AI and Ethics (2021) 1:395–403
https://doi.org/10.1007/s43681-021-00068-x

OPINION PAPER

Facebook's ethical failures are not accidental; they are part of the business model

David Lauer[1]

Received: 13 April 2021 / Accepted: 29 May 2021

Facebook's stated mission is "to give people the power to build community and bring the world closer together." But a deeper look at their business model suggests that it is far more profitable to drive us apart. By creating "filter bubbles"—social media algorithms designed to increase engagement and, consequently, create echo chambers where the most inflammatory content achieves the greatest visibility—Facebook profits from the proliferation of extremism, bullying, hate speech, disinformation, conspiracy theory, and rhetorical violence. Facebook's problem is not a technology problem. It is a business model problem. This is why solutions based in technology have failed to stem the tide of problematic content. If Facebook employed a business model focused on efficiently providing accurate information and diverse views, rather than addicting users to highly engaging content within an echo chamber, the algorithmic outcomes would be very different.

Facebook's failure to check political extremism, [15] willful disinformation, [39] and conspiracy theory [43] has been well-publicized, especially as these unseemly elements have penetrated mainstream politics and manifested as deadly, real-world violence. So it naturally raised more than a few eyebrows when Facebook's Chief AI Scientist Yann LeCun tweeted his concern [32] over the role of right-wing personalities in downplaying the severity of the COVID-19 pandemic. Critics were quick to point out [29] that Facebook has profited handsomely from exactly this brand of disinformation. Consistent with Facebook's recent history on such matters, LeCun was both defiant and unconvincing.

In response to a frenzy of hostile tweets, LeCun made the following four claims:

1. Facebook does not cause polarization or so-called "filter bubbles" and that "most serious studies do not show this."
2. Critics [30] who argue that Facebook is profiting from the spread of misinformation—are "factually wrong."[1]
3. Facebook uses AI-based technology to filter out [33]:
    a. Hate speech;
    b. Calls to violence;
    c. Bullying; and
    d. Disinformation that endangers public safety or the integrity of the democratic process.
4. Facebook is not an "arbiter of political truth" and that having Facebook "arbitrate political truth would raise serious questions about anyone's idea of ethics and liberal democracy."

Absent from the claims above is acknowledgement that the company's profitability depends substantially upon the polarization LeCun insists does not exist.

Facebook has had a profound impact on our access to ideas, information, and one another. It has unprecedented global reach, and in many markets serves as a de-facto monopolist. The influence it has over individual and global affairs is unique in human history. Mr. LeCun has been at Facebook since December 2013, first as Director of AI Research and then as Chief AI Scientist. He has played a leading role in shaping Facebook's technology and approach. Mr. LeCun's problematic claims demand closer examination. What follows, therefore, is a response to these claims which will clearly demonstrate that Facebook:

- Elevates disinformation campaigns and conspiracy theories from the extremist fringes into the mainstream, fostering, among other effects, the resurgent anti-vaccination movement, broad-based questioning of basic public

✉ David Lauer
dave@urvin.ai

[1] Urvin AI, 413 Virginia Ave, Collingswood, NJ 08107, USA

[1] Facebook executives have, themselves, acknowledged that Facebook profits from the spread of misinformation: https://www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news.

health measures in response to COVID-19, and the proliferation of the Big Lie of 2020—that the presidential election was stolen through voter fraud [16];
- Empowers bullies of every size, from cyber-bullying in schools, to dictators who use the platform to spread disinformation, censor their critics, perpetuate violence, and instigate genocide;
- Defrauds both advertisers and newsrooms, systematically and globally, with falsified video engagement and user activity statistics;
- Reflects an apparent political agenda espoused by a small core of corporate leaders, who actively impede or overrule the adoption of good governance;
- Brandishes its monopolistic power to preserve a social media landscape absent meaningful regulatory oversight, privacy protections, safety measures, or corporate citizenship; and
- Disrupts intellectual and civil discourse, at scale and by design.

1 I deleted my Facebook account

I deleted my account years ago for the reasons noted above, and a number of far more personal reasons. So when LeCun reached out to me, demanding evidence for my claims regarding Facebook's improprieties, it was via Twitter. What proof did I have that Facebook creates filter bubbles that drive polarization?

In anticipation of my response, he offered the claims highlighted above. As evidence of his claims, he directed my attention to a single research paper [23] that, on closer inspection, does not appear at all to reinforce his case.

The entire exchange also suggests that senior leadership at Facebook still suffers from a massive blindspot regarding the harm that its platform causes—that they continue to "move fast and break things" without regard for the global impact of their behavior.

LeCun's comments confirm the concerns that many of us have held for a long time: Facebook has declined to resolve its systemic problems, choosing instead to paper over these deep philosophical flaws with advanced, though insufficient, technological solutions. Even when Facebook takes occasion to announce its triumphs in the ethical use of AI, such as its excellent work [8] detecting suicidal tendencies, its advancements pale in comparison to the inherent problems written into its algorithms.

