked about harassment. "It's an investment that's not just in words." She highlighted as an example the company's recent "Kindness prom," where teen influencers ate free In-N-Out burgers and danced the evening away under a balloon arch that declared the space a BULLY FREE ZONE. Last week, Instagram announced a set of new features to limit bullying and "spread kindness," including comment filters on live videos, a "kindness camera effect to spread positivity," and the deployment of machine-learning technology to better detect bullying in photos.

When approached for comment, a Facebook spokesperson referred me to Instagram's communications team. Instagram declined to speak on the record about many of the particulars of its anti-harassment efforts.

"We want people to come to Instagram and have a positive experience—to make friends, find interests, and do all the things that make Instagram such a positive place," Newton wrote in an October statement to *The Atlantic*. "Bullying and

harassment are completely counter to the experience we work to create. We want to stop this behavior, and we want people to feel safe on Instagram, but we know we have a lot more to do."

But in interviews, three current or former Instagram employees told *The Atlantic* that they do not believe the company has done enough to protect users from large-scale harassment, and that projects that would seem to tackle the issue are understaffed and unprioritized.

"There's an effort called 'kindness,' which is to reduce bullying and harassment, but there's not that many people working on it," said Alex, a current Instagram employee who asked to be referred to by a gender-neutral pseudonym. "Generally, what you'll find is a lot of these efforts on harassment or bullying, or there's a new feature to track how much time you spend—they're mostly done for PR." Another Instagram employee told me nearly the same thing: that Instagram's anti-bullying rhetoric "doesn't seem connected to what's actually going on in the company."

Users say harassment on Instagram can come from nearly any direction, or nearly any reason. For Riley, a 14-year-old who asked to be referred to by a pseudonym, it was on an account devoted to American Girl dolls; after she posted a pro-LGBT hashtag, trolls found her phone number and called it nonstop for days, threatening to find her house "and do horrible things" to her, she said. The harassment became so overwhelming that she deleted her account.

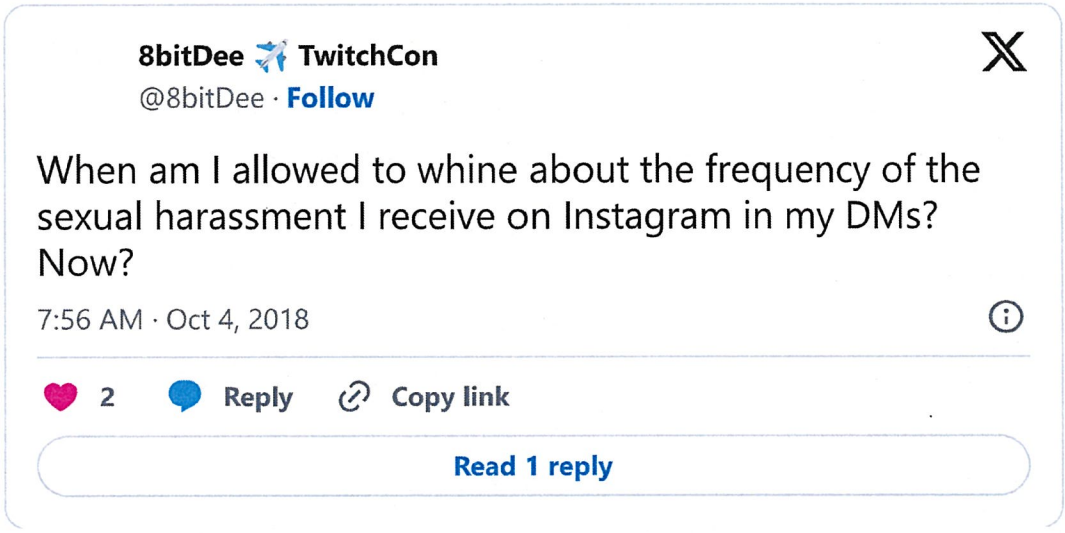
[Read: The teens who post the same thing every day on Instagram]

Sarah, a Montreal-based woman who runs the feminist Instagram account @douconsideryourselfafeminist, said, "There's not a day that goes by without death threads, rape threats, insults." She said she flags the harassment, but "most of the time you get a report saying they don't violate anything."

