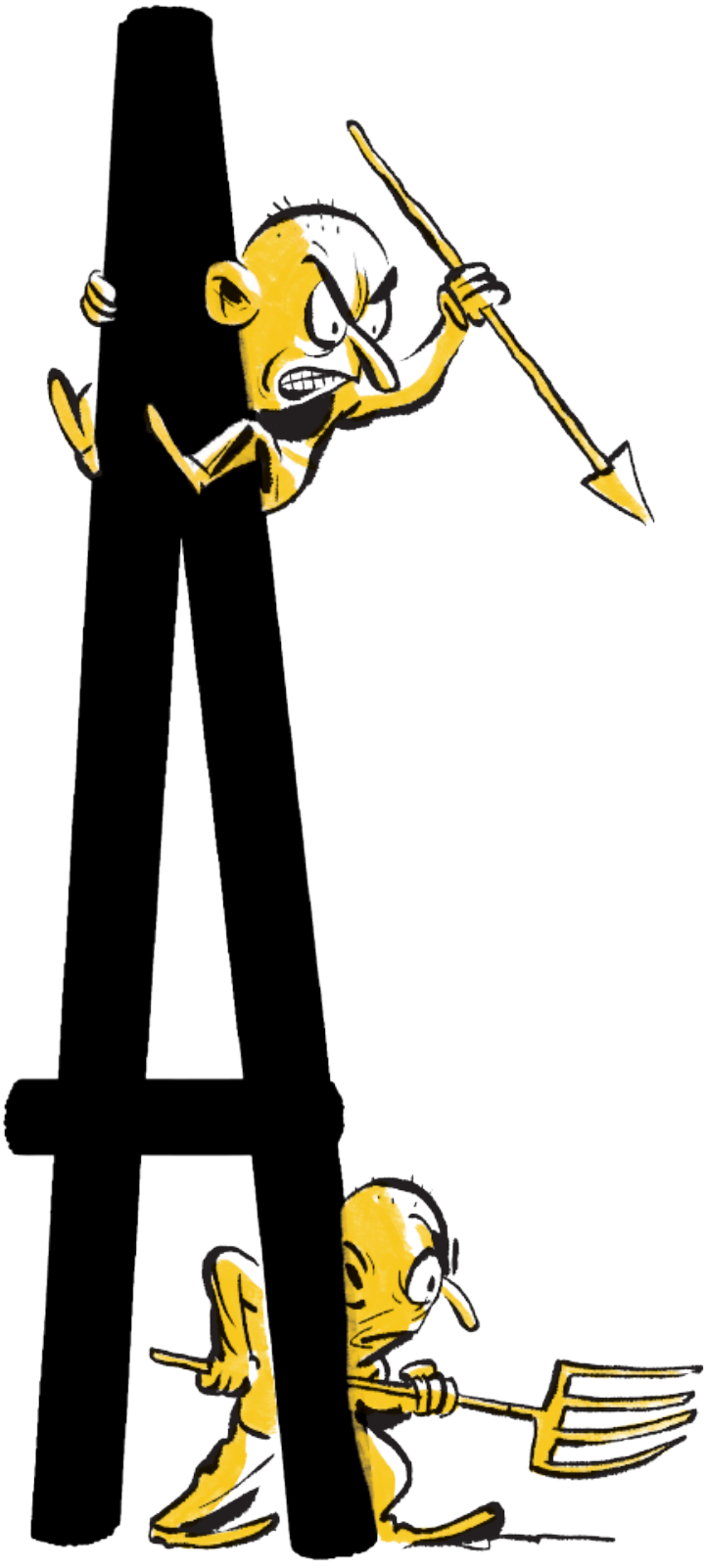
EXHIBIT 129

Instagram’s Kevin Systrom wants to clean up the &#%$@! internet.

by Nicholas Thompson | photographs by Art Streiber | illustration by Zohar Lazar | lettering by Braulio Amado

08.14.17

Kevin Systrom, the CEO of Instagram, was at Disneyland last June when he decided the internet was a cesspool that he had to clean up. His company was hosting a private event at the park as part of VidCon 2016, an annual gathering that attracts social media virtuosos, and Systrom was meeting with some Instagram stars. They were chatting and joking and posing for one another’s phone cameras. But the influencers were also upset. Instagram is supposed to be a place for self-expression and joy. Who wants to express themselves, though, if they’re going to be mocked, harassed, and shamed in the comments below a post? Instagram is a bit like Disneyland—if every now and then the seven dwarfs hollered at Snow White for looking fat.



