

BRIEF

How Instagram's Algorithm Pushes Potentially Harmful Eating Disorder Content to Users¹

In just two days, an Instagram account registered as a 17-year-old was targeted with over 150 profiles that appear to promote harmful eating disorder content

Findings and Methodology

Avaaz conducted a two-day experiment (from October 6 - 7, 2021) to understand if and how Instagram's algorithm pushes young users toward potentially harmful eating disorder-related content.

Our research indicates that finding Instagram profiles with potentially harmful and triggering eating disorder-related content is as easy as searching for #anorexia², and that Instagram's "suggestion" algorithm can continue to push users further and further into the world of such content through suggesting a seemingly endless stream of "related" profiles. This is despite [Instagram's stated ban](#) on content that "promotes, encourages, coordinates, or provides instructions for eating disorders", as well as [their policy](#) that certain "triggering" posts about eating disorders -- even if they are not actively encouraging them -- "may not be eligible for recommendations" because it "impedes our [Instagram's] ability to foster a safe community."³

For this research, Avaaz created a brand new Instagram account registered as a 17-year-old (gender unspecified). After searching for #anorexia, Instagram suggested different posts associated with the hashtag.⁴ Our team quickly found a profile that had shared a photo of a very thin body, whose bio said: "Ed, thinspo" (meaning "eating disorder" and "thinspiration", a

¹ **Document as a whole is not for distribution or publication. If this research is used, Avaaz must be informed and can be cited as follows, "Preliminary research from global civic organization Avaaz shows/suggests." As part of Avaaz's ongoing non-partisan investigation into disinformation relevant to the US, EU, and critical public policy debates, our team will share snapshots of our findings that serve the public interest. Findings presented below may be updated as our investigation continues. After reviewing our reporting, the social media platform of concern may take moderation actions against the content described in this brief, such as flagging it or removing it from circulation.**

² The organization SumOfUS [documented](#) 22 different hashtags, including #anorexia, that promote eating disorders on Instagram. Additionally, 2019 [research from The Guardian](#) "found thousands of hashtags and accounts promoting anorexia."

³ According to [The Guardian](#), Instagram wants to prevent potentially triggering content while still allowing users in recovery from eating disorders to discuss their experiences. Because of this, Instagram allows users to share their own experiences of eating disorders, provided they are not intended to promote it as a desirable outcome, but such posts "may not be eligible for recommendations."

⁴ After searching for #anorexia, a box initially appears that says: "When it comes to sensitive topics about body image, we want to support our community. We've gathered some resources that may be helpful", which links to resources and a helpline. However, our account could easily bypass this intervention by clicking on "show posts."

[commonly used term](#) among those “who have identified eating disorders as a lifestyle choice, rather than an illness”).

Avaaz followed that profile, and from there Instagram fed our account more and more profiles that shared potentially harmful and triggering content, including extreme dieting, behaviors like bingeing and purging, photos of very thin or emaciated people, descriptions of self-harm, “current weights” versus “goal weights”, and “accountability” posts that promised to fast or exercise in exchange for likes and comments.

Over the course of two days⁵ following these types of profiles, our team uncovered some additional troubling elements about Instagram’s suggestion algorithm, notably that it suggested **private user profiles** to follow as well as **profiles seemingly belonging to minors**.⁶

While Avaaz could not see the posts of private user profiles, our threshold for including these in our dataset was based on explicit or coded language used in their publicly available bios that strongly indicated their posts may contain potentially harmful and triggering content. Such bios often included terms like “CM/GW/UGW” (current weight, goal weight, ultimate goal weight) or “TW ED/ANA/MIA/SH” (trigger warning eating disorder, anorexia, bulimia, self-harm) -- acronyms that [experts](#) have documented as commonly used by people both suffering from or promoting eating disorder behaviors.

In total, Avaaz documented:

- **153 recommended profiles with a combined 94,762 followers**⁷ whose bios and/or posts contained harmful eating disorder-related content;
- **The majority (66%) of recommended profiles were private** (101 out of 153), while only 34% were public (52 out of 153);
- **12 recommended profiles appeared to belong to minors** under the age of 18;
- **The majority (85%) of recommended profiles were “small accounts”** with fewer than 1,000 followers. Of these, 25% had fewer than 100 followers.

Avaaz’s work follows the recent research and findings of [Senator Richard Blumenthal’s team](#), whose Instagram account registered as a 13-year-old girl was increasingly fed more and more extreme dieting accounts by the Instagram algorithm over time, after initially following some dieting and pro-eating disorder accounts.