This is because, fundamentally, their problem is not a failure of technology, nor a shortcoming in their AI filters. Facebook's problem is its business model. Facebook makes superficial technology changes, but at its core, profits chiefly from engagement and virality. Study after study has found that "lies spread faster than the truth," [47] "conspiracy theories spread through a more decentralized network," [41] and that "politically extreme sources tend to generate more interactions from users."[2] Facebook knows that the most efficient way to maximize profitability is to build algorithms that create filter bubbles and spread viral misinformation.

This is not a fringe belief or controversial opinion. This is a reality acknowledged even by those who have lived inside of Facebook's leadership structure. As the former director of monetization for Facebook, Tim Kendall explained in his Congressional testimony, "social media services that I, and others have built, have torn people apart with alarming speed and intensity. At the very least we have eroded our collective understanding—at worst, I fear we are pushing ourselves to the brink of a civil war." [38]

2 Facebook's black box

To effectively study behavior on Facebook, we must be able to study Facebook's algorithms and AI models. Therein lies the first problem. The data and transparency to do so are simply not there. Facebook does not practice transparency—they do not make comprehensive data available on their recommendation and filtering algorithms, or their other implementations of AI. One organization attempting to study the spread of misinformation, NYU's Cybersecurity for Democracy, explains, "[o]ur findings are limited by the lack of data provided by Facebook…. Without greater transparency and access to data, such research questions are out of reach."[3]

Facebook's algorithms and AI models are proprietary, and they are intentionally hidden from us. While this is normal for many companies, no other company has 2.85 billion monthly active users. Any platform that touches so many lives must be studied so that we can truly understand its impact. Yet Facebook does not make the kind of data available that is needed for robust study of the platform.

Facebook would likely counter this, and point to their partnership with Harvard's Institute for Quantitative Social Science (Social Science One) as evidence that they are making data available to researchers [19]. While this partnership is one step in the right direction, there are several problems with this model:

[2] Cybersecurity for Democracy. (March 3, 2021). "Far-right news sources on Facebook more engaging." https://medium.com/cybersecurity-for-democracy/far-right-news-sources-on-facebook-more-engaging-e04a01efae90.

[3] Ibid.

- The data are extremely limited. At the moment it consists solely of web page addresses that have been shared on Facebook for 18 months from 2017 to 2019.
- Researchers have to apply for access to the data through Social Science One, which acts as a gatekeeper of the data.
- If approved, researchers have to execute an agreement directly with Facebook.

This is not an open, scientific process. It is, rather, a process that empowers administrators to cherry-pick research projects that favor their perspective. If Facebook was serious about facilitating academic research, they would provide far greater access to, availability of, and insight into the data. There are legitimate privacy concerns around releasing data, but there are far better ways to address those concerns while fostering open, vibrant research.

3 Does Facebook cause polarization?

LeCun cited a single study as evidence that Facebook does not cause polarization. But do the findings of this study support Mr. LeCun's claims?

The study concludes that "polarization has increased the most among the demographic groups least likely to use the Internet and social media." The study does not, however, actually measure this type of polarization directly. Its primary data-gathering instrument—a survey on polarization—did not ask whether respondents were on the Internet or if they used social media. Instead, the study estimates whether an individual respondent is likely to be on the Internet based on an index of demographic factors which suggest "predicted" Internet use. As explained in the study, "the main predictor [they] focus on is age" [23]. Age is estimated to be negatively correlated with social media usage. Therefore, since older people are also shown to be more politically polarized, LeCun takes this as evidence that social media use does not cause polarization.

This assumption of causality is flawed. The study does not point to a causal relationship between these demographic factors and social media use. It simply says that these demographic factors drive polarization. Whether these factors have a correlational or causative relationship with the Internet and social media use is complete conjecture. The author of the study himself caveats any such conclusions, noting that "[t]hese findings do not rule out any effect of the internet or social media on political polarization." [5].

Not only is LeCun's assumption flawed, it is directly refuted by a recent Pew Research study [3] that found an overwhelmingly high percentage of US adults age 65 + are on Facebook (50%), the most of any social network. If anything, older age is actually more clearly correlated with Facebook use relative to other social networks.

Moreover, in 2020, the MIS Quarterly journal published a study by Steven L. Johnson, et al. that explored this problem and found that the "more time someone spends on Facebook, the more polarized their online news consumption becomes. This evidence suggests Facebook indeed serves as an echo chamber especially for its conservative users" [24].