Violet Paley, an actress who accused James Franco of sexual misconduct early this year, said that since then, she's been relentlessly targeted on Instagram. "I get tagged in Stories, like 'this dumb bitch,' calling me a whore," she said. She showed me screenshots: One Instagram user offered $1,000 for her home address, promising to "teach these lying whores a lesson." Another is devoted to Photoshopping her in offensive and degrading positions. Yet another posted her

address and phone number; the same user "started messaging all my friends and acquaintances, saying, 'I'm going to kill Violet,'" Paley said. At one point, the situation got so bad that she reached out to the FBI.

Paley said that aside from the time someone posted her address, Instagram has ignored the reports she's filed against her harassers. "When I report things, I think they just go into somewhere and they ignore it," she said. "Nothing ever happens."



Even lifestyle Instagrammers, long considered the platform's bread and butter, have begun questioning their place on it as a result of rampant harassment. In June, fashion-and-beauty Instagrammer Suzanne Jackson, who has more than 237,000 followers, spoke out, saying that she and other influencers are "no longer ignoring" the abuse they receive on the platform. Jeanette Johnson, a fashion blogger, told me that she has experienced harassment nonstop since joining Instagram several years ago, but that it's gotten worse in the past year. Recently, someone posted her home address in the comment section of a photo as an apparent threat. Johnson deleted the comment and reported the man to Instagram, but his account is still active. Johnson told me she no longer feels safe in her home.

Even those with a less glamorous Instagram presence aren't immune to attacks. "I'm a 42-year-old athlete, and I get harassed all the time. It's outrageous, but it's pervasive," said C. C. Rowe, a triathlete with just over 1,000 followers. "All of my friends have gotten it, and I would say Instagram is the worst."

SEVERAL THINGS make Instagram a uniquely fertile breeding ground for harassment. First of all, it's huge: More than 1 billion people use the platform every month. That's much larger than the other default-public social network, Twitter, which has 335 million monthly active users, and which has unrolled sophisticated anti-harassment measures amid a sustained wave of bad press. But Instagram's privacy settings are less comprehensive and less granular than Facebook's, the other social network with more than a billion users. On Instagram, your profile and all its content are either public to the world or limited to your approved followers; many people, especially those looking to grow their personal brand or small business, say they feel compelled to remain public because there's no middle ground. In other words, Instagram is large and public enough to invite harassment, but unregulated enough to let it fester.

The platform is also a powerful discovery engine: On Instagram, it's easy to search by hashtag or location and pull up thousands of people's profiles and public images, and it's simple for anyone who wants to mobilize an army to encourage trolls to pile on a specific person by tagging them in an image or story. These trolls are often marshaled via Instagram's robust and sometimes ferocious fan culture, in which celebrities and dedicated fan accounts reach scales much greater than on other platforms. A single Justin Bieber fan account, for instance, has more than 1.6 million followers, and celebrities such as Selena Gomez and Ariana Grande have several fan accounts with more than 100,000 followers each.

One wrong comment and these fans can rally by the millions to attack. "When you're hanging out on Instagram, it's easy to feel part of a big, friendly, happy group," said Zoe Fraade-Blanar, author of the book *Superfandom*. "But being part of an angry mob with those same people is also a lot of fun, and that's why you get these huge uprisings."

Not found

> Hey @instagram, I saw what the picture was of, and I saw the message telling my wife to choke to death on it. If that's not a

— Daniel Kibblesmith ☃ (@kibblesmith) October 3, 2018

Last summer, Skye, an English 14-year-old, posted something about drama between the bands Fifth Harmony and 5 Seconds of Summer on the fan account she ran for the latter. Her post got picked up by other accounts, and within hours she was receiving hundreds of cruel comments. She tried to report each one to Instagram, she said, but for every account she blocked or reported, more would crop up.

[Read: Why Instagram questions became so annoying]

Timothy Heller saw the wrath of Instagram fans after she accused her former friend, the pop star and former *Voice* contestant Melanie Martinez, of sexual assault in December 2017. Martinez's fans came after her on every major social platform—but Heller said it's on Instagram that the harassment has had the most lasting damage on her career, even almost a year later. "I model for brands sometimes, and these people will go to the brand Instagram account and mass comment on the pic of me," she said. She now sometimes warns people she works with that they should prepare to be trolled.

