House Committee on Energy and Commerce
Disinformation Online and a Country in Crisis

Hany Farid, Ph.D.

Testimony

Overview

Technology and the internet have had a remarkable impact on our lives and society. Many educational, entertaining, and inspiring things have emerged from the past two decades in innovation. At the same time, many horrific things have emerged: a massive proliferation of child sexual abuse material [21], the spread and radicalization of domestic and international terrorists [13], the distribution of illegal and deadly drugs [36, 11], the proliferation of mis- and disinformation campaigns designed to sow civil unrest, incite violence, and disrupt democratic elections [5, 28], the proliferation of dangerous, hateful, and deadly conspiracy theories [33, 27, 10], the routine harassment of women and under-represented groups in the form of threats of sexual violence and revenge and non-consensual pornography [16], small- to large-scale fraud [40], and spectacular failures to protect our personal and sensitive data [17].

How, in 20 short years, did we go from the promise of the internet to democratize access to knowledge and make the world more understanding and enlightened, to this litany of daily horrors? Due to a combination of naivete, ideology, willful ignorance, and a mentality of growth at all costs, the titans of tech have simply failed to install proper safeguards on their services.

We can and we must do better when it comes to contending with some of the most violent, harmful, dangerous, hateful, and fraudulent content online. We can and we must do better when it comes to contending with the misinformation apocalypse that has emerged over the past few years.

Facebook's Mark Zuckerberg has tried to frame the issue of reining in mis- and disinformation as not wanting to be "the arbiter of truth" [26]. This entirely misses the point. The point is not about truth or falsehood, but about algorithmic amplification. The point is that social media decides every day what is relevant by recommending it to their billions of users. The point is that social media has learned that outrageous, divisive, and conspiratorial content increases engagement [18]. The point is that online content providers could simply decide that they value trusted information over untrusted information, respectful over hateful, and unifying over divisive, and in turn fundamentally change the divisiveness-fueling and misinformation-distributing machine that is social media today.

By way of highlighting the depth and breadth of these problems, I will describe two recent case studies that reveal a troubling pattern of how the internet, social media, and

more generally, information, is being weaponized against society and democracy. I will conclude with a broad overview of interventions to help avert the digital dystopia that we seem to be hurtling towards.

COVID-related Misinformation

The COVID-19 global pandemic has been an ideal breeding ground for online misinformation: Social-media traffic has reached an all-time record [9] as people are forced to remain at home, often idle, anxious, and hungry for information [32], while at the same time, social-media services are unable to fully rely on human moderators to enforce their rules [7]. The resulting spike in COVID-19 related misinformation is of grave concern to health professionals [3]. The World Health Organization has listed the need for surveys and qualitative research about the infodemic in its top priorities to contain the pandemic [43].

Between April 11, and April 21, 2020, we launched a U.S.-based survey to examine the belief in 20 prevalent COVID-19 related false statements, and 20 corresponding true statements [28]. Our results, from 500 respondents, reveal a troubling breadth and depth of misinformation that is highly partisan, and fueled by social media.

On average, 55.7%/29.8% of true/false statements reached participants, of which 57.8%/10.9% are believed. When participants are asked if they know someone that believes or is likely to believe a statement, 71.4%/42.7% of the true/false statements are believed by others known to the participant. The median number of true/false statements that reached a participant is 11/6; the median number of true/false statements believed by a participant is 12/2; the median number of true/false statements believed by others known to the participant is 15/8; and 31% claimed to believe at least one false conspiracy (cf. [14]).

It is generally encouraging that true statements have a wider reach and wider belief than false statements. The reach and belief in false statements, however, is still troubling, particularly given the potentially deadly consequences that might arise from misinformation. Even more troubling is the partisan divide that emerges upon closer examination of the data.

We found that political leaning and main news source had a significant effect on the likelihood of the number of false statements that are believed. The number of false statements believed by those with social media as their main source of news is 1.41 times greater than those who cited another main news source. The number of false statements believed by those on the right of the political spectrum is 2.15 times greater than those on the left. And, those on the right were 1.12 times less likely to believe true statements than those on the left.