After the chat, Systrom, who is 33, posted a Boomerang video of himself crouched among the celebrities. It's an ebullient shot of about 20 young people swaying, waving, bobbing, and smiling. In the lower right corner, a young woman bangs her knees together and waves her hand like she's beating eggs for a soufflé.

The comments on that post started out with a heart emoji, a "Hoooooo," and "So fun!" Soon, though, the thread, as so often happens online, turned rancid, with particular attention focused on the young woman in the lower right. "Don't

close  wait  just  wait  OPEN  them leg  baby," "cuck," "succ," "cuck," "Gimme ze suc." "Succ4succ." "Succme." "Go to the window and take a big L E A P out of it." A number of comments included watermelon emoji, which, depending on context, can be racist, sexist, or part of picnic planning. The newly resurgent alt-right proclaimed over and over again that "#memelivesmatter." There was a link in Arabic to a text page about economic opportunities in Dubai. Another user asked Systrom to follow him—"Follback @kevin." And a few brave people piped up to offer feedback on Instagram's recent shift to ordering posts by relevancy rather than recency: "BRING BACK THE CHRONOLOGICAL ORDER!"

Systrom is a tall, lean man with a modest bearing. His handshake is friendly, his demeanor calm. He's now a billionaire, but he doesn't seem to play the alpha male games of his peers. There is no yacht; there are no visits to the early primary states; there is no estranged former partner with an NDA. Systrom's personal Instagram feed is basically dogs, coffee, bikes, and grinning celebrities. A few years ago, Valleywag described his voice as "the stilted monotone of a man reading his own obituary," but he's become much smoother of late. If he has a failing, his critics say, it's that he's a sucker: He and his cofounder, Mike

Krieger, sold Instagram to Facebook too soon. They'd launched it a few years after graduating from Stanford, and it went into orbit immediately. They got $1 billion for it. Snap, which spurned an offer from Facebook, is now worth roughly $17 billion.

Related Stories

- **How Instagram Explore Became the Realest Place on the Web**

By David Pierce

- **Upworthy's Quest to Engineer Optimism for an Anxious Age**

By Zachary Karabell

- **Tristan Harris Wants to Rescue Our Minds From Our Phones**

By Nicholas Thompson

Systrom takes pride in this reputation for kindness and considers it a key part of Instagram's DNA. When the service launched in 2010, he and Krieger deleted hateful comments themselves. They even personally banned users in an effort Systrom called "pruning the trolls." He notes that Krieger "is always smiling and always kind," and he says he tries to model his behavior after that of his wife, "one of the nicest people you'll ever meet." Kevin Systrom really does want to be the sunny person on display in @kevin's feed.

So when Systrom returned from VidCon to Instagram's headquarters, in Menlo Park, he told his colleagues that they had a new mission. Instagram was going to become a kind of social media utopia: the nicest darn place online. The engineers needed to head to their whiteboards. The next image he posted on Instagram, just before Independence Day, was of some sumptuous homemade pretzels.

"Nice buns like yur mum," @streamlinedude commented. @Juliamezi added, "If you stop reading this you will die." She, or it, then added, oddly, "If u don't post this on 20 photos I will sleep with you forever."



Technology platforms, the conventional wisdom now goes, are not neutral. Their design and structure encourage certain behaviors, and then their algorithms control us even more. We may feel like we're paddling our own boats, but the platform is the river and the algorithms are the current.

As the CEO of a service with 700 million users, Systrom recognizes that he's something like the benevolent dictator of a country more than twice the size of the US. The choices he makes affect the lives of all his users—some of whom are insecure teens, some of whom are well-adjusted adults, some of whom are advertisers, and some of whom are pop singers dealing with an infestation of snakes.

In mid July 2016, just after VidCon, Systrom was faced with just such an ophiological scourge. Somehow, in the course of one week, Taylor Swift had lost internet fights with Calvin Harris, Katy Perry, and Kim Kardashian. Swift was accused of treacherous perfidy, and her feed quickly began to look like the Reptile Discovery Center at the National Zoo. Her posts were followed almost entirely by snake emoji: snakes piled on snakes, snakes arranged numerically, snakes alternating with pigs. And then, suddenly, the snakes started to vanish. Soon Swift's feed was back to the way she preferred it: filled with images of her and her beautiful friends in beautiful swimsuits, with commenters telling her how beautiful they all looked.