Shutterstock / Paul Spella / The Atlantic

In June, the actress Kelly Marie Tran of *Star Wars: The Last Jedi* deleted all her Instagram posts after months of relentless harassment on the platform. The following month, Pete Davidson quit because he was being harassed by fans of his then-fiancée, Ariana Grande, who herself temporarily left the platform as a result of harassment. Also this summer, the *Titans* star Anna Diop quit the platform because of out-of-control trolling. The actress Daisy Ridley quit for a time back in 2016 for the same reason, as did Justin Bieber. In September, Khloé Kardashian was forced to restrict permissions on her account after a flood of racist comments about her daughter, who is six months old.



Like Twitter, Instagram enables the easy setup of endless anonymous accounts: All you need is an email address, and you can start posting within minutes. Abusers leverage this functionality to create armies of fake accounts to attack people. But while Twitter now allows users to protect themselves—by muting replies from people who don't follow you, whom you don't follow, who aren't verified, who haven't confirmed an email address, and more—Instagram has implemented only some of these controls. For instance, on Instagram, you can hide comments from people you don't follow or who don't follow you—but you can't hide them from people who have a default avatar or haven't confirmed their email address, two hallmarks of burner accounts.

Earlier this month, the former *Bachelor* contestant Ashley Spivey was issued a stream of vivid death threats on Facebook and Instagram by fans of fellow

Instagram star @GirlWithNoJob after the two had a dustup on the latter platform. Spivey went through the standard reporting process on Instagram and heard nothing, so she begged for help on her Instagram Story. She said a Facebook employee who follows her reached out, offering to help. Spivey doesn't know whether the Facebook employee had a direct impact, but the harassment on Facebook finally abated. On Instagram, it's still going strong.

[Read: Teens Are being bullied 'constantly' on Instagram]

"A long time ago, I stopped being as active on Twitter because I thought that was the place where I would get a lot of harassment," Spivey said. "But lately I get way more harassment on Instagram than I ever did on Twitter."

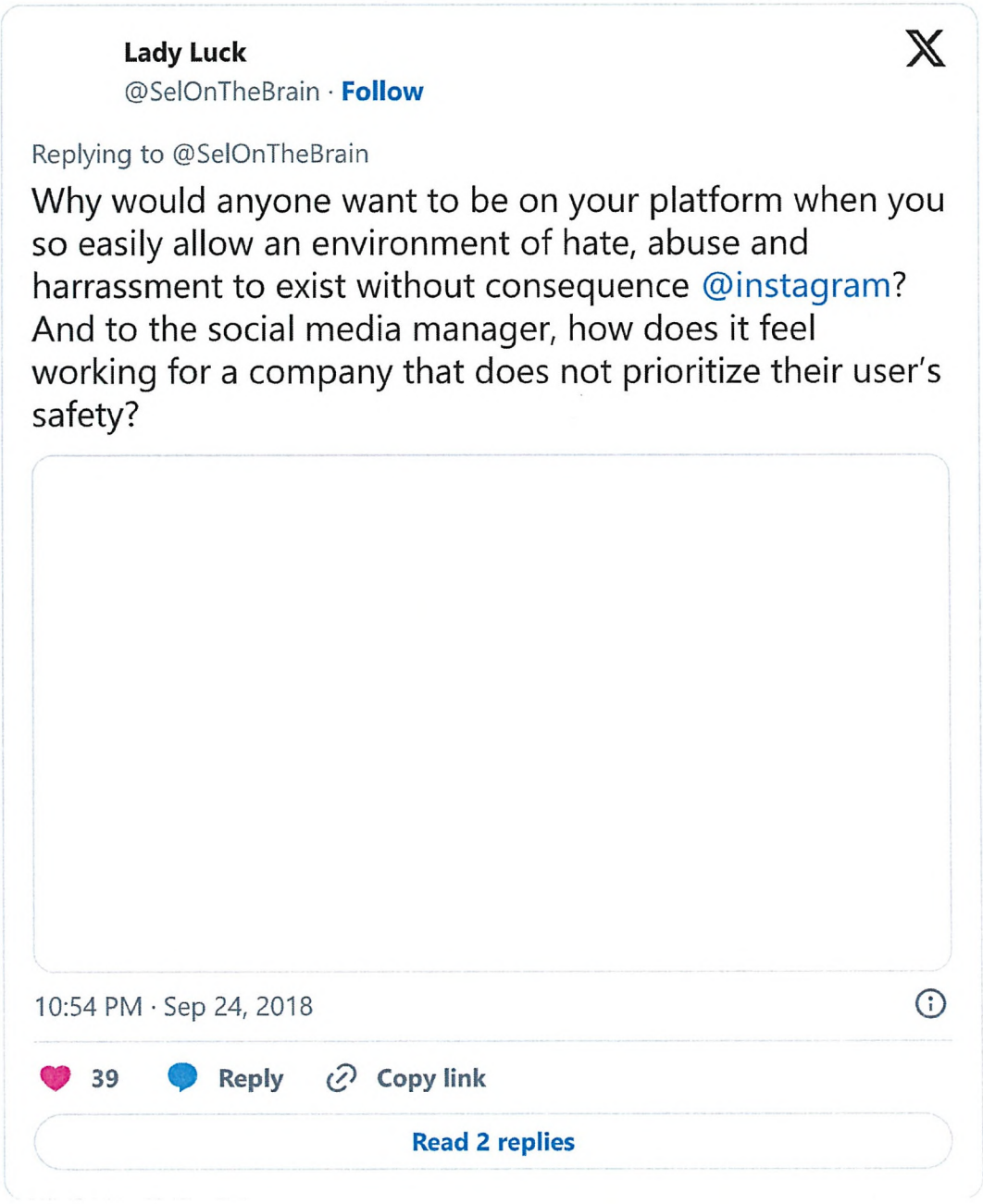
Users like Spivey who try to call out harassment problems say it's often their attempts at raising the issue that are punished rather than the original offender.