Allcott, et al. also explores this question in "The Welfare Effects of Social Media" in November, 2019, beginning with a review of other studies confirming a relationship between social media use, well-being and political polarization [1]:

> More recent discussion has focused on an array of possible negative impacts. At the individual level, many have pointed to negative correlations between intensive social media use and both subjective well-being and mental health. Adverse outcomes such as suicide and depression appear to have risen sharply over the same period that the use of smartphones and social media has expanded. Alter (2018) and Newport (2019), along with other academics and prominent Silicon Valley executives in the "time well-spent" movement, argue that digital media devices and social media apps are harmful and addictive. At the broader social level, concern has focused particularly on a range of negative political externalities. Social media may create ideological "echo chambers" among like-minded friend groups, thereby increasing political polarization (Sunstein 2001, 2017; Settle 2018). Furthermore, social media are the primary channel through which misinformation spreads online (Allcott and Gentzkow 2017), and there is concern that coordinated disinformation campaigns can affect elections in the US and abroad.

Allcott's 2019 study uses a randomized experiment in the run-up to the November 2018 midterm elections to examine how Facebook affects several individual and social welfare measures. They found that:

> deactivating Facebook for the four weeks before the 2018 US midterm election (1) reduced online activity, while increasing offline activities such as watching TV alone and socializing with family and friends; (2) reduced both factual news knowledge and political polarization; (3) increased subjective well-being; and (4) caused a large persistent reduction in post-experiment Facebook use.

In other words, not using Facebook for a month made you happier and resulted in less future usage. In fact, they say that "deactivation significantly reduced polarization of views on policy issues and a measure of exposure to polarizing

news." None of these findings would come as a surprise to anybody who works at Facebook.

"A former Facebook AI researcher" confirmed that they ran "'study after study' confirming the same basic idea: models that maximize engagement increase polarization" [21]. Not only did Facebook know this, but they continued to design and build their recommendation algorithms to maximize user engagement, knowing that this meant optimizing for extremism and polarization.[4]

Facebook understood what they were building according to Tim Kendall's Congressional testimony in 2020. He explained that "we sought to mine as much attention as humanly possible and turn [sic] into historically unprecedented profits" [38]. He went on to explain that their inspiration was "Big Tobacco's playbook … to make our offering addictive at the outset." They quickly figured out that "extreme, incendiary content" directly translated into "unprecedented engagement—and profits." He was the director of monetization for Facebook—few would have been better positioned to understand Facebook's motivations, findings and strategy.

4 Engagement, filter bubbles, and executive compensation

The term "filter bubble" was coined by Eli Pariser who wrote a book with that title, exploring how social media algorithms are designed to increase engagement and create echo chambers where inflammatory posts are more likely to go viral. Filter bubbles are not just an algorithmic outcome; often we filter our own lives, surrounding ourselves with friends (online and offline) who are more likely to agree with our philosophical, religious and political views.

Social media platforms capitalize on our natural tendency toward filtered engagement. These platforms build algorithms, and structure executive compensation, [27] to maximize such engagement. By their very design, social media curation and recommendation algorithms are engineered to maximize engagement, and thus, are predisposed to create filter bubbles.

Facebook has long attracted criticism for its pursuit of growth at all costs. A recent profile of Facebook's AI efforts details the difficulty of getting "buy-in or financial support when the work did not directly improve Facebook's growth." [21]. Andrew Bosworth, a Vice President at Facebook said in a 2016 memo that nothing matters but growth, and that "all the work we do in growth is justified" regardless of whether "it costs someone a life by exposing someone to bullies" or if "somebody dies in a terrorist attack coordinated on our tools" [31].

Bosworth and Zuckerberg went on to claim [36] that the shocking memo was merely an attempt at being provocative. Certainly, it succeeded in this aim. But what else could they really say? It's not a great look. And it looks even worse when you consider that Facebook's top brass really do get paid more when these things happen. The above-referenced report is based on interviews with multiple former product managers at Facebook, and shows that their executive compensation system is largely based around their most important metric–user engagement. This creates a perverse incentive. And clearly, by their own admission, Facebook will not allow a few casualties to get in the way of their executive compensation.

5 Is it incidental or intentional?

Yaël Eisenstat, a former CIA analyst who specialized in counter-extremism went on to work at Facebook out of concern that the social media platform was increasing radicalization and political polarization. She explained in a TED talk [13] that the current information ecosystem is manipulating its users, and that "social media companies like Facebook profit off of segmenting us and feeding us personalized content that both validates and exploits our biases. Their bottom line depends on provoking a strong emotion to keep us engaged, often incentivizing the most inflammatory and polarizing voices." This emotional response results in more than just engagement—it results in addiction.

Eisenstat joined Facebook in 2018 and began to explore the issues which were most divisive on the social media platform. She began asking questions internally about what was causing this divisiveness. She found that "the largest social media companies are antithetical to the concept of reasoned discourse … Lies are more engaging online than truth, and salaciousness beats out wonky, fact-based reasoning in a world optimized for frictionless virality. As long as algorithms' goals are to keep us engaged, they will continue to feed us the poison that plays to our worst instincts and human weaknesses."

She equated Facebook's algorithmic manipulation to the tactics that terrorist recruiters use on vulnerable youth. She offered Facebook a plan to combat political disinformation and voter suppression. She has claimed that the plan was rejected, and Eisenstat left after just six months.

