

This week,

A former Facebook employee has gotten a lot of press and a senate hearing for revealing tens of thousands of internal Facebook documents about how the company fuels negative well-being in children, hate, and polarization. So let's talk about this in context of antisemitism — what the problems are, and how we fix it.

I'm Lev Gringauz, and welcome to The Jews Are Tired, your podcast about Jewish news.

Alright, let's set the context for this: Jewish media doesn't really tend to cover social media unless there's a report about how much antisemitism there is; or a complaint from a Jewish organization about how much antisemitism there is; or social media companies promise that they'll do a better job of removing antisemitic posts, comments, and images from their platforms.

Which is to say, most of the time, we — and Jews in general — are focused on content and content moderation being the problem on social media. That often is expressed with some version of the question, 'why is Holocaust denial allowed on Facebook?'

But what's notable about the recent news dump around Facebook, and the former employee, Frances Haugen [HAW-GEN], is that it's reframing the conversation. Bad content like hate speech, and content moderation, isn't actually the core problem, nor is it even fixable. The real issue is how Facebook prioritizes content to show it to more people.

And I'll just say the uncomfortable kicker now: What this conversation now boils down to is that we can't stop Holocaust denial and other terrible things on social media — but we can drastically slow down how many people see those terrible things, and as a result, how many people believe them.

But if we, the American Jews and other citizens of this country, don't understand why this is the case, we will fail to push legislators to fix what can actually be fixed. And the worst case scenario is that, if we are misinformed, our legislators will only make this situation worse.

Like many things, this is a multi-faceted issue. So let's start with Frances Haugen.

Haugen has worked at a lot of major tech companies, like Google and Pinterest, and, most importantly now, Facebook, where she was a project manager on the Civic Integrity team from mid-2019 until early 2021. Her background is in algorithms, and how algorithms push content to people online. By the way, Facebook also owns Instagram and Whatsapp.

But Haugen joined the Civic Integrity team not to push content to people, but to study how Facebook was being used to spread misinformation and violence-inciting rhetoric so that Facebook could, theoretically, put a stop to that kind of activity, particularly when it came to elections.

Haugen was driven by a personal reason to do this sort of work: A close friend had become a white nationalist after reading white supremacist conspiracy theories online. In her own words, as she told The Wall Street Journal: QUOTE “It’s one thing to study misinformation, it’s another to lose someone to it.” ENDQUOTE

But what Haugen found at Facebook was a culture where problems were researched, but solutions were largely rejected and the people working on solutions were understaffed to the point of being practically invisible.

Why? Because fixing hate speech and misinformation on Facebook is bad for business.

Facebook’s business model is about selling people’s attention to advertisers, which it did last year for \$86 billion in revenue. To do that, Facebook needs three things: A lot of people; those people spending as much time as possible on Facebook so that they are more likely to see ads; and ads that are targeted to people so that they are more likely to click on those ads.

To manage that, Facebook has a bunch of machine-learning algorithms, which are complex sets of calculations and equations that can learn what data to provide based on context, and to then provide that data consistently in the future. Now, I’m not a computer science journalist, so here’s a good explanation on these algorithms from a story, linked to in the podcast notes, from the MIT Technology Review:

QUOTE “An algorithm trained on ad click data, for example, might learn that women click on ads for yoga leggings more often than men. The resultant model will then serve more of those ads to women.

Facebook’s massive amounts of user data...[allowed the company to] develop models that learned to infer the existence not only of broad categories like “women” and “men,”

but of very fine-grained categories like “women between 25 and 34 who liked Facebook pages related to yoga,” and targeted ads to them. The finer-grained the targeting, the better the chance of a click, which would give advertisers more bang for their buck.”
ENDQUOTE

So that answers how Facebook tries to provide relevant ads to people. And Facebook has the people to sell the ads to, roughly 2.8 billion people a month that use it in 2021, thanks to it being, well, a free service for all of us to connect with each other on.

But Facebook’s problem in this profit trifecta is that we the people have started to use it less. By early 2018, internal documents showed clearly that the number of likes, comments, and shares were all declining. And if we don’t use Facebook as much, or engage with posts that much, we’re a step closer to not using Facebook at all. Which means Facebook wouldn’t be able to use our attention to sell as many ads and make money.

So the question became, how to keep people more engaged with Facebook. The answer, in early 2018, came with a change to the algorithms that decide what content should be in the Facebook newsfeed, the central endless scroll of posts and stuff that most of us look at when using the platform. If the content interests us, we’ll be more engaged.

The change put more emphasis on what Facebook calls “meaningful social interactions,” which were ranked by points. As an example of some of that, a like on a post was one point, a reaction other than a like was five points, and certain kinds of comments and reshares were rated at 30 points. So posts that were more likely to inspire a lot of reactions and reshares would be boosted to more people’s timelines for them to see it. And as they reacted and reshared, Facebook could slip in some advertisements to look at, too.

But what kind of posts and content are likely to get a lot of reactions, and thus get shown to more and more people? Inflammatory, negative, polarizing content. Stuff like hate speech, conspiracy theories, anti-vaccine content, antisemitism, and Holocaust denial. Which the algorithms, not knowing any better, now had extra reason to spread.

And Facebook has known it this whole time. In late 2018, Mark Zuckerberg, founder and CEO of Facebook, wrote in a policy note that he published to his Facebook account that

QUOTE “One of the biggest issues social networks face is that, when left unchecked, people will engage disproportionately with more sensationalist and provocative content.

This is not a new phenomenon. It is widespread on cable news today and has been a staple of tabloids for more than a century. At scale it can undermine the quality of public discourse and lead to polarization.” ENDQUOTE

AKA people like paying attention to bad and outrageous stuff. The new algorithms, at the end of the day, were just designed to give people what we already wanted — trash, and lots of it. But the new algorithms also work, as Facebook did see a rise in more engagement, which means they had less reason to worry about keeping our attention for profit.