Late last year, Joanie Diana Goss, a physiotherapist who uses her account to document her workouts, started posting screenshots of some of the awful messages she receives on Instagram—messages that she said Instagram has "never once" addressed when she asked. Her account was then penalized for sharing the

offensive content in the messages, while the accounts that originally sent them remain active.

While Instagram does auto-filter certain words from comments, users say trolls simply add an extra letter or symbol to escape the filters. "People always find new ways to spell terrible things," said Katie, a plus-size fashion Instagrammer with hundreds of thousands of followers who asked to be referred to by a pseudonym because she fears retaliation from both Instagram and her harassers. She said trolls have called her employer, attempted to sabotage her day job, threatened to mutilate her body, and more.



Sara Mills, a beauty-focused Instagrammer with more than half a million followers, tried using comment filters to silence certain terms—but blocking terms doesn't block the person who used them. "I feel like it just reinforces the behavior," said Mills. "Instagram should say, 'These words are blocked, you can't do this.' With all the psychology they put into making you spend more time on the app, or share more photos, or figuring out what color to make a button, you'd think they could put some of that into figuring out how to get people to treat each other like human beings."

Mills said the type of micro-celebrity that Instagram has so famously created are particularly vulnerable. "I always hear from people, 'Well, you put yourself out there. You have to deal with this.' Let me tell you, I don't make millions of dollars like a celebrity. I don't have a team to protect me. I don't have an admin running my page. I'm not four steps removed," she said. "In these people's mind, they're talking to a picture of someone. They don't realize that in the influencer world, most of us manage our own accounts."

The most upsetting thing, Mills said, is that she feels like the company—which is expected to make some $8 billion in revenue this year selling advertising against the kind of content she and people like her produce for free—has given her no recourse. "You can't reach out to Instagram—there's no reaching out," she said. "Even as someone that's verified, I've never been able to get an actual person to help me."

Shutterstock / Thanh Do / The Atlantic

Some users have even found more creative ways to beg for help. "@Kevin I have a pretty big file of insta screenshots of a woman bullying me and my coworkers and threatening to come after me, but your report bots keep coming back to me saying we're cool with this," user @RubiStudios wrote in a comment on one of Systrom's own Instagram photos. Systrom did not respond.

ACCORDING TO an Instagram spokesperson, Facebook and Instagram share a team of 20,000 people working on safety and security across both platforms. Of that team, 7,500 people—a mix of contractors, full-time employees, and staff from "partner companies"—are tasked with reviewing content from the more than 1 billion people who use Instagram, and 2 billion who use Facebook, every month around the world. It's a Sisyphean task, according to the people who do it.