As noted earlier, LeCun flatly denies [34] that Facebook creates filter bubbles that drive polarization. In sharp contrast, Eisenstat explains that such an outcome is a feature of their algorithm, not a bug. The Wall St. Journal reported that in 2018, senior executives at Facebook were informed of the following conclusions during an internal presentation [22]:

[4] Ibid.

- "Our algorithms exploit the human brain's attraction to divisiveness… [and] if left unchecked," Facebook would feed users "more and more divisive content in an effort to gain user attention and increase time on the platform."
- The platform aggravates polarization and tribal behavior.
- Some proposed algorithmic changes would "disproportionately affect[] conservative users and publishers."
- Looking at data for Germany, an internal report found "64% of all extremist group joins are due to our recommendation tools … Our recommendation systems grow the problem."

These are Facebook's own words, and arguably, they provide the social media platform with an invaluable set of marketing prerogatives. They are reinforced by Tim Kendall's testimony as discussed above.

"Most notably," reported the WSJ, "the project forced Facebook to consider how it prioritized 'user engagement'—a metric involving time spent, likes, shares and comments that for years had been the lodestar of its system." As noted in the section above, executive compensation was tied to "user engagement," which meant product developers at Facebook were incentivized to design systems in this very way.[5]

Mark Zuckerberg and Joel Kaplan reportedly [22] dismissed the conclusions from the 2018 presentation, calling efforts to bring greater civility to conversations on the social media platform "paternalistic." Zuckerberg went on to say that he would "stand up against those who say that new types of communities forming on social media are dividing us." Kaplan reportedly "killed efforts to build a classification system for hyperpolarized content." Failing to address this has resulted in algorithms that, as Tim Kendall explained, "have brought out the worst in us. They have literally rewired our brains so that we are detached from reality and immersed in tribalism" [38].

Facebook would have us believe that it has made great strides in confronting these problems over just the last two years, as Mr. LeCun has claimed. But at present, the burden of proof is on Facebook to produce the full, raw data so that independent researchers can make a fair assessment of his claims.

[5] Facebook claims to have since broadened the metrics it uses to calculate executive pay, but to what extent this might offset the prime directive of maximizing user engagement is unclear.

6 The AI filter

According to LeCun's tweets cited at the beginning of this paper, Facebook's AI-powered filter cleanses the platform of:

1. Hate speech;
2. Calls to violence;
3. Bullying; and
4. Disinformation that endangers public safety or the integrity of the democratic process

These are his words, so we will refer to them even while the actual definitions of hate speech, calls to violence, and other terms are potentially controversial and open to debate.

These claims are provably false. While "AI" (along with some very large, manual curation operations in developing countries) may effectively filter *some* of this content, at Facebook's scale, *some* is not enough.

Let's examine the claims a little closer.

6.1 Does Facebook actually filter out hate speech?

An investigation by the UK-based counter-extremist organization ISD (Institute for Strategic Dialog) found that Facebook's algorithm "actively promotes" Holocaust denial content [20]. The same organization, in another report, documents how Facebook's "delays or mistakes in policy enforcement continue to enable hateful and harmful content to spread through paid targeted ads." [17]. They go on to explain that "[e]ven when action is taken on violating ad content, such a response is often reactive and delayed, after hundreds, thousands, or potentially even millions of users have already been served those ads on their feeds."[6]

Zuckerberg admitted in April 2018 that hate speech in Myanmar was a problem, and pledged to act. Four months later, Reuters found more than "1000 examples of posts, comments, images and videos attacking the Rohingya or other Myanmar Muslims that were on Facebook" [45]. As recently as June 2020 there were reports [7] of troll farms using Facebook to intimidate opponents of Rodrigo Duterte in the Philippines with death threats and hateful comments.

6.2 Does Facebook actually filter out calls to violence?

The Sri Lankan government had to block access to Facebook "amid a wave of violence against Muslims … after Facebook ignored years of calls from both the government and civil society groups to control ethnonationalist accounts

[6] Ibid.

that spread hate speech and incited violence." [42] A report from the Center for Policy Alternatives in September 2014 detailed evidence of 20 hate groups in Sri Lanka, and informed Facebook. In March of 2018, Buzzfeed reported that "16 out of the 20 groups were still on Facebook".[7]

When former President Trump tweeted, in response to Black Lives Matters protests, when "the looting starts, the shooting starts," the message was liked and shared hundreds of thousands of times across Facebook and Instagram, even as other social networks such as Twitter flagged the message for its explicit incitement of violence [48] and prevented it from being retweeted.

Facebook played a pivotal role in the planning of the January 6th insurrection in the US, providing an unchecked platform for proliferation of the Big Lie, radicalization around this lie, and coordinated organization around explicitly-stated plans to engage in violent confrontation at the nation's capital on the outgoing president's behalf. Facebook's role in the deadly violence was far greater and more widespread than the role of Parler and the other fringe right-wing platforms that attracted so much attention in the aftermath of the attack [11].

