

**ALGORITHMS AND AMPLIFICATION: HOW
SOCIAL MEDIA PLATFORMS' DESIGN CHOICES
SHAPE OUR DISCOURSE AND OUR MINDS**

HEARING

BEFORE THE

SUBCOMMITTEE ON PRIVACY,
TECHNOLOGY, AND THE LAW

OF THE

COMMITTEE ON THE JUDICIARY
UNITED STATES SENATE

ONE HUNDRED SEVENTEENTH CONGRESS

FIRST SESSION

APRIL 27, 2021

Serial No. J-117-14

Printed for the use of the Committee on the Judiciary



www.judiciary.senate.gov
www.govinfo.gov

U.S. GOVERNMENT PUBLISHING OFFICE

WASHINGTON : 2026

COMMITTEE ON THE JUDICIARY

RICHARD J. DURBIN, Illinois, *Chair*

PATRICK J. LEAHY, Vermont	CHARLES E. GRASSLEY, Iowa, <i>Ranking Member</i>
DIANNE FEINSTEIN, California	LINDSEY O. GRAHAM, South Carolina
SHELDON WHITEHOUSE, Rhode Island	JOHN CORNYN, Texas
AMY KLOBUCHAR, Minnesota	MICHAEL S. LEE, Utah
CHRISTOPHER A. COONS, Delaware	TED CRUZ, Texas
RICHARD BLUMENTHAL, Connecticut	BEN SASSE, Nebraska
MAZIE K. HIRONO, Hawaii	JOSH HAWLEY, Missouri
CORY A. BOOKER, New Jersey	TOM COTTON, Arkansas
ALEX PADILLA, California	JOHN KENNEDY, Louisiana
JON OSSOFF, Georgia	THOM TILLIS, North Carolina
	MARSHA BLACKBURN, Tennessee

JOSEPH ZOGBY, *Chief Counsel and Staff Director*

KOLAN L. DAVIS, *Republican Chief Counsel and Staff Director*

SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW

CHRISTOPHER A. COONS, Delaware, *Chair*

SHELDON WHITEHOUSE, Rhode Island	BEN SASSE, Nebraska, <i>Ranking Member</i>
AMY KLOBUCHAR, Minnesota	LINDSEY O. GRAHAM, South Carolina
MAZIE K. HIRONO, Hawaii	JOSH HAWLEY, Missouri
JON OSSOFF, Georgia	JOHN KENNEDY, Louisiana
	MARSHA BLACKBURN, Tennessee

SOPHIE BRILL, *Democratic Staff Counsel*

WILLIAM PAYNE, *Republican Staff Counsel*

CONTENTS

OPENING STATEMENTS

	Page
Coons, Hon. Christopher A.	1
Sasse, Hon. Ben	2
Durbin, Hon. Richard J.	3

WITNESSES

Bickert, Monika	6
Prepared statement	43
Questions submitted with no response returned	75
Culbertson, Lauren	9
Prepared statement	48
Questions submitted with no response returned	84
Donovan, Joan	13
Prepared statement	52
Questions submitted with no response returned	89
Harris, Tristan	11
Prepared statement	59
Questions submitted with no response returned	91
Veitch, Alexandra	8
Prepared statement	61
Responses to written questions	93

**ALGORITHMS AND AMPLIFICATION: HOW
SOCIAL MEDIA PLATFORMS' DESIGN CHOICES
SHAPE OUR DISCOURSE AND OUR MINDS**

TUESDAY, APRIL 27, 2021

UNITED STATES SENATE,
SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW,
COMMITTEE ON THE JUDICIARY,
Washington, DC.

The Subcommittee met, pursuant to notice, at 10:03 a.m., in Room 226, Dirksen Senate Office Building, Hon. Christopher A. Coons, Chair of the Subcommittee, presiding.

Present: Senators Coons [presiding], Klobuchar, Hirono, Ossoff, Sasse, Hawley, Kennedy, and Blackburn.

Also present: Senators Durbin, Grassley, and Blumenthal.