"The people who I work with, who make $15 an hour, have to have expertise on regulation, fraud, bullying, hate speech," said Andy, who moderates content for Facebook and Instagram and who asked to use a pseudonym. Even he finds the reporting mechanism lacking: "I've used the Instagram app as a user just a little bit, and the couple times I've reported something, the experience has been totally frustrating," he said.

Andy said that the few times he's used Instagram, he's seen content that, as a moderator, he "knew should have been deleted." Indeed, it's not hard to find: I recently came across an account that was posting memes about "kill black and Jewish people" and "rape disabled women." It also posted multiple threats to carry out school shootings. The account had 6,400 followers, and all its posts had hundreds of likes. It was active for more than a year and had been reported to Instagram many times using its internal reporting tool, but wasn't taken down until I flagged it to Instagram's PR department.



> so does anything on instagram ever meet the guidelines for harassment or @instagram
>
> — manal (@cliffordoze) September 24, 2018

A September *Business Insider* investigation revealed that Instagram's new TV service recommended videos of potential child abuse showing graphic violence and genital mutilation. By the time they were removed, only after being flagged to Instagram's PR team, they had generated more than a million views. Four days later, *The Washington Post* reported that Instagram had failed to curb illegal drug sales on the platform, noting that illicit content had flourished there. Following a drug dealer, for instance, filled "up a person's feed with posts for drugs, suggesting other sellers to follow and introducing new hashtags, such as #xansforsale," *The Post* reported.

When it comes to moderation, context is often critical. Instagram is available in every country except China and North Korea; its users speak dozens of languages and interact with one another in ways specific to the nuances of innumerable cultures. They use regional slang, tell inside jokes, and communicate in code. But Instagram's current reporting pathway doesn't allow users to explain exactly why something is offensive, leaving moderators to guess.

[Read: Posting Instagram sponsored content is the new summer job.]

"There could be all sorts of things that the user understands that the moderator doesn't," Andy said. "So many of my co-workers are old, people who did not grow up thinking like anything like this would ever happen. They got hired because their résumé says, 'I have a Facebook account,' but you need a Ph.D. in 4chan slang sometimes, and stuff that's specific to Instagram, in order to understand what someone means when they post something. We just have no context about the stuff that we get related to harassment, and it makes it a lot harder to interpret who is attacking."

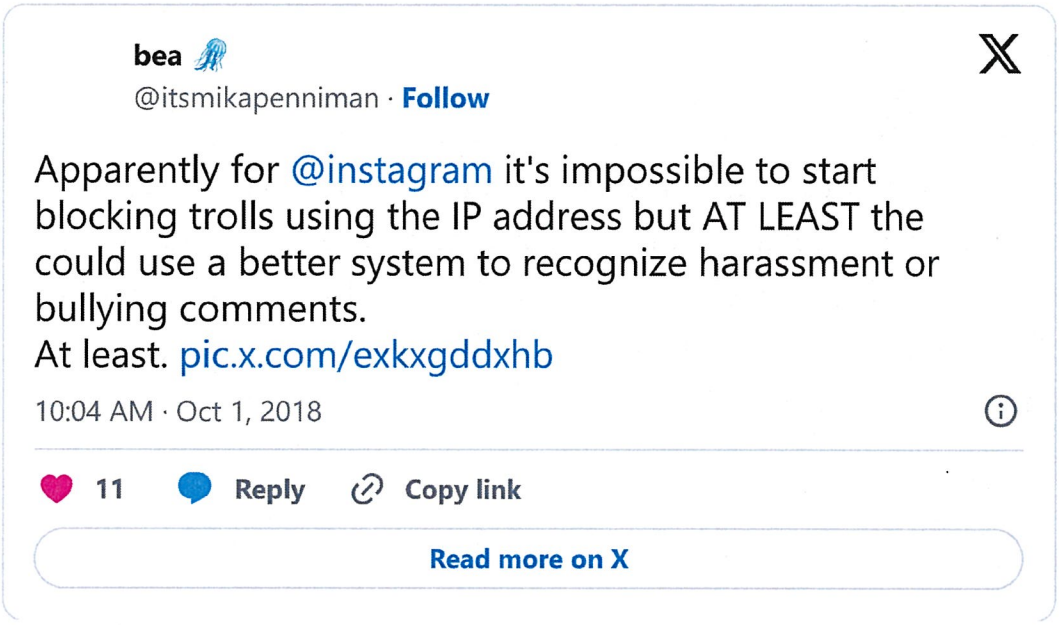
An increasing amount of content on Instagram is also just screenshots from Twitter, so moderators feel like they need to be experts in multiple social platforms to understand the nuances of each post, they told me. "We get it with YouTube, too—like the Logan Paul YouTube business or whatever is happening in the YouTube world, it obviously crosses over to Instagram, and we have to figure out what it means," Alex said.