6.3 Does Facebook actually filter out cyberbullying?

According to Enough Is Enough, a non-partisan, non-profit organization whose mission is "making the Internet safer for children and families," the answer is a resounding no. According to their most recent cyberbullying statistics, [10] 47% of young people have been bullied online, and the two most prevalent platforms are Instagram at 42% and Facebook at 37%.

In fact, Facebook is failing to protect children on a global scale. According to a UNICEF poll of children in 30 countries, one in every three young people says that they have been victimized by cyberbullying. And one in five says the harassment and threat of actual violence caused them to skip school. According to the survey, conducted in concert with the UN Special Representative of the Secretary-General (SRSG) on Violence against Children, "almost three-quarters of young people also said social networks, including Facebook, Instagram, Snapchat and Twitter, are the most common place for online bullying" [49].

6.4 Does Facebook actually filter out "disinformation that endangers public safety or the integrity of the democratic process?"

To list the evidence contradicting this point would be exhausting. Below are just a few examples:

- The Computational Propaganda Research Project found in their 2019 Global Inventory of Organized Social Media Manipulation that 70 countries had disinformation campaigns organized on social media in 2019, with Facebook as the top platform [6].
- A Facebook whistleblower produced a 6600 word memo detailing case after case of Facebook "abdicating responsibility for malign activities on its platform that could affect the political fate of nations outside the United States or Western Europe." [44]
- Facebook is ground-zero for anti-vaccination and pandemic misinformation, with the 26-min conspiracy theory film "Plandemic" going viral on Facebook in April 2020 and garnering tens of millions of views. Facebook's attempt to purge itself of anti-vaccination disinformation was easily thwarted when the groups guilty of proliferating this content removed the word "vaccine" from their names. In addition to undermining public health interests by spreading provably false content, these anti-vaccination groups have obscured meaningful discourse about the actual health concerns and risks that may or may not be connected to vaccinations. A paper from May 2020 attempts to map out the "multi-sided landscape of unprecedented intricacy that involves nearly 100 million individuals" [25] that are entangled with anti-vaccination clusters. That report predicts that such anti-vaccination views "will dominate in a decade" given their explosive growth and intertwining with undecided people. According to the Knight Foundation and Gallup, [26] 75% of Americans believe they "were exposed to misinformation about the election" on Facebook during the 2020 US presidential election. This is one of those rare issues on which Republicans (76%), Democrats (75%) and Independents (75%) agree–Facebook was the primary source for election misinformation.

If those AI filters are in fact working, they are not working very well.

All of this said, Facebook's reliance on "AI filters" misses a critical point, which is that you cannot have AI ethics without ethics [30]. These problems cannot be solved with AI. These problems cannot be solved with checklists, incremental advances, marginal changes, or even state-of-the-art deep learning networks. These problems are caused by the company's entire business model and mission. Bosworth's

[7] Ibid.

provocative quotes above, along with Tim Kendall's direct testimony demonstrate as much.

These are systemic issues, not technological ones. Yael Eisenstat put it best in her TED talk: "as long as the company continues to merely tinker around the margins of content policy and moderation, as opposed to considering how the entire machine is designed and monetized, they will never truly address how the platform is contributing to hatred, division and radicalization."

7 Facebook does not want to be the arbiter of truth

We should probably take comfort in Facebook's claim that it does not wish to be the "arbiter of political truth." After all, Facebook has a troubled history with the truth. Their ad buying customers proved as much when Facebook was forced to pay $40 million to settle a lawsuit alleging that they had inflated "by up to 900 percent—the time it said users spent watching videos." [4] While Facebook would neither admit nor deny the truth of this allegation, they did admit to the error in a 2016 statement [14].

This was not some innocuous lie that just cost a few firms some money either. As Slate explained in a 2018 article, "many [publications] laid off writers and editors and cut back on text stories to focus on producing short, snappy videos for people to watch in their Facebook feeds." [40] People lost their livelihoods to this deception.

Is this an isolated incident? Or is fraud at Facebook systemic? Matt Stoller describes the contents of recently unsealed legal documents [12] in a lawsuit alleging Facebook has defrauded advertisers for years [46]:

> The documents revealed that Facebook COO Sheryl Sandberg directly oversaw the alleged fraud for years. The scheme was simple. Facebook deceived advertisers by pretending that fake accounts represented real people, because ad buyers choose to spend on ad campaigns based on where they think their customers are. Former employees noted that the corporation did not care about the accuracy of numbers as long as the ad money was coming in. Facebook, they said, "did not give a shit."
>
> The inflated statistics sometimes led to outlandish results. For instance, Facebook told advertisers that its services had a potential reach of 100 million 18–34-year-olds in the United States, even though there are only 76 million people in that demographic. After employees proposed a fix to make the numbers honest, the corporation rejected the idea, noting that the "revenue impact" for Facebook would be "significant." One Facebook employee wrote, "My question lately is: how long can we get away with the reach overestimation?"
>
> According to these documents, Sandberg aggressively managed public communications over how to talk to advertisers about the inflated statistics, and Facebook is now fighting against her being interviewed by lawyers in a class action lawsuit alleging fraud.