The five largest effects were based on political leaning, where, as compared to those on the left, those on the right are:

- 15.44 times more likely to believe that "asymptomatic carriers of COVID-19 who die of other medical problems are added to the coronavirus death toll to get the numbers up to justify this pandemic response."

- 14.70 times more likely to believe that "House Democrats included $25 million to boost their own salaries in their proposal for the COVID-19 related stimulus package."

- 9.44 times more likely to believe that "COVID-19 was man-made in a lab and is not thought to be a natural virus."

- 7.41 times more likely to believe that "Silver solution kills COVID-19."

- 6.97 times more likely to believe that "COVID stands for *Chinese Originated Viral Infectious Disease.*"

It remains unclear the extent to which COVID-19 misinformation is a result of coordinated attacks, or has arisen organically through mis-understanding and fear. It also remains unclear if the spread and belief in this misinformation is on the rise or decline, and how it has impacted other parts of the world. We and others are actively pursuing answers to each of these questions. What is clear is that, fueled by social media, COVID-related misinformation has spread widely and deeply, and the online content providers have done a poor job of controlling the spread of misinformation, and are complicit in fueling its spread.

YouTube's Conspiracies

By allowing for a wide range of opinions to coexist, social media has allowed for an open exchange of ideas. There have, however, been concerns that the recommendation engines which power these services amplify sensational content because of its tendency to generate more engagement. The algorithmic promotion of conspiracy theories by YouTube's recommendation engine, in particular, has recently been of growing concern to academics [25, 4, 39, 31, 35, 41, 2], legislators [24], and the public [15, 1, 30, 34, 8, 3]. In August 2019, the FBI introduced fringe conspiracy theories as a domestic terrorist threat, due to the increasing number of violent incidents motivated by such beliefs [12].

Some 70% of watched content on YouTube is recommended content [38], in which YouTube algorithms promote videos based on a number of factors including optimizing for user-engagement or view-time. Because conspiracy theories generally feature novel and provoking content, they tend to yield higher than average engagement [19]. The recommendation algorithms are thus vulnerable to sparking a reinforcing feedback loop [44] in which more conspiracy theories are recommended and consumed [6].

We analysed more than 8 million recommendations from YouTube's watch-next algorithm over 20 months. Recommendations were collected daily, starting from the most recent videos posted by a set of 1000+ of the most popular news and informational channels in the U.S. The recommended videos were then fed to a classifier trained to detect conspiratorial content.

Shown in Figure 1 is our estimate of the percentage of conspiratorial videos recommended by YouTube on information channels, between October 2018 and June 2020. Starting in April 2019, we monitored a consistent decrease in conspiratorial recommendations until the beginning of June 2019 when the frequency briefly hit a low point of 3%. Since that time, however, the frequency has hovered between 3% and 5%, and saw a steady uptick starting in March of 2020 as the scale of the global COVID-19 pandemic became apparent.