Our combined research indicates that Instagram has not thoroughly investigated and addressed in earnest how its algorithm feeds potentially harmful and triggering eating disorder content to young users. To the contrary, its algorithm has learned to associate

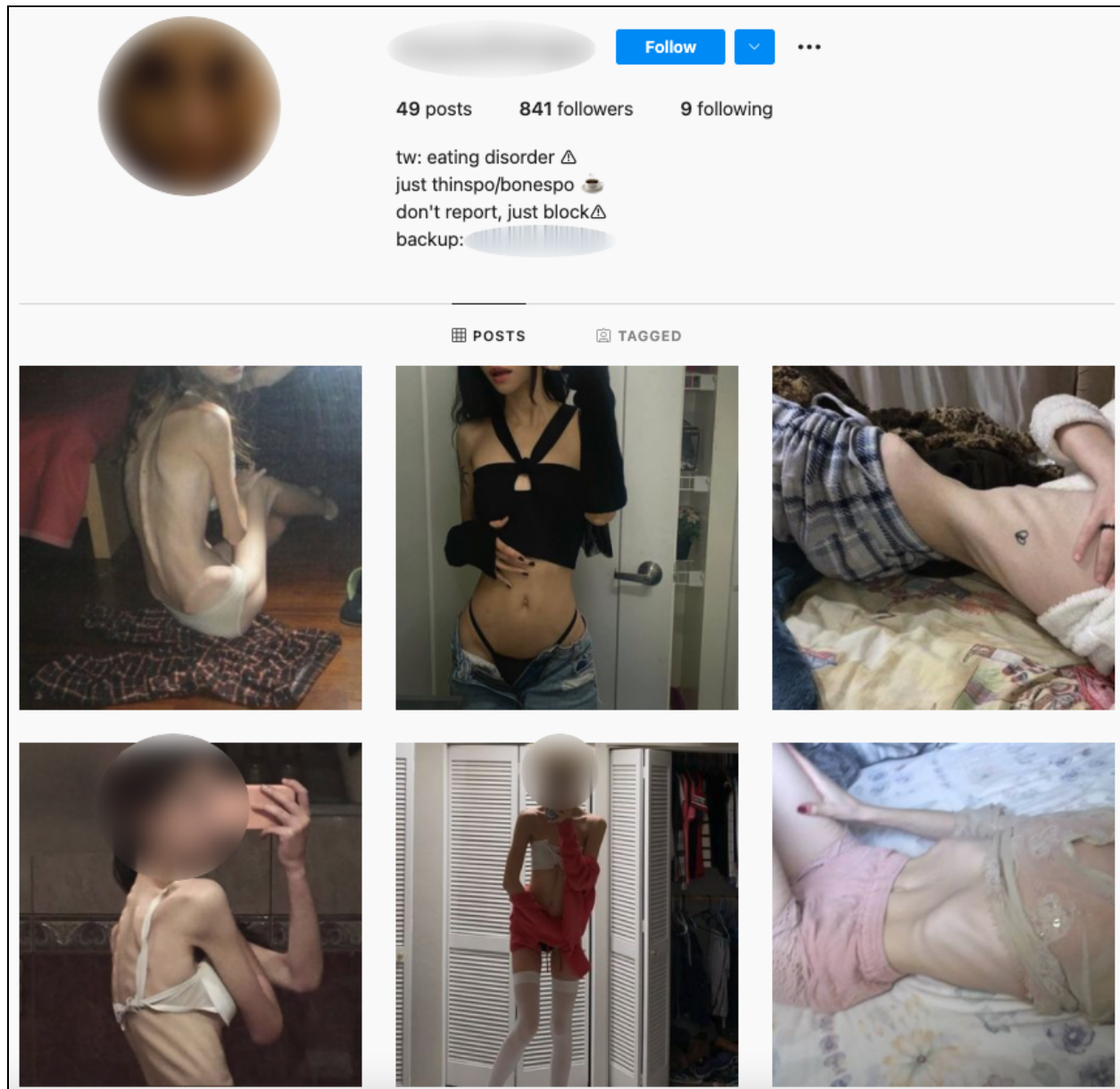
⁵ Avaaz’s research period was limited by the duration of its research account, which was deactivated by Instagram without prior notification.

⁶ Unlike Facebook, Instagram does not display a user’s birthday. Instead, Avaaz identified profiles that we reasonably assume belong to minors based on their bios, in which the user chose to disclose their age.

⁷ As of October 14, 2021, this follower count is up-to-date. Additionally, as of this date, four of these accounts have been removed.