But Instagram can't build that world with simple technical fixes like automated snake emoji deletion.

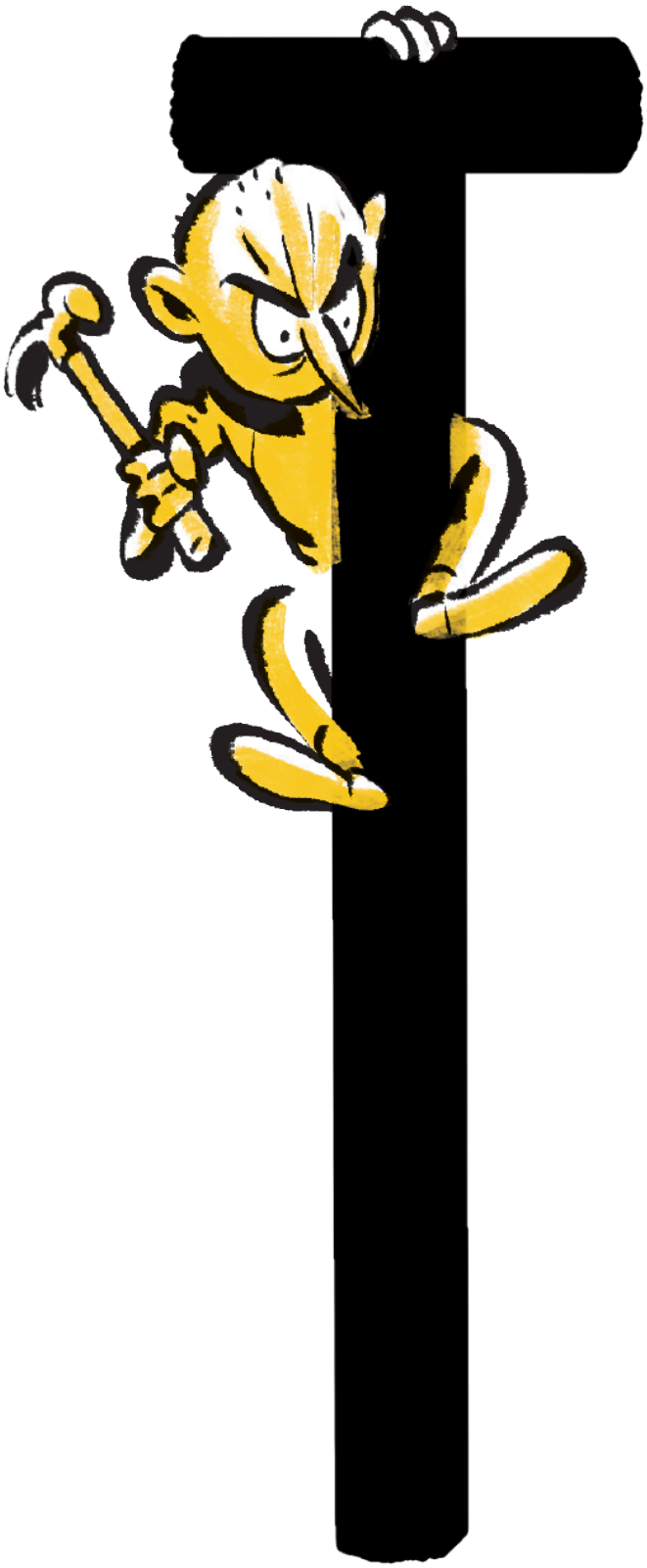
This was no accident. Over the previous weeks, Systrom and his team at Instagram had quietly built a filter that would automatically delete specific words and emoji from users' feeds. Swift's snakes became the first live test case. In September, Systrom announced the feature to the world. Users could click a button to "hide inappropriate comments," which would block a list of words the company had selected, including racial slurs and words like whore. They could also add custom keywords or even custom emoji, like, say, snakes.

The engineers at Instagram were just getting started. In October, the service launched a series of tools that roughly model what would happen if an empathetic high school guidance counselor hacked your phone. If you type in the word suicide, you'll be met first with a little box that says, "If you're going through something difficult, we'd like to help." The next screen offers support, including a number for a suicide-prevention hotline. Two months later, in December 2016, the company gave users the ability to simply turn off commenting for any given post. It's for those times when you want a monologue, not a conversation.