This longitudinal study revealed a troubling frequency of active promotion of conspiratorial content, and, despite previous claims to the contrary, YouTube was able to reduce the prevalence of this content by adjusting their recommendation algorithms. This reduction, however, does not make the problem of radicalization on YouTube obsolete nor fictional, as some have claimed [22]. Aggregate data hide very different realities for individuals, and although radicalization is a serious issue, it is only relevant for a fraction of the users. Those
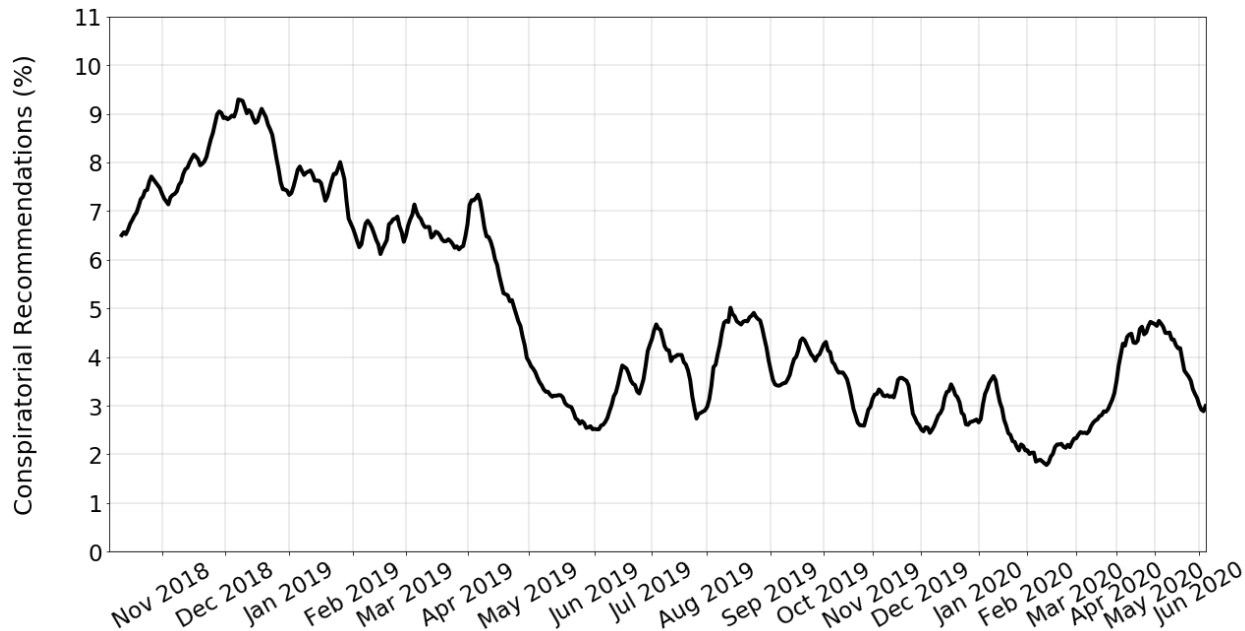
Figure 1: Longitudinal trend of conspiratorial recommendations from informational channels on YouTube, in which each data point corresponds to a rolling seven day average.

with a history of watching conspiratorial content can certainly still experience YouTube as a filter-bubble, reinforced by personalized recommendations and channel subscriptions.

With two billion monthly active users on YouTube, the design of the recommendation algorithm has more impact on the flow of information than the editorial boards of traditional media. The role of this engine is made even more crucial in the light of (1) the increasing use of YouTube as a primary source of information, particularly among the youth [37]; (2) the nearly monopolistic position of YouTube on its market; and (3) the ever-growing weaponization of YouTube to spread misinformation and partisan content around the world [5].

Political Misinformation

The misinformation and conspiracy studies described above are the canaries in the coal mine as we approach the 2020 U.S. national election. The internet, and social media in particular, has been weaponized by individuals, organizations, and nation-states to sow civil unrest and interfere with democratic elections. We should have a vigorous debate about the issues that we face as a country, but we should not allow social media to be weaponized against our society and democracy in the form of voter suppression, and spreading and amplifying lies, conspiracies, hate, and fear.

Interventions

The internet, and social media in particular, is failing us on an individual, societal, and democratic level. Online content providers have prioritized growth, profit, and market dominance over creating a safe and healthy online ecosystem. These providers have taken the position that they are simply in the business of hosting user-generated content and are not

and should not be asked to be "the arbiters of truth." This, however, defies the reality of social media today where the vast majority of delivered content is actively promoted by content providers based on their algorithms that are designed in large part to maximize engagement and revenue. These algorithms have access to highly detailed profiles of billions of users, allowing the algorithm to micro-target content in unprecedented ways. These algorithms have learned that divisive, hateful, and conspiratorial content engages users and so this type of content is prioritized, leading to rampant misinformation and conspiracies and, in turn, increased anger, hate, and intolerance, both online and offline.