**OPENING STATEMENT OF HON. CHRISTOPHER A. COONS,
A U.S. SENATOR FROM THE STATE OF DELAWARE**

Chair COONS. This hearing will come to order.

Thank you to all of our witnesses for participating today, and a particular thank you to Mr. Harris, who I understand is joining us from Hawaii where it is 4 o'clock in the morning. I would also like to thank Ranking Member Senator Ben Sasse for working with me to put this hearing together. I am truly grateful we have been able to work together on this important topic. It is too important to let it fall victim to the trap of the typical partisan gridlock here in Washington. Thank you as well to Chairman Durbin for attending today.

Generally, when people hear the term "algorithm," you might think of some very complicated mathematical formula or piece of computer code. As many of us have become increasingly aware, algorithms impact what literally billions of people read and watch, and impact what they think every day. Facebook, Twitter, YouTube, the three major tech companies represented in today's hearing, use algorithms to determine what appears on your screen when you open and engage with their applications, and there is nothing inherently wrong about that.

With billions or even trillions of pieces of content to choose from on each platform, it makes sense that they should have a way to help us sift through what they think their users are looking for and what we are actually seeking. Advances in machine learning that made this technology possible have led to enormous good in other contexts. Machine learning has driven innovation across many industries from medical science to public transportation, and has al-

lowed companies to deliver better services. Many have also recently argued this advanced technology is harnessed into algorithms designed to attract our time and attention on social media, and the results can be harmful to our kids' attention spans, to the quality of our public discourse, to our public health, and even to our democracy itself.

What happens when algorithms become so good at amplification, at showing you content that a computer thinks you will like, that you or your kids or your family members end up spending hours each day engaged, staring at the screen? What happens algorithms become so hyper tailored to you and your habits and interests that you stop being exposed to ideas you might find disagreeable, or even so different from you—yours as to be offensive? What happens when they amplify content that might be very popular, but is also hateful or just plain false?

As I noted, Ranking Member Sasse and I worked on this hearing, and one of the main reasons for that is because we truly do not see these as partisan questions and do not come to this hearing with a specific regulatory or legislative agenda, but this is an area that requires urgent attention. As Mark Zuckerberg himself recently put it, and I quote, “When left unchecked, people will engage disproportionately with sensationalist and provocative content, which can undermine the quality of public discourse and lead to civic polarization.” If we are so polarized and angry and we can no longer hear each other's points of view, then our democracy itself suffers.

As quaint as some might think it, Ranking Member Sasse and I plan to use this hearing as an opportunity to learn about how these companies' algorithms work, what steps may have been taken to reduce algorithmic amplification that is harmful, and what can be done better so we can build on that knowledge in considering a potential path forward, whether voluntary, regulatory, or legislatively.

I look forward to hearing from the representatives of Facebook, Twitter, and YouTube, who have agreed to testify. Each of these platforms has taken a number of measures in recent years to curb some of the harms that algorithmic amplification can cause. It is also my hope that these platforms can build upon good practices, learn from each other, and make a significant difference. We will also hear from two outside experts, who can help us ask some bigger-picture questions and to narrow in on some of the strategies and tactics we could or should follow moving forward, including whether and how legislation might improve the practices that these and many other platforms use.

Thank you, and I am now going to turn to my Ranking Member, Senator Sasse, for his opening remarks.

**OPENING STATEMENT OF HON. BEN SASSE,
A U.S. SENATOR FROM THE STATE OF NEBRASKA**

Senator SASSE. Thank you, Chairman Coons. Congratulations on having a gavel for the first time in 6 years. Hopefully you do not get to keep it long, but I have enjoyed the preparation for this hearing with you and with your team. They have been thoughtful to deal with, and I appreciate your opening statement. I guess I should acknowledge the witnesses, too. Thank you to all four of

you. Mr. Harris, it is actually 85 degrees this afternoon in DC, so you did not have to avoid us in Hawaii and have to testify at 4 a.m., but thank you for participating in the pre-dawn hours there, nonetheless.