A cynic may note that these changes are as good for business as they are for the soul. Advertisers like spending money in places where people say positive things, and celebrities like places where they won't be mocked. Teenagers will make their accounts public if they feel safe, and if their parents don't tell them to get off their phones.

Still, talking to people at the company, from Systrom on down, you get the sense that this is a campaign felt in the heart, not just in the pocket. Nicky Jackson Colaco, Instagram's director of public policy, speaks of her own children and the many teenagers whose first experience with the swamp of social media is on Instagram. "I think what we're saying is, we want to be in a different world," she says.

But Instagram can't build that world with relatively simple technical fixes like automated snake emoji deletion. So, even amid a bevy of product launches last fall, Instagram's engineers began work on something much more complex.



Trying to sort rubbish from reason on the internet has long been a task for humans. But thanks to artificial intelligence, the machines are getting better at the job. Last June, around the time Systrom visited VidCon, Facebook announced that it had built a tool to help computers interpret language. The system, named DeepText, is based on a machine learning concept called word embeddings. When the system encounters a new word, it tries to deduce meaning from the other words around it. If a watermelon emoji is always surrounded by right-wing memes, that means something. The more data the classification engine analyzes, the smarter it gets. Like us, it learns over time; unlike us, it doesn't get exhausted or depressed reading the word cuck 72 times in a row.

One way to think of DeepText is that it's like the brain of an adult whose entire memory has been wiped and who will now devote themself to whatever linguistic task you assign. Facebook essentially has an icebox filled with these empty brains, which it gives to its engineering teams. Some are taught to recognize whether a Messenger user needs a taxi; others are taught to guide people selling bikes on Marketplace.



“What we’re saying is that we want to be in a different world,” says Nicky Jackson Colaco,

Instagram's public policy director.

After learning about DeepText, Systrom realized that his engineers could train it to fight spam on Instagram. First, though, like a child learning language, it would need some humans to teach it. So Systrom gathered a team to sort through massive piles of bilge, buffoonery, and low-grade extortion on the platform.

They labeled comments as spam or not spam and then fed everything into DeepText. The machines studied the categories and came up with rules to identify genuine economic offers in Dubai and whether it's a friend or a bot who wants a follow-back. Once DeepText was able to classify spam with sufficient accuracy, the engineers signaled the go-ahead, and the company quietly launched the product last October.

Then Systrom had an even more complicated idea: What if Instagram could use DeepText to knock out mean comments? Forget about the succs and the follbacks. Could the AI learn to filter out more ambiguous content? "Go to the window and take a big L E A P out of it" is definitely hostile, but it doesn't include any particularly hostile words. "Don't close  wait  just  wait  OPEN  them leg  baby" is gross. But can a computer tell? "Nice buns like yur mum" is rude and off topic. But it could be charming if it came from a childhood friend who truly appreciated your mother's German pretzels.

Other social media companies had worked to filter spam, but Instagram's new plan to make the whole platform kinder was vastly more ambitious. Systrom told his team to press ahead.



Instagram is a relatively small company. It has only about 500 employees—roughly one person for every 1.5 million active users. And the team that has trained the machines to be kind is tiny too. When I visited in late June, there were about 20 people, split equally among standing desks and sitting desks, surrounded by scattered boxes of sanitizing wipes. Everyone seemed young; the group seemed diverse. A woman in a head scarf sat near a white guy in a Tim Lincecum jersey. Their job is to pore through comments and determine if each one complies with Instagram's Community Guidelines, either specifically or, as a spokesperson for the company says, "in spirit." The guidelines, which Instagram first drafted in 2012, serve as something like a constitution for the social media platform, and

there's a relatively simple 1,200-word version available to the public. (In short: Always be respectful and please keep your clothes on.) The raters, though, have access to a much longer, secret set of guidelines, and they use it to determine what's naughty and what's nice. There are dozens of raters, all of whom are at least bilingual. They have analyzed more than 2 million comments, and each comment has been rated at least twice.



Nuance is crucial even when dealing with the most offensive words. "If you're using the N-word as a slur, then it's not allowed in our platform," says James Mitchell, Instagram's director of content operations, who manages the raters. "The exceptions are if you're using the word in a self-referential way or if you're recounting a story or an experience you had where you were discriminated against."