Many want to frame the issue of content moderation as an issue of freedom of speech. It is not. First, private companies have the right to regulate content on their services without running afoul of the first amendment, as many routinely do when they ban legal adult pornography. Second, the issue of content moderation should focus not on content removal but on the underlying algorithms that determine what is relevant and what we see, read, and hear online. It is these algorithms that are at the core of the misinformation amplification. Snap's CEO, Evan Spiegel, for example, recently announced "We will make it clear with our actions that there is no gray area when it comes to racism, violence and injustice. We simply cannot promote accounts in America that are linked to people who incite racial violence, whether they do so on or off our platform." But Mr. Spiegel also announced that the company would not necessarily remove specific content or accounts [20]. It is precisely this amplification and promotion that should be at the heart of the discussion of online content moderation.

If online content providers prioritized their algorithms to value trusted information over untrusted information, respectful over hateful, and unifying over divisive, we could move from a divisiveness-fueling and misinformation-distributing machine that is social media today, to a healthier and more respectful online ecosystem. If advertisers, that are the fuel behind social media, took a stand against online abuses, they could withhold their advertising dollars to insist on real change.

Standing in the way of this much needed change is a lack of corporate leadership, a lack of competition, a lack of regulatory oversight, and a lack of education among the general public. Responsibility, therefore, falls on the private sector, government regulators, and we the general public.

Biography

Hany Farid is a Professor at the University of California, Berkeley with a joint appointment in Electrical Engineering & Computer Sciences and the School of Information. His research focuses on digital forensics, image analysis, and human perception. He received his undergraduate degree in Computer Science and Applied Mathematics from the University of Rochester in 1989, his Ph.D. in Computer Science from the University of Pennsylvania in 1997. Following a two-year post-doctoral fellowship in Brain and Cognitive Sciences at MIT, he joined the faculty at Dartmouth College in 1999 where he remained until 2019. He is the recipient of an Alfred P. Sloan Fellowship, a John Simon Guggenheim Fellowship, and is a Fellow of the National Academy of Inventors.

References

[1] Fiction is outperforming reality.

[2] J. Allgaier. Strategically distorted communications in online videos on climate change and climate engineering. *Frontiers in Communication*, 4:36, 2019.

[3] Avaaz. Why is YouTube broadcasting climate misinformation to millions?, Jan 2020.

[4] A. Bessi, F. Zollo, M. D. Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi. User's polarization on Facebook and YouTube. *Plos One*, 11(8), 2016.

[5] S. Bradshaw and P. Howard. The global disinformation order. *Computational Propaganda Research Project*, Sep 2019.

[6] G. Chaslot. How YouTube's A.I. boosts alternative facts. *Medium*, Apr 2017.

[7] E. Dwoskin and N. Tiku. Facebook sent home thousands of human moderators due to the coronavirus. now the algorithms are in charge. The Washington Post, Mar 2020.

[8] E. G. Ellis. Online conspiracy theories. *Wired*, Oct 2018.

[9] Facebook Newsroom. Keeping our services stable and reliable during the COVID-19 outbreak, Apr 2020.

[10] M. Faddoul, G. Chaslot, and H. Farid. A longitudinal analysis of YouTube's promotion of conspiracy videos. *arXiv*, (2003.03318), 2020.

[11] H. Farid and M. Falco. Google is not cracking down on the most dangerous drug in America. Newsweek, June 2020.

[12] FBI Phoenix Field Office. Anti-government, identity based, and fringe political conspiracy theories very likely motivate some domestic extremists to commit criminal, sometimes violent activity, May 2019.

[13] M. Fisher and A. Taub. How everyday social media users become real-world extremists. New York Times, Apr 2018.

[14] D. Freeman, F. Waite, L. Rosebrock, A. Petit, C. Causier, A. East, L. Jenner, A.-L. Teale, L. Carr, S. Mulhall, and et al. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in england. *Psychological Medicine*, page 1–30, 2020.