Facebook's embrace of deception extends from its ad-buying fraud to the content on its platforms. For instance:

- Those who would "aid[] and abet[] the spread of climate misinformation" on Facebook benefit from "a giant loophole in its fact-checking program." Evidently, Facebook gives its staff the power to overrule climate scientists by deeming climate disinformation "opinion." [2].
- The former managing editor of Snopes reported that Facebook was merely using the well-regarded fact-checking site for "crisis PR," that they did not take fact checking seriously and would ignore concerns [35]. Snopes tried hard to push against the Myanmar disinformation campaign, amongst many other issues, but its concerns were ignored.
- ProPublica recently reported [18] that Sheryl Sandberg silenced and censored a Kurdish militia group that "the Turkish government had targeted" in order to safeguard their revenue from Turkey.
- Mark Zuckerberg and Joel Kaplan intervened [37] in April 2019 to keep Alex Jones on the platform, despite the right-wing conspiracy theorist's lead role in spreading disinformation about the 2012 Sandy Hook elementary school shooting and the 2018 Parkland high school shooting.

Arguably, Facebook's executive team has not only ceded responsibility as an "arbiter of truth," but has also on several notable occasions, intervened to ensure the continued proliferation of disinformation.

8 How do we disengage?

Facebook's business model is focused entirely on increasing growth and user engagement. Its algorithms are extremely effective at doing so. The steps Facebook has taken, such as building "AI filters" or partnering with independent fact checkers, are superficial and toothless. They cannot begin to untangle the systemic issues at the heart of this matter, because these issues are Facebook's entire reason for being.

So what can be done? Certainly, criminality needs to be prosecuted. Executives should go to jail for fraud. Social media companies, and their organizational leaders, should face legal liability for the impact made by the content on

their platforms. One effort to impose legal liability in the US is centered around reforming section 230 of the US Communications Decency Act. It, and similar laws around the world, should be reformed to create far more meaningful accountability and liability for the promotion of disinformation, violence, and extremism.

Most importantly, monopolies should be busted. Existing antitrust laws should be used to break up Facebook and restrict its future activities and acquisitions.

The matters outlined here have been brought to the attention of Facebook's leadership in countless ways that are well documented and readily provable. But the changes required go well beyond effective leveraging of AI. At its heart, Facebook will not change because they do not want to, and are not incentivized to. Facebook must be regulated, and Facebook's leadership structure must be dismantled.

It seems unlikely that politicians and regulators have the political will to do all of this, but there are some encouraging signs, especially regarding antitrust investigations [9] and lawsuits [28] in both the US and Europe. Still, this issue goes well beyond mere enforcement. Somehow we must shift the incentives for social media companies, who compete for, and monetize, our attention. Until we stop rewarding Facebook's illicit behavior with engagement, it's hard to see a way out of our current condition. These companies are building technology that is designed to draw us in with problematic content, addict us to outrage, and ultimately drive us apart. We no longer agree on shared facts or truths, a condition that is turning political adversaries into bitter enemies, that is transforming ideological difference into seething contempt. Rather than help us lead more fulfilling lives or find truth, Facebook is helping us to discover enemies among our fellow citizens, and bombarding us with reasons to hate them, all to the end of profitability. This path is unsustainable.

The only thing Facebook truly understands is money, and all of their money comes from engagement. If we disengage, they lose money. If we delete, they lose power. If we decline to be a part of their ecosystem, perhaps we can collectively return to a shared reality.