So that, ultimately, is Facebook’s tradeoff. Its business succeeds by making our world worse, in many ways. And when researchers and employees said, hey, you can make things better by reversing these algorithm changes so that we’re not just pushing whatever gets the most reactions and outrage, Facebook didn’t want to do that. Because reducing the toxicity of Facebook, also means reducing engagement on Facebook, which means less money.

And bringing it all the way back to the Facebook whistleblower, Frances Haugen, she saw and learned from internal documents that Facebook knew all this, and very little was being done to fix it. So when Haugen left Facebook earlier this year, she decided to bring all the recent research with her, and she gave those documents to The Wall Street Journal to publish, and to the federal government to investigate.

The Wall Street Journal’s series of stories on this, published over the past month, are linked in the podcast notes, by the way. They, and in-depth reporting from Wired and the MIT Technology Review, form the backbone of this episode. There’s also a lot of additional stories and elements to this that the original reporting covers, which I am not going into on this episode. So I highly recommend you read what’s linked in the podcast notes if you have the time.

Now let’s finish off this episode by addressing what may be the elephant in the room: I said that all of this information leaves us with the reality that bad stuff on Facebook, for example, Holocaust denial posted every day, can’t be stopped — but its exposure to people can be drastically slowed down if these algorithms are fixed.

But why *can’t* Facebook just take down all those antisemitic posts, comments, and images and keep them off the platform for good? Unfortunately, the answer is that doing so is just not possible.

Facebook has billions of posts and comments uploaded every day in every language. There aren't enough people to train to monitor every single piece of content that users put on Facebook. And algorithms are a mixed bag.

On one hand, Facebook has some really good content identifying algorithms. According to a Wired story from 2019, Facebook catches more than 99% of spam and posts supporting terrorism. Nudity is identified 96% of the time. But for hate speech, QUOTE "Facebook finds just 52 percent before users do." ENDQUOTE

That's because it's really hard to train algorithms to identify hate speech. Here's how that MIT Technology Review explains it.

QUOTE "Misinformation and hate speech constantly evolve. New falsehoods spring up; new people and groups become targets. To catch things before they go viral, content-moderation models must be able to identify new unwanted content with high accuracy. But machine-learning models do not work that way. An algorithm that has learned to recognize Holocaust denial can't immediately spot, say, Rohingya genocide denial.

"It must be trained on thousands, often even millions, of examples of a new type of content before learning to filter it out. Even then, users can quickly learn to outwit the model by doing things like changing the wording of a post or replacing incendiary phrases with euphemisms, making their message illegible to the AI while still obvious to a human. This is why new conspiracy theories can rapidly spiral out of control." ENDQUOTE

And doing all this, again, across different languages, is really difficult, especially when the algorithms already struggle just in English. A perfect example comes from The Wall Street Journal, which reported that earlier this year, as Facebook tried to cut down on anti-vaccine content and misinformation, it really struggled. One anti-vaccine post that said vaccines QUOTE "are all experimental & you are in the experiment," was left up to get 53,000 reshares and three million views. Apparently Facebook didn't take this post down because its systems thought the post was written in Romanian, and thus didn't flag it as an issue.

Frances Haugen raised the language issue as well during her Senate hearing on Tuesday this week when explaining why trying to moderate specific content would never succeed.

She said QUOTE “In the case of Ethiopia there are 100 million people and six languages. Facebook only supports two of those languages for integrity systems...this strategy of focusing on language-specific, content-specific systems for AI to save us is doomed to fail.” ENDQUOTE

And there’s also another issue at play, which is that of privacy. Even if Facebook theoretically did have the ability to monitor every single thing anyone posts or comments or direct messages on its platform, that would mean constant surveillance on everything we do in our digital lives. If not on Facebook, then on Instagram — which Facebook owns. If not on Instagram, then on Whatsapp — which Facebook owns. That’s also a problem for those of us who care about how much data social media companies collect about us, and use for their own profits.

And if you want to venture into the territory of direct government oversight of what content should or shouldn’t be allowed on social media, you have to be extra careful. The First Amendment of the U.S. Constitution, which, among other things, guarantees the right to freedom of speech, only makes sure that the government guarantees freedom of speech. As a private company, Facebook doesn’t have to guarantee freedom of speech at all to anybody. In fact, they have a First Amendment right to block or not block whoever or whatever they want. And it would be serious overreach for the federal government to suddenly ignore the First Amendment and dictate what speech is or isn’t ok to social media companies.

Which leads us back to the best solution being the one that is possible to actually do — pressure Facebook into fixing its algorithms and the system that helps terrible content be seen by too many people. Part of the reason Haugen is whistleblowing on Facebook is so that Congress can figure out a decent way to regulate these sorts of algorithms. All of which can be done without attempting the impossible and trying to regulate content directly.

What comes next remains to be seen. Congress typically sucks at understanding the internet, let alone how to regulate problems online. But again, it feels like the conversation is clearer now, and focused not on individual bad posts and how to stop them, but the larger systems that are responsible for inciting hate, misinformation, and violence. To leave off, here’s an excerpt from Haugen’s opening statement at the Senate hearing on Tuesday that sums up what’s at stake, edited down from its original 9 minutes.

End:

This has been this week's The Jews Are Tired podcast, I'm Lev Gringauz, don't forget to subscribe and share, and hopefully next week, the Jews will get some rest.

The Jews Are Tired is a product of Jewfolk, Inc. For more information, go to TCJewfolk.com, or email the show at podcast@tcjewfolk.com. A link to the transcript of this episode is available in the podcast notes, along with links to any news stories or reports referenced for this episode.