After the raters had sorted through the data, four-fifths of the text they classified was fed into DeepText. The machines studied all the comments, looking for patterns in the data classified as good versus the ones classified as bad. Eventually, Instagram's engineers, working with DeepText, came up with a set of rules for identifying negative comments, based on the content of the posts and other factors, such as the relationship between the author and the commenter. The company also uses a metric that the engineers, internally, call a "karma score," which captures the quality of the user's historic posts. The rules were then tested on the one-fifth of the data that hadn't been given to DeepText, to see how well the machines matched the humans' evaluations.

Humans are boiling stews of biases and contradictions, and computers don't have emotions. But machines are only as good as the rules built into them.

The machines give each comment a score between 0 and 1—say .41 or .89—which is a measure of Instagram's confidence that the comment is offensive or inappropriate. The

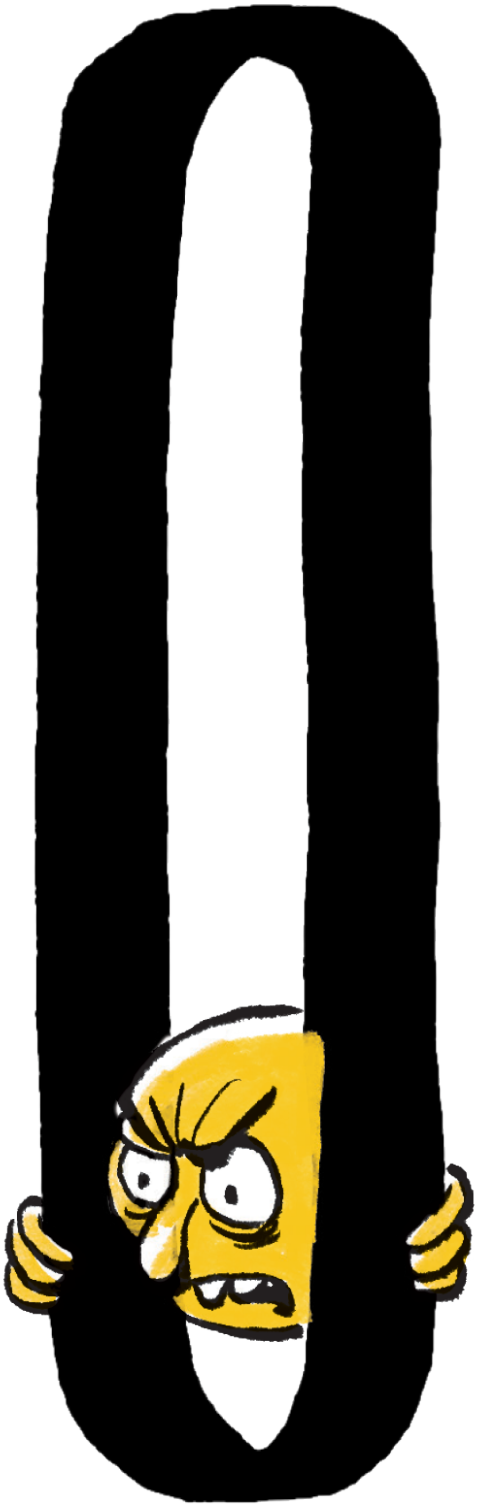
higher the score, the worse the comment, and above a certain threshold, the comment gets zapped. Instagram has tuned the system so that it has a false-positive rate of 1 percent, meaning that 1 percent of the comments deleted by the machines will be ones the humans would have waved along.

Sometimes using a machine can feel like the purest way to solve a problem. Humans are boiling stews of biases and contradictions, and computers don't have emotions. But machines are only as good as the rules built into them. Earlier this year, the chief scientist of the text analytics company Luminoso, Rob Speer, built an algorithm based on word embeddings to try to understand the sentiment of text posts. He applied the algorithm to restaurant reviews and found, oddly, that Mexican restaurants all seemed to do poorly. Stumped, he dug into the data. People like Mexican food, but the system suggested they didn't. Ultimately, he found the culprit: "The reason was that the system had learned the word 'Mexican' from reading the Web," he wrote. And on the internet, the word Mexican is often associated with the word illegal, which, to the algorithm, meant something bad.

What happens if genuine arguments and thoughtful criticism start to appear less frequently?