References

1. Allcot, H., et al.: "The Welfare Effects of Social Media." (2019). https://web.stanford.edu/~gentzkow/research/facebook.pdf
2. Atkin, E.: Facebook creates fact-checking exemption for climate deniers. *Heated*. (2020). https://heated.world/p/facebook-creates-fact-checking-exemption
3. Auxier, B., Anderson, M.: Social Media Use in 2021. Pew Research Center. (2021). https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf
4. Baron, E.: Facebook agrees to pay $40 million over inflated video-viewing times but denies doing anything wrong. *The Mercury News*. (2019). https://www.mercurynews.com/2019/10/07/facebook-agrees-to-pay-40-million-over-inflated-video-viewing-times-but-denies-doing-anything-wrong/
5. Boxell, L.: "The internet, social media, and political polarisation." (2017). https://voxeu.org/article/internet-social-media-and-political-polarisation
6. Bradshaw, S., Howard, P.N.: The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation. Working Paper 2019.2. Oxford: Project on Computational Propaganda. (2019)
7. Cabato, R.: Death threats, clone accounts: Another day fighting trolls in the Philippines. *The Washington Post*. (2020). https://www.washingtonpost.com/world/asia_pacific/facebook-trolls-philippines-death-threats-clone-accounts-duterte-terror-bill/2020/06/08/3114988a-a966-11ea-a43b-be9f6494a87d_story.html
8. Card, C.: "How Facebook AI Helps Suicide Prevention." Facebook. (2018). https://about.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/
9. Chee, F.Y.: "Facebook in EU antitrust crosshairs over data collection." Reuters. (2019). https://www.reuters.com/article/us-eu-facebook-antitrust-idUSKBN1Y625J
10. Cyberbullying Statistics. Enough Is Enough. https://enough.org/stats_cyberbullying
11. Dwoskin, E.: Facebook's Sandberg deflected blame for Capitol riot, but new evidence shows how platform played role. *The Washington Post*. (2021). https://www.washingtonpost.com/technology/2021/01/13/facebook-role-in-capitol-protest
12. DZ Reserve and Cain Maxwell v. Facebook, Inc. (2020). https://www.economicliberties.us/wp-content/uploads/2021/02/2021.02.17-Unredacted-Opp-to-Mtn-to-Dismiss.pdf
13. Eisenstat, Y.: Dear Facebook, this is how you're breaking democracy [Video]. *TED*. (2020). https://www.ted.com/talks/yael_eisenstat_dear_facebook_this_is_how_you_re_breaking_democracy#t-385134
14. Fischer, D.: Facebook Video Metrics Update. *Facebook*. (2016). https://www.facebook.com/business/news/facebook-video-metrics-update
15. Fisher, M., Taub, A.: "How Everyday Social Media Users Become Real-World Extremists." *New York Times*. (2018). https://www.nytimes.com/2018/04/25/world/asia/facebook-extremism.html
16. Frenkel, S.: "How Misinformation 'Superspreaders' Seed False Election Theories". *New York Times*. (2020). https://www.nytimes.com/2020/11/23/technology/election-misinformation-facebook-twitter.html
17. Gallagher, A.: *Profit and Protest: How Facebook is struggling to enforce limits on ads spreading hate, lies and scams about the Black Lives Matter protests*. The Institute for Strategic Dialogue (2020)
18. Gillum, J., Ellion, J.: Sheryl Sandberg and Top Facebook Execs Silenced an Enemy of Turkey to Prevent a Hit to the Company's Business. *ProPublica*. (2021). https://www.propublica.org/article/sheryl-sandberg-and-top-facebook-execs-silenced-an-enemy-of-turkey-to-prevent-a-hit-to-their-business
19. Gonzalez, R.: "Facebook Opens Its Private Servers to Scientists Studying Fake News." *Wired*. (2018). https://www.wired.com/story/social-science-one-facebook-fake-news/
20. Guhl, J., Davey, J.: *Hosting the 'Holohoax': A Snapshot of Holocaust Denial Across Social Media*. The Institute for Strategic Dialogue (2020).
21. Hao, K.: "How Facebook got addicted to spreading misinformation". *MIT Technology Review*. (2021). https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation
22. Horwitz, J., Seetharaman, D.: "Facebook Executives Shut Down Efforts to Make the Site Less Divisive." *Wall St Journal* (2020)