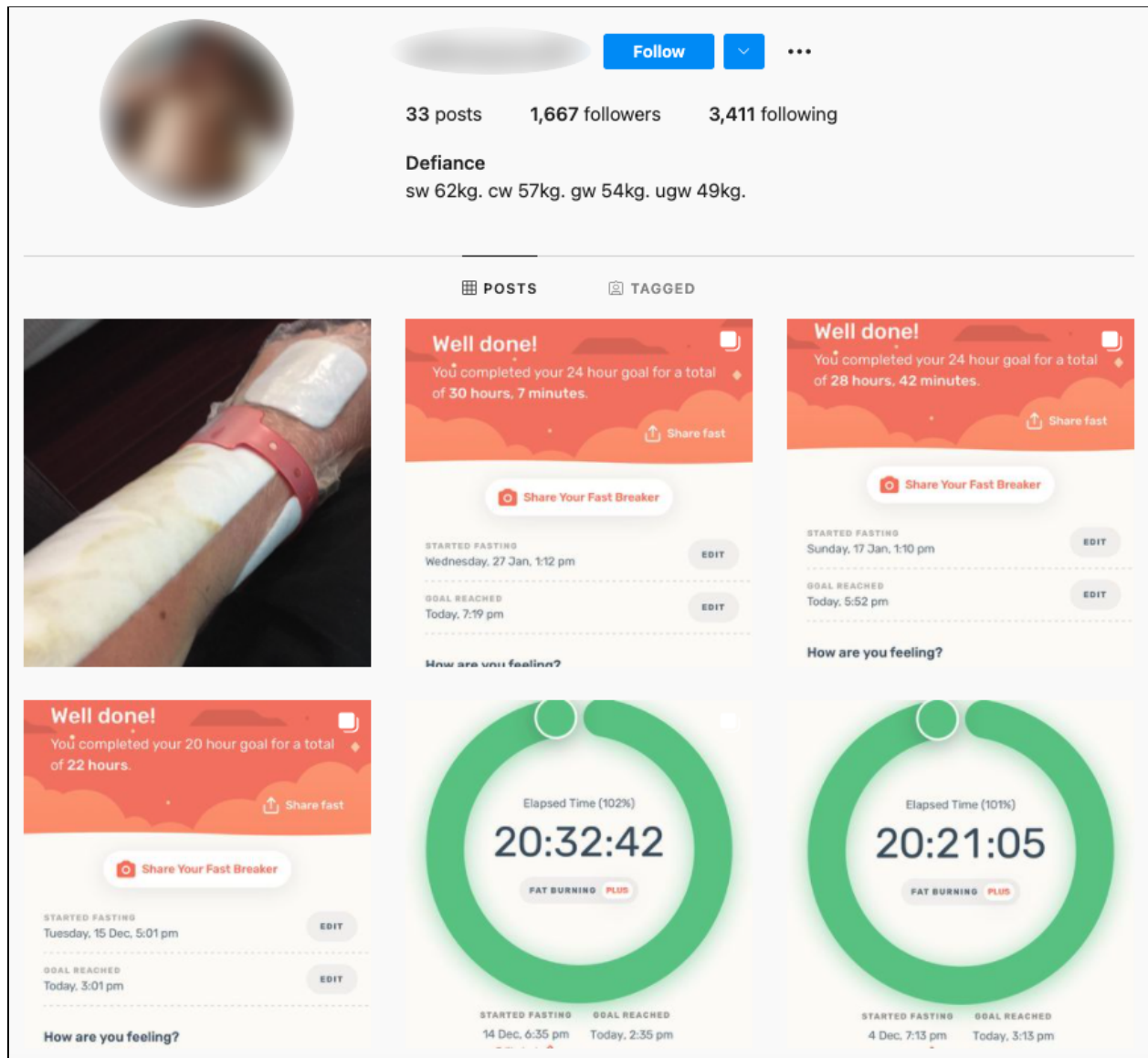
triggering and at times quite graphic profiles with one another and push them to young users to follow, even if those profiles are private and/or seemingly belong to minors. This puts certain demographics of users at increased risk for self-harm, as eating disorder experts have reported that such content can [“act as validation for users already predisposed to unhealthy behaviors.”](#)