When I tell Systrom this story, he responds quickly, "That sounds horrible." He then points out that his ratings won't be based on such algorithms; they'll be based on the judgments of his human raters. But might the raters be biased in some way? What if the guy in the Tim Lincecum jersey is really sensitive to nasty things said to tall women but not to short ones? What if he thinks that #memelivesmatter is hateful? Or empowering? The cumulative biases of the people in that room, clicking away on their MacBooks, will shape the biases of a filter that will mediate the world for 700 million people.

The product debuted in late June, and so far user response seems reasonably positive. Few people, in fact, have noticed. The filter isn't perfect, though. It has trouble, for example, with words that mean different things across cultures. It stopped sentences that used the word fag in a way that clearly referred to the British slang for cigarettes. It stopped the sentence "You need to check your BMI index" when it came from a new account but not from a verified one. It also had trouble recognizing Kanye West lyrics. Every line in this sequence got banned when it was put through: "For my southside niggas that know me best / I feel like me and Taylor might still have sex / Why, I made that bitch famous." It was entirely at ease, however, with more creative Kanye insults like "You left your fridge open / somebody just took a sandwich."



One big risk for Instagram is that the filter will slowly change the tenor of the platform. Instagram is mostly pictures, of course, but what happens if genuine arguments and thoughtful criticism start to appear less frequently? The best ideas don't always come from being positive and friendly. Instagram, as is often noted, wouldn't exist without the iPhone. And Steve Jobs was known to say a few things of the sort that might not get by the filter. It's possible that the world Systrom is trying to create may not just feel nice—it may feel sanitized. It may feel like, say, Disneyland.

When he announced the product in late June, Systrom published an image composed of letters in the shape of a heart and explained what he was doing. The response was mostly positive, rapturous almost. "How great it is!!!!!!," "




," "Amazing

," "Thank you!," "

," "Bravo!!!"

But some critical comments didn't bother the filter. There were complaints, again, about the nonchronological timeline. Other people just thought the whole thing odd. But a few readers, including someone named @futurestrader, noted the most germane concern: "While I do agree IG and social media in general has become a cesspool of trolling, I hope this doesn't push us further down the line of sensorship [sic] in general with comments and ideas that ain't agreed with."



James Mitchell, Instagram’s director of content operations, oversees a team of in-house

raters who have analyzed more than 2 million comments.

The line of censorship in Silicon Valley—what it is and what it isn't—is a crooked one, or at least a blurry one. The government has constitutional limits on its right to censor you, but private platforms do not. Still, the idea of free speech has long been a central concern in Silicon Valley. The hacker ethic, the loose code that early computer programmers embraced, prized the free flow of information. In 2009, Facebook declared its mission to "make the world more open and connected." That was the same year that the thwarted Green Revolution in Iran was briefly called the Twitter Revolution. "We wanted to maximize the number of opinions online, because it was our belief that more speech was better speech," Jason Goldman, a member of the founding team at Twitter, told me. In 2012, an executive at the company referred to the platform as the "free speech wing of the free speech party."

Now that period in Twitter's corporate lifetime looks like a moment of naive idealism: the creation of young men who didn't understand the depths to which sexism, and maybe even fascism, lurk within the human id. Back then, calls for free speech came from people who wanted to bring down dictatorships. Now they seem to come from people demanding the right to say racist stuff without being called racist.

To Systrom, it's pretty simple: Freedom of speech does not mean the freedom to shitpost.

And so the notion of free speech is shifting at the companies that run the internet. Facebook had a reckoning after false stories on its News Feed—free speech, in a sense—may have helped elect Donald Trump. Perhaps not coincidentally, the company changed its mission statement this past June. Now Facebook wants to "give people the power to build community and bring the world closer together." The original mission statement had echoes of James Madison. This one has echoes of a Coke ad.

To Systrom, it's pretty simple: Freedom of speech does not mean the freedom to shitpost. His network isn't a public square; it's a platform people can choose to use or not. When pressed on the matter, he asks, "Is it free speech to just be mean to someone?" Jackson Colaco makes the same point more sharply. "If toxicity on a platform gets so bad that people don't even want to post a comment, they don't even want to share an idea, you've actually threatened expression."