23. Internet use and political polarization, Boxell, L., Gentzkow, M., Shapiro, J.M.: Proc Natl. Acad. Sci. **114**(40), 10612–10617 (2017). https://doi.org/10.1073/pnas.1706588114
24. Johnson, S.L., et al.: Understanding echo chambers and filter bubbles: the impact of social media on diversification and partisan shifts in news consumption. MIS Q. (2020). https://doi.org/10.25300/MISQ/2020/16371
25. Johnson, N.F., Velásquez, N., Restrepo, N.J., et al.: The online competition between pro- and anti-vaccination views. Nature **582**, 230–233 (2020). https://doi.org/10.1038/s41586-020-2281-1
26. Jones, J.: In Election 2020, How Did The Media, Electoral Process Fare? Republicans, Democrats Disagree. *Knight Foundation*. (2020). https://knightfoundation.org/articles/in-election-2020-how-did-the-media-electoral-process-fare-republicans-democrats-disagree
27. Kantrowitz, A.: "Facebook Is Still Prioritizing Scale Over Safety." *Buzzfeed.News*. (2019). https://www.buzzfeednews.com/article/alexkantrowitz/after-years-of-scandal-facebooks-unhealthy-obsession-with
28. Kendall, B., McKinnon, J.D.: "Facebook Hit With Antitrust Lawsuits by FTC, State Attorneys General." Wall St. Journal. (2020). https://www.wsj.com/articles/facebook-hit-with-antitrust-lawsuit-by-federal-trade-commission-state-attorneys-general-11607543139
29. Lauer, D.: [@dlauer]. *And yet people believe them because of misinformation that is spread and monetized on facebook* [Tweet]. Twitter. (2021). https://twitter.com/dlauer/status/1363923475040251905
30. Lauer, D.: You cannot have AI ethics without ethics. AI Ethics **1**, 21–25 (2021). https://doi.org/10.1007/s43681-020-00013-4
31. Lavi, M.: Do Platforms Kill? Harvard J. Law Public Policy. **43**(2), 477 (2020). https://www.harvard-jlpp.com/wp-content/uploads/sites/21/2020/03/Lavi-FINAL.pdf
32. LeCun, Y.: [@ylecun]. *Does anyone still believe whatever these people are saying? No one should. Believing them kills* [Tweet]. Twitter. (2021). https://twitter.com/ylecun/status/1363923178519732230
33. LeCun, Y.: [@ylecun]. *The section about FB in your article is factually wrong. For starter, AI is used to filter things like hate speech, calls to violence, bullying, child exploitation, etc. Second, disinformation that endangers public safety or the integrity of the democratic process is filtered out* [Tweet]. Twitter. (2021). https://twitter.com/ylecun/status/1364010548828987393
34. LeCun, Y.: [@ylecun]. As attractive as it may seem, this explanation is false. [Tweet]. Twitter. (2021). https://twitter.com/ylecun/status/1363985013147115528
35. Levin, S.: 'They don't care': Facebook factchecking in disarray as journalists push to cut ties. *The Guardian*. (2018). https://www.theguardian.com/technology/2018/dec/13/they-dont-care-facebook-fact-checking-in-disarray-as-journalists-push-to-cut-ties
36. Mac, R.: "Growth At Any Cost: Top Facebook Executive Defended Data Collection In 2016 Memo—And Warned That Facebook Could Get People Killed." *Buzzfeed.News*. (2018). https://www.buzzfeednews.com/article/ryanmac/growth-at-any-cost-top-facebook-executive-defended-data
37. Mac, R., Silverman, C.: "Mark Changed The Rules": How Facebook Went Easy On Alex Jones And Other Right-Wing Figures. *BuzzFeed.News*. (2021). https://www.buzzfeednews.com/article/ryanmac/mark-zuckerberg-joel-kaplan-facebook-alex-jones
38. Mainstreaming Extremism: Social Media's Role in Radicalizing America: Hearings before the Subcommittee on Consumer Protection and Commerce of the Committee on Energy and Commerce, 116th Cong. (2020) (testimony of Tim Kendall)
39. Meade, A.: "Facebook greatest source of Covid-19 disinformation, journalists say". *The Guardian*. (2020). https://www.theguardian.com/technology/2020/oct/14/facebook-greatest-source-of-covid-19-disinformation-journalists-say
40. Oremus, W.: The Big Lie Behind the "Pivot to Video". *Slate*. (2018). https://slate.com/technology/2018/10/facebook-online-video-pivot-metrics-false.html
41. Propagating and Debunking Conspiracy Theories on Twitter During the 2015–2016 Zika Virus Outbreak, Michael J. Wood, Cyberpsychology, Behavior, and Social Networking. **21**(8), (2018). https://doi.org/10.1089/cyber.2017.0669
42. Rajagopalan, M., Nazim, A.: "We Had To Stop Facebook": When Anti-Muslim Violence Goes Viral. *BuzzFeed.News*. (2018). https://www.buzzfeednews.com/article/meghara/we-had-to-stop-facebook-when-anti-muslim-violence-goes-viral
43. Rosalsky, G.: "Are Conspiracy Theories Good For Facebook?". *Planet Money*. (2020). https://www.npr.org/sections/money/2020/08/04/898596655/are-conspiracy-theories-good-for-facebook
44. Silverman, C., Mac, R.: "I Have Blood on My Hands": A Whistleblower Says Facebook Ignored Global Political Manipulation. *BuzzFeed.News*. (2020). https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo
45. Stecklow, S.: Why Facebook is losing the way on hate speech in Myanmar. *Reuters*. (2018). https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/
46. Stoller, M.: *Facebook: What is the Australian law? And why does FB keep getting caught for fraud?*. Substack. (2021). https://mattstoller.substack.com/p/facecrook-dealing-with-a-global-menace
47. The spread of true and false news online, Soroush Vosoughi, Deb Roy, Sinan Aral, Science. **359**(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
48. The White House 45 Archived [@WhiteHouse45]: *"These THUGS are dishonoring the memory of George Floyd, and I won't let that happen. Just spoke to Governor Tim Walz and told him that the Military is with him all the way. Any difficulty and we will assume control but, when the looting starts, the shooting starts. Thank you!"* [Tweet]. Twitter. (2020) https://twitter.com/WhiteHouse45/status/1266342941649506304
49. UNICEF.: UNICEF poll: More than a third of young people in 30 countries report being a victim of online bullying. (2019). https://www.unicef.org/press-releases/unicef-poll-more-third-young-people-30-countries-report-being-victim-online-bullying