Chris, I want to applaud your opening statement. It is too easy in DC for us to take any complicated issue and reduce it immediately to heroes and villains, and whatever the regulatory or legislative predetermined tool was, to then slam it down on the newly to-be-defined problem. I think—I think you underscored a number of really important points. The simplest one is that algorithms, like almost all technologies that are new, have costs and benefits. Algorithms can make the world a better place. Algorithms can make the world a worse place.

One of the most fundamental questions before us as a people is not, first and foremost, governmental, or legislative, or regulatory, though those issues do exist. The first one is, in the new digital economy or the attention economy, the old adage holds that if a product is free, you are probably the product. The American people need to understand, we, parents and neighbors need to understand that we are being given access to these unbelievably powerful tools that can be used for lots and lots of good. In most cases, because it is free, there is somebody who would really like to capture our attention, shorten our attention spans, and drive us into often poisonous echo chambers. Algorithms have great potential for good. They can also be misused, and we, the American people, need to be reflective and thoughtful about that first and foremost.

To the tech companies who showed up today and to those of you who are also adjacent to the Silicon Valley conversation, thank you for your interest and attention to this conversation. I think it is very important for us to push back on the idea that really complicated, qualitative problems have easy, quantitative solutions. In some hearings that are not narrowly on this topic, but other technology-related Big Tech hearings that we have had over the course of the last 2 or 3 years in this Committee, sometimes really hard, nettlesome problems we have wrestled with, we have been told that as soon as the supercomputers were better, they would solve these problems. The truth is we need to distinguish between qualitative and quantitative problems.

I appreciate the Chairman's perspective on the way we are beginning this hearing, which is this is not a rush to pretend politicians know a lot more about these problems than we really do. It is an acknowledgement that there are some big problems and challenges in this area, and prudence, and humility, and transparency are the best way to begin. I am grateful for the Chairman's leadership of this Committee and this particular hearing.

Chair COONS. Thank you, Senator Sasse. I will now turn to Chairman Durbin for his opening remarks.

**OPENING STATEMENT OF HON. RICHARD J. DURBIN,
A U.S. SENATOR FROM THE STATE OF ILLINOIS**

Chair DURBIN. Thanks. I will be brief, and I appreciate the opportunity to join you and Senator Sasse and make a statement. Congratulations, Senator Coons, for taking the reins as Chair of the Privacy, Technology, and Law Subcommittee, which I was

pleased to reconstitute in this Congress. You have already demonstrated significant leadership. I look forward to your work and the cooperative efforts of Senator Sasse in bringing, I hope, some policy and legislation before the Full Committee.

This country stands at a crossroads as we grapple with the role of technology and social media in our lives and culture. I think Senator Sasse summarize it: it is plus and it is minus. We are trying to look to the minus side, but should never overlook the plus side. For example, the right to privacy, especially for children, is one of the persistent concerns I share with many Members of this Committee. Every day, internet companies collect reams of personal data on Americans, including kids, but we cannot expect children to fully understand the consequences of their internet use and this collection process.

Kids deserve, I believe, a chance to request a clean slate once they are old enough to appreciate the nature of internet data collection. That is why later this week, I will be reintroducing the Clean Slate for Kids Online Act, which would give every American an enforceable legal right to demand that website companies delete all personal information collected from or about the person when he or she was a child under the age of 13.

The right to privacy and access to one's data could keep this Subcommittee completely occupied. There is a lot more to explore, including the subject of today's hearing, which will examine how social media platforms use highly targeted algorithms to captivate and persuade us in our every aspect of our life. Algorithms influence what we read, watch, buy, and how we engage, and they do not just affect our personal lives. They infect—they affect us on a global basis.

For example, an independent civil rights audit last year found that Facebook is not sufficiently attuned to how its algorithms “fuel extreme and polarizing content,” and can drive people toward self-reinforcing echo chambers of extremism. Following the recent release of that audit, Chairman Coons wrote a letter to Facebook, which I was proud to join, that called on the company to do more to mitigate the spread of anti-Muslim extremism and bigotry on their platform. Last November, when Facebook CEO, Mark Zuckerberg, testified in this Committee, I asked him about recent incidents where hate and conspiracy groups use Facebook to plan and recruit, including the organizer of the conspiracy to kidnap Michigan's Governor, Gretchen Whitmer, and the so-called Kenosha Guard Militia, which posted a “call to arms” on Facebook in the aftermath of the shooting of Jacob Blake in Kenosha, Wisconsin.