Her comment raises a new problem, though: How exactly do you know when restricting speech helps speech? When is less more and when is less less? These are not questions with easy answers, and Systrom brushes them aside by noting that Instagram is really only targeting the very worst comments. "We're talking about the lower 5 percent. Like the really, really bad stuff. I don't think we're trying to play anywhere in the area of gray."



A decade ago, Tristan Harris and Mike Krieger both studied at Stanford's famous Persuasive Technology Lab, where they learned how technology can shape behavior. Together, the two of them mocked up an app called Send the Sunshine, a rough Crayola sketch of what Instagram is working on now. The app would prompt friends in sunny places to send photographs to friends in places with gloomy weather.

Harris spent several years working at Google, and he now runs a nonprofit called Time Well Spent, from which he has launched a battle against social media. His weapon of choice: long-form journalism. He started gaining significant attention after being profiled in The Atlantic, which called him the "closest thing Silicon Valley has to a conscience." His stature grew after a segment on 60 Minutes and then an interview on Sam Harris' podcast. "Everything they do is about increasing engagement," he tells me, referring to the big tech companies. "You can just substitute 'addiction' for 'engagement,' and it means the same thing."

Harris has partnered with an app called Moment, which measures the amount of time people spend in the other apps on their phones, and then asks them whether they're pleased with that use. Ninety-nine percent of users are happy that they spend time in Google Calendar, and 96 percent are happy with the time they spend on Waze. More than half of all users, though, express unhappiness with the time they spend on Instagram, which averages about 54 minutes a day. Facebook does even worse—it's the app people feel the third worst about using, trailing only Grindr and Candy Crush Saga.

"I think the idea that anyone here is trying to design something maliciously addictive is far-fetched."

When I mention Harris in an interview, Systrom smiles and bats back the critique before it has been fully offered. "Sorry, I'm laughing just because I think the idea that anyone here tries to design something that is maliciously addictive is so far-fetched. We try to solve problems for people, and if that means they like to use the product, I think we've done our job well."

The two Stanford classmates fundamentally agree on one really important thing: The way people use technology can be quite unhealthy. Harris, though, wants companies to "stop hijacking people's minds for the sake of engagement." Systrom wants the engagement to include more sunshine.



After launching the comment filter in late June, he got to work on a related task: elevating high-quality comments in users' feeds. Systrom wouldn't say exactly how he'll measure quality, but the change will roughly be like when the company re-sorted the main feed based on relevancy, not chronology. The idea is to accelerate something called the mimicry effect. When people see others saying nice things, they say nice things too. The contractors in the head scarves and Lincecum jerseys may soon be turning their attention to something new.

Systrom's grand ambition isn't just to fix Instagram. His first goal is to clean up the platform he runs. But, at a time when our national conversation gets darker by the day, he also wants to show the rest of the internet that toxicity online isn't ineluctable. "Maybe trying sends a signal to other companies that this is a priority, and starts a national and international conversation that we should all be having about creating safe and inclusive online communities, not only for our kids but for our friends and our families," he says. "I think that will be success."

When pressed on whether his ultimate goal is to make Instagram better, or in some subtle way to make humans better, he demurs: "I actually think it's about making the internet better."

Is a calmer, kinder internet a better one? Maybe. Instagram, fully filtered, isn't going to knock down dictatorships. And maybe it won't even be that much fun. There's a certain pleasure in seeing Swift's perfectly curated summer soiree interrupted by a green snake emoji. Still, it's hard to make the counterargument—particularly in the era of Trump, and particularly on a service used by so many teens—that the internet should be meaner.

The point may have been proven soon after the product launched, in mid-July, when tragedy struck Instagram. Systrom's close friend, Joy-Vincent Niemantsverdriet, one of the company's senior designers, died in a swimming accident. A distraught Systrom cried openly in staff meetings. Instagram is mostly a dreamscape of lives that seem better than the lives we actually live. But the world has a way of breaching the barriers we build around ourselves. Less than a week later, Systrom posted a tribute to his former colleague: "JV stood for everything IG stands for. He stood for craft, showing kindness without expectation of gain, and looking at the world through a lens of curiosity."

Whether because the filters work, or because humans can be empathetic sometimes, even online, there wasn't a single snarky line or joke in the first few days. Just sympathy, sadness, and even insight—with no one joining to shout it down.

Nicholas Thompson (@nxthompson) is editor in chief of WIRED.

This article appears in the September issue. Subscribe now.

All Grooming by James Anthony