> Just went to have a look at Megan's Instagram cos they mentioned it on Aftersun. The comments. 😢
> She is a human being, no one deserves this kind of misogynistic abuse. Social media needs to verify everyone & hold people to account for such hateful harassment! 😡 #LoveIsland pic.twitter.com/apY804ex72
>
> — Samantha Pressdee (@SammiePressdee) July 30, 2018

When Instagram introduces new features, the moderation-team members receive no warning, Andy said. Consequently, they are left scrambling to understand how they work and what constitutes harassment on each format. "When the Questions feature rolled out, same way as every other new feature, we had no idea," he said. "We didn't know which part is the question, which is the answer, who says what? That makes such a big difference on whether you're going to delete or ignore the post. The mods are just totally not kept up to date on how people use features."

Alex, the current Instagram employee who asked to be referred to by a pseudonym, said the company prioritizes growth above all else, often at costs to user experience. "The focus is still on getting people to spend more time, getting more users, getting more revenue. That doesn't change much internally," Alex said. "There's been a lot of effort to shape the narrative, but the reality is that it doesn't drive business impact."



At Instagram and Facebook, Alex said, "features can make whatever progress … but can't hurt the other metrics. A feature might decrease harassment 10 percent, but if it decreases users by 1 percent, that's not a trade-off that will fly. Internally right now, no one is willing to make that trade-off."

Allie, a former employee at Instagram, agreed. "Instagram has terrible tools. I think people haven't really focused on it much because so many harassment campaigns are just more visible on other platforms," she said. Throughout her time there, she said, "many of the efforts to reduce harassment were oriented toward PR, but very few engineering and community resources were put toward actually decreasing harassment."

She said that the company's focus on growth has crippled its ability to understand the deep problems within the communities the platform has created. "When you work in growth products for so long, you just kind of don't learn to understand community concerns," she said.

Even easy fixes were ignored during her time at the company, she said. "For almost a year, there was a link to a help page on the reporting flow that led to a broken page. It was like that for at least six months. How much do you have to drop the ball that on your harassment-reporting flow you have a broken page? It just shows how little resources and attention they have put toward it," she said.

Instagram has touted its automated moderation tools as a supplement to its human moderators. "We are working on AI technology that can find bullying or harassing content, which our teams can then review and remove," Instagram's Newton wrote in a statement. "Our aim is that in time we will be able to detect and remove this content faster than ever before so we can preserve the kind, safe environment our community deserves."

Earlier this year, for instance, Instagram announced it would be leveraging Facebook's machine-learning system, DeepText, to help moderate the platform. But Facebook has also struggled to stem harassment on its platform. It was Facebook's very own moderation system that led to women being locked out of their accounts or banned from posting on Facebook when they negatively commented about men in the wake of the #MeToo movement.

"Let's just say Facebook is totally screwing up their content-moderation policy on Facebook.com, and it's kind of sad that Instagram might suffer because of that," Andy said.

As Allie sees it, the issue is not that Instagram has anti-harassment features that are ready to be rolled out but haven't been; it's that no one is even given time or resources to develop these features in the first place.

"It's not even like, 'Oh, we decided we wanted to do a product, but we didn't," she said. "It's like, 'Oh, we didn't even have the resources to start.'"

For the time being, though, many users are bracing themselves every time they open the app. Others have left Instagram altogether, or taken their accounts private. Farbstein, now 19, is still public, and trying to stay positive. "For me it's really important to spread my message of empowerment and authenticity with the public," he said. "Especially to those who really need to hear it."

And instead of playing whack-a-mole with the abusive comments, he has a new strategy: He's leaving them up. "I want people to see the ugliness that exists in this world," he said. "It's gotten so easy to be a troll."