

🕒 This article was published more than **3 years ago**

Democracy Dies in Darkness

Internet ‘algospeak’ is changing our language in real time, from ‘nip nops’ to ‘le dollar bean’

To avoid angering the almighty algorithm, people are creating a new vocabulary

April 8, 2022



By [Taylor Lorenz](#)

“Algospeak” is becoming increasingly common across the Internet as people seek to bypass content moderation filters on social media platforms such as TikTok, YouTube, Instagram and Twitch.

Algospeak refers to code words or turns of phrase users have adopted in an effort to create a brand-safe lexicon that will avoid getting their posts removed or down-ranked by content moderation systems. For instance, in many online videos, it’s common to say “unalive” rather than “dead,” “SA” instead of “sexual assault,” or “spicy eggplant” instead of “vibrator.”

As the pandemic pushed more people to communicate and express themselves online, algorithmic content moderation systems have had an unprecedented impact on the words we choose, particularly on TikTok, and given rise to a new form of internet-driven Aesopian language.

Unlike other mainstream social platforms, the primary way content is distributed on TikTok is through an algorithmically curated “For You” page; having followers doesn’t guarantee people will see your content. This shift has led average users to tailor their videos primarily toward the algorithm, rather than a following, which means abiding by content moderation rules is more crucial than ever.

When the pandemic broke out, people on TikTok and other apps began referring to it as the “Backstreet Boys reunion tour” or calling it the “panini” or “panda express” as platforms down-ranked videos mentioning the pandemic by name in an effort to combat misinformation. When young people began to discuss struggling with mental health, they talked about “becoming unalive” in order to have frank conversations about suicide without algorithmic punishment. Sex workers, who have long been censored by moderation systems, refer to themselves on TikTok as “accountants” and use the corn emoji as a substitute for the word “porn.”

As discussions of major events are filtered through algorithmic content delivery systems, more users are bending their language. Recently, in discussing the invasion of Ukraine, people on YouTube and TikTok have used the sunflower emoji to signify the country. When encouraging fans to follow them elsewhere, users will say “blink in lio” for “link in bio.”

Euphemisms are especially common in radicalized or harmful communities. Pro-anorexia eating disorder communities have long adopted variations on moderated words to evade restrictions. One paper from the School of Interactive Computing, Georgia Institute of Technology found that the complexity of such variants even increased over time. Last year, anti-vaccine groups on Facebook began changing their names to “dance party” or “dinner party” and anti-vaccine influencers on Instagram used similar code words, referring to vaccinated people as “swimmers.”

Tailoring language to avoid scrutiny predates the Internet. Many religions have avoided uttering the devil’s name lest they summon him, while people living in repressive regimes developed code words to discuss taboo topics.

Early Internet users used alternate spelling or “leetspeak” to bypass word filters in chat rooms, image boards, online games and forums. But algorithmic content moderation systems are more pervasive on the modern Internet, and often end up silencing marginalized communities and important discussions.

During YouTube’s “adpocalypse” in 2017, when advertisers pulled their dollars from the platform over fears of unsafe content, LGBTQ creators spoke about having videos demonetized for saying the word “gay.” Some began using the word less or substituting others to keep their content monetized. More recently, users on TikTok have started to say “cornucopia” rather than “homophobia,” or say they’re members of the “leg booty” community to signify that they’re LGBTQ.

“There’s a line we have to toe, it’s an unending battle of saying something and trying to get the message across without directly saying it,” said Sean Szolek-VanValkenburgh, a TikTok creator with over 1.2 million followers. “It disproportionately affects the LGBTQIA community and the BIPOC community because we’re the people creating that verbiage and coming up with the colloquiums.”

Conversations about women’s health, pregnancy and menstrual cycles on TikTok are also consistently down-ranked, said Kathryn Cross, a 23-year-old content creator and founder of Anja Health, a start-up offering umbilical cord blood banking. She replaces the words for “sex,” “period” and “vagina” with other words or spells them with symbols in the captions. Many users say “nip nops” rather than “nipples.”

“It makes me feel like I need a disclaimer because I feel like it makes you seem unprofessional to have these weirdly spelled words in your captions,” she said, “especially for content that’s supposed to be serious and medically inclined.”

Because algorithms online will often flag content mentioning certain words, devoid of context, some users avoid uttering them altogether, simply because they have alternate meanings. “You have to say ‘saltines’ when you’re literally talking about crackers now,” said Lodane Erisian, a community manager for Twitch creators (Twitch considers the word “cracker” a slur). Twitch and other platforms have even gone so far as to remove certain emotes because people were using them to communicate certain words.