That call to arms spread widely and was read by a 17-year-old vigilante named Kyle Rittenhouse, who traveled from Illinois to Wisconsin, where he allegedly shot and killed two people on the streets of Kenosha on August 25th, 2020. That militia page was reportedly flagged at least 455 times to Facebook. However, Facebook found the page did not violate standards, so it was left up. The response from Mr. Zuckerberg at the hearing was, and I quote, “It was a mistake. It was certainly an issue, and we are debriefing and figuring out how we can do better.”

Unfortunately, it is clear that they did not figure out how to do better quickly enough. Not even two months later, a mob of domestic terrorists and violent extremists stormed this Capitol Building in the January 6th coup attempt, fueled by widespread lies and conspiracy theories that claimed the election had been stolen from the former President. While the efforts to overturn a free and fair election were ultimately unsuccessful, the trauma of that harrowing day lingers on.

After January 6th, the consequences of rampant hate and misinformation on social media platforms has never been clearer. We need social media companies to finally take real action to address the abuse and misuse of their platforms and the role that algorithms play in amplifying it. I look forward to hearing from the witnesses, and I am hopeful that this Subcommittee can accomplish under Chairman Coon's leadership what we are expecting as an opportunity for this country to move in the right direction.

Chair COONS. Thank you, Mr. Chairman. I will now briefly introduce our witnesses for today and then swear them in.

Up first is Monika Bickert, Facebook's vice president of content policy. She originally joined Facebook in 2012 as lead security counsel, advising the company on child safety and law enforcement. Prior to joining Facebook, Ms. Bickert served as resident legal advisor at the U.S. Embassy in Bangkok where she specialized in Southeast Asia rule of law, development, and response to child exploitation and human trafficking. She also served as a prosecutor with the Department of Justice for 11 years in Washington.

Lauren Culbertson is Twitter's head of U.S. Public Policy based in Washington, DC, leads the company's Federal and State public policy teams and initiatives. Serves as Twitter's global lead for intermediary liability policy, and spearheads the company's efforts to help combat the opioid crisis. Previously, Ms. Culbertson worked in the U.S. Senate for my friend, Senator Johnny Isakson of Georgia. She also founded a business, Millennial Bridge, to promote public policy.

Alexandra Veitch leads YouTube's government affairs and public policy for the Americas where she advises the company in public policy issues around online and user-generated content. She previously served as special assistant to President Obama and as Deputy Assistant Secretary for the Department of Homeland Security. Before that, she served as a member of Speaker Pelosi's senior staff and began her career working for Senator Barbara Mikulski of Maryland. Ms. Veitch's private sector experience also includes leading North American government affairs for Tesla and CSRA.

Tristan Harris has spent his career studying today's major technology platforms and how they have increasingly become the social fabric by which we live, and think, and communicate. Mr. Harris is the cofounder and president of the Center for Humane Technology, which aims to catalyze a shift toward humane technology that operates for the common good. Mr. Harris was the primary subject of the Netflix documentary, "The Social Dilemma." Mr. Harris also led the Time Well Spent movement, which sparked product changes at Facebook, Apple, and Google.

Dr. Joan Donovan is a leading public scholar and disinformation researcher specializing in media manipulation, critical internet

studies, and online extremism. She is the research director at the Harvard Kennedy School Shorenstein Center, and director of the Technology and Social Change Project. Dr Donovan is a cofounder of Harvard Kennedy School's Misinformation Review. Her research can also be found in peer-reviewed academic journals, such as Social Media Plus Society, the Journal of Contemporary Ethnography, Information, Communication and Society, and Social Studies of Science. She is a columnist at MIT Technology Review.

You are all virtual, which makes this next step just a little different or novel for me. Would our four witnesses please stand to be sworn and raise your right hand? Since I cannot