(*Trigger Warning*) Profile and Content Examples

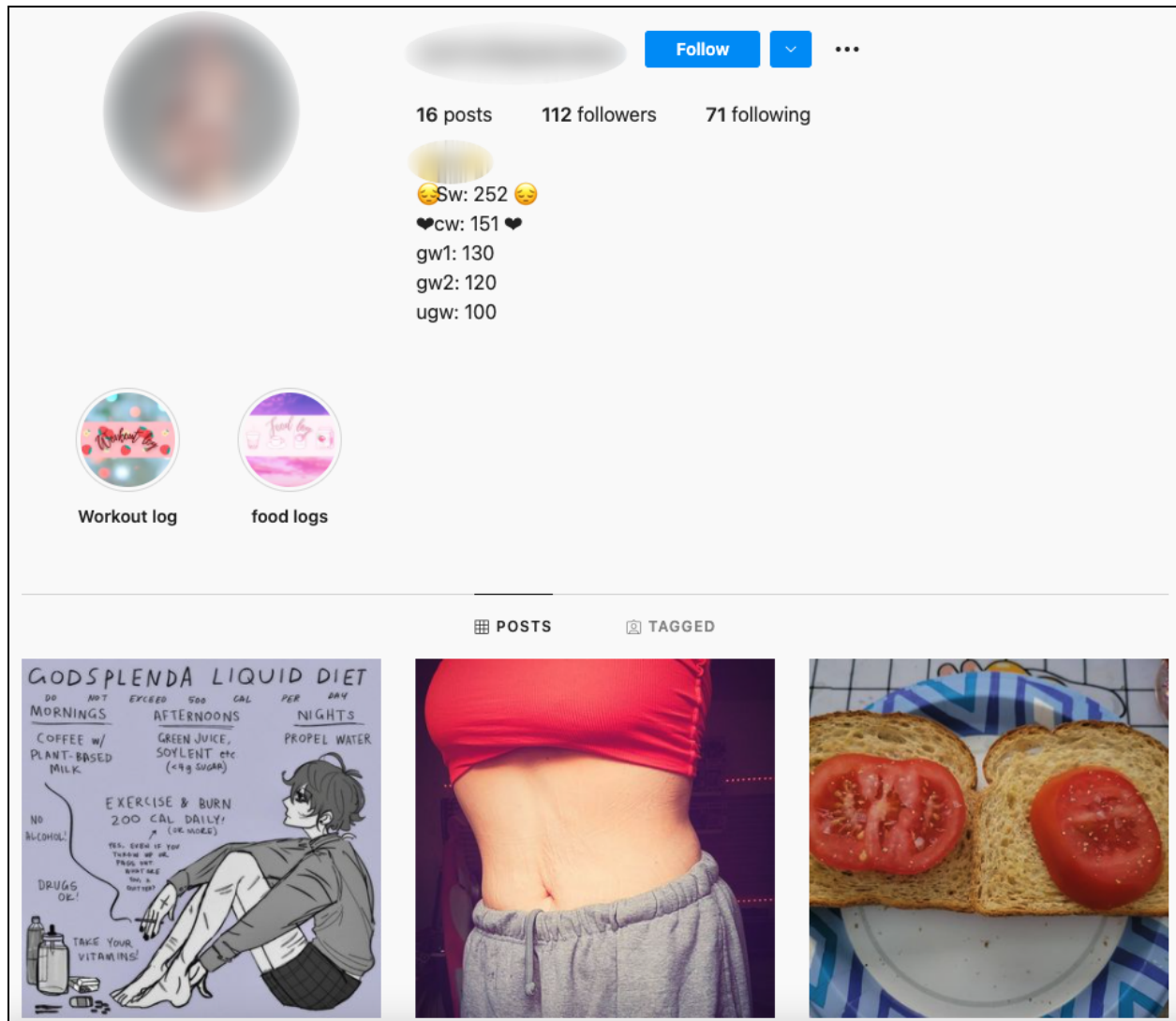


This user's bio says: "tw [trigger warning]: eating disorder" and "just thinspo/bonespo" ("thinspo" and "bonespo" are abbreviated versions of "thinspiration" and "bonespiration", respectively).

These two terms are associated with posting photos of extremely thin people as “inspiration”. As is common with other accounts in this dataset, the user asks people not to report them, but block instead, to presumably avoid detection from Instagram.



This user’s bio includes their sw (meaning “start weight”), cw (current weight), gw (goal weight), and ugw (ultimate goal weight). These abbreviations can be thought of as internal “code” within the Instagram eating disorder community, signalling to others what kind of content is likely present on their profile. This user consistently shares photos of their fasting “achievements”, anywhere from 20 to 30 hours. Their most recent photo is apparently from the hospital after committing self-harm.



Similar to the previous profile, this user's bio shares their starting, current, goal, and ultimate goal weights. Their Instagram Stories Highlights include a workout log and food log, the latter of which details how to have a 156-calorie breakfast and 128-calorie lunch. This user consistently posts photos showing extreme low-calorie meals; for instance, their most recent post says: "do not exceed 500 cal per day" and details how to achieve this through a mostly liquid-based diet.

Policy Recommendation

To ensure that users are adequately protected against harmful content on Instagram Avaaz advocates for **transparency and audits**: The government, researchers and the public must have the tools to understand how social media platforms work and their cumulative impact. The platforms must be required to provide comprehensive reports on disinformation, measures taken against it, and the design, operation, and impact of their curation algorithms (while

respecting trade secrets). Platforms' algorithms must also be continually, independently audited to measure impact and to improve design, operation and outcomes.

END