Black and trans users, and those from other marginalized communities, often use algospeak to discuss the oppression they face, swapping out words for “white” or “racist.” Some are too nervous to utter the word “white” at all and simply hold their palm toward the camera to signify White people.

“The reality is that tech companies have been using automated tools to moderate content for a really long time and while it’s touted as this sophisticated machine learning, it’s often just a list of words they think are problematic,” said Ángel Díaz, a lecturer at the UCLA School of Law who studies technology and racial discrimination.

In January, Kendra Calhoun, a postdoctoral researcher in linguistic anthropology at UCLA, and Alexia Fawcett, a doctoral student in linguistics at UC Santa Barbara, gave a presentation about language on TikTok. They outlined how, by self-censoring words in the captions of TikToks, new algospeak code words emerged.

TikTok users now use the phrase “le dollar bean” instead of “lesbian” because it’s the way TikTok’s text-to-speech feature pronounces “Le\$bian,” a censored way of writing “lesbian” that users believe will evade content moderation.



Evan Greer, director of Fight for the Future, a digital rights nonprofit advocacy group, said that trying to stomp out specific words on platforms is a fool's errand.

“One, it doesn’t actually work,” she said. “The people using platforms to organize real harm are pretty good at figuring out how to get around these systems. And two, it leads to collateral damage of literal speech.” Attempting to regulate human speech at a scale of billions of people in dozens of different languages and trying to contend with things such as humor, sarcasm, local context and slang can’t be done by simply down-ranking certain words, Greer argues.

“I feel like this is a good example of why aggressive moderation is never going to be a real solution to the harms that we see from big tech companies’ business practices,” she said. “You can see how slippery this slope is. Over the years we’ve seen more and more of the misguided demand from the general public for platforms to remove more content quickly regardless of the cost.”

Big TikTok creators have created shared Google docs with lists of hundreds of words they believe the app's moderation systems deem problematic. Other users keep a running tally of terms they believe have throttled certain videos, trying to reverse engineer the system.

"Zuck Got Me For," a site created by a meme account administrator who goes by Ana, is a place where creators can upload nonsensical content that was banned by Instagram's moderation algorithms. In a manifesto about her project, she wrote: "Creative freedom is one of the only silver linings of this flaming online hell we all exist within ... As the algorithms tighten it's independent creators who suffer."

She also outlines how to speak online in a way to evade filters. "If you've violated terms of service you may not be able to use swear words or negative words like 'hate', 'kill', 'ugly', 'stupid', etc.," she said. "I often write, 'I *opposite of love* xyz' instead of 'I hate xyz.'"

The Online Creators' Association, a labor advocacy group, has also issued a list of demands, asking TikTok for more transparency in how it moderates content. "People have to dull down their own language to keep from offending these all-seeing, all-knowing TikTok gods," said Cecelia Gray, a TikTok creator and co-founder of the organization.

TikTok offers an online resource center for creators seeking to learn more about its recommendation systems, and has opened multiple transparency and accountability centers where guests can learn how the app's algorithm operates.

Vince Lynch, chief executive of IV.AI, an AI platform for understanding language, said in some countries where moderation is heavier, people end up constructing new dialects to communicate. "It becomes actual sub languages," he said.

But as algospeak becomes more popular and replacement words morph into common slang, users are finding that they're having to get ever more creative to evade the filters. "It turns into a game of whack-a-mole," said Gretchen McCulloch, a linguist and author of "Because Internet," a book about how the Internet has shaped language. As the platforms start noticing people saying "seggs" instead of "sex," for instance, some users report that they believe even replacement words are being flagged.

"We end up creating new ways of speaking to avoid this kind of moderation," said Díaz of the UCLA School of Law, "then end up embracing some of these words and they become common vernacular. It's all born out of this effort to resist moderation."

This doesn't mean that all efforts to stamp out bad behavior, harassment, abuse and misinformation are fruitless. But Greer argues that it's the root issues that need to be prioritized. "Aggressive moderation is never going to be a real solution to the harms that we see from big tech companies' business practices," she said. "That's a task for policymakers and for building better things, better tools, better protocols and better platforms."

Ultimately, she added, "you'll never be able to sanitize the Internet."

What readers are saying

The comments reflect a range of opinions on social media, language censorship, and algorithms. Many users express frustration with content moderation and censorship, arguing for more freedom in

language use and criticizing algorithms that amplify conflict for profit. Some... [Show more](#)

This summary is AI-generated. AI can make mistakes and this summary is not a replacement for reading the comments.