[15] B. Geurkink. Our recommendation to YouTube. *Mozilla Foundation*, 2019.

[16] S. Haynes. 'a toxic place for women.' a new study reveals the scale of abuse on Twitter. Time, Dec 2018.

[17] V. Ho. Facebook's privacy problems: a roundup. The Guardian, Dec 2018.

[18] J. Horwitz and D. Seetharaman. Facebook executives shut down efforts to make the site less divisive. Wall Street Journal, May 2020.

[19] M. N. Hussain, S. Tokdemir, N. Agarwal, and S. Al-Khateeb. Analyzing disinformation and crowd manipulation tactics on YouTube. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug 2018.

[20] C. Kang and K. Conger. Snap says it will no longer promote Trump's account. New York Times, Jun 2020.

[21] M. Keller and G. Dance. The internet is overrun with images of child sexual abuse. what went wrong? New York Times, Sep 2019.

[22] M. Ledwich and A. Zaitsev. Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *arXiv*, (1912.11211), 2019.

[23] S. Lewandowsky, U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3):106–131, 2012.

[24] P. Lewis. Senator warns YouTube algorithm may be open to manipulation by 'bad actors'. *The Guardian*, Feb 2018.

[25] R. Lewis. Alternative influence. *Data & Society*, Sep 2018.

[26] T. McCarthy. Zuckerberg says Facebook won't be 'arbiters of truth' after Trump threat. The Guardian, May 2020.

[27] J. Nicas. Can YouTube quiet its conspiracy theorists? New York Times, March 2020.

[28] S. Nightingale, M. Faddoul, and H. Farid. Quantifying the reach and belief in COVID-19 misinformation, 2020.

[29] B. Nyhan and J. Reifler. When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330, 2010.

[30] C. O'Donovan, C. Warzel, L. McDonald, B. Clifton, and M. Woolf. We followed YouTube's recommendation algorithm down the rabbit hole. *BuzzFeed News*, Jan 2019.

[31] D. O'Callaghan, D. Greene, M. Conway, J. Carthy, and P. Cunningham. Down the (white) rabbit hole: The extreme right and online recommender systems. *Social Science Computer Review*, 33(4):459–478, 2014.

[32] A. Pakpour and M. Griffiths. The fear of COVID-19 and its role in preventive behaviors. *Journal of Concurrent Disorders*, 2020.

[33] B. Resnick. Social media's conspiracy theory problem isn't going away. Vox, Aug 2019.

[34] K. Roose. Youtube unleashed a conspiracy theory boom, can it be contained? *The New York Times*, Feb 2019.

[35] M. Samory and T. Mitra. Conspiracies online: User discussions in a conspiracy community following dramatic events. *ICWSM 2018*, 2018.

[36] D. Scott. This is how easy it is to order deadly opioids over the internet. Vox, Jan 2018.

[37] A. Smith, S. Toor, and P. v. Kessel. Many turn to YouTube for children's content, News, How-To Lessons. *Pew Research Center: Internet, Science and Tech*, Nov 2018.

[38] J. E. Solsman. YouTube's AI is the puppet master over most of what you watch. *CNET*, Jan 2018.

[39] M. Y.-J. Song and A. Gruzd. Examining sentiments and popularity of pro- and anti-vaccination videos on YouTube. *Proceedings of the 8th International Conference on Social Media and Society*, Jul 2017.

[40] S. Tompor. Scammers are fooling millennials out of millions of dollars: Here's how. Detroit Free Press, Oct 2019.

[41] Z. Tufekci. YouTube, the great radicalizer. *The New York Times*, Mar 2018.

[42] S. Vosoughi, D. Roy, and S. Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[43] World Health Organisation. A coordinated global research roadmap, 2020.

[44] Z. Zhao, L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy, X. Yi, and E. Chi. Recommending what video to watch next: A multitask ranking system. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 43–51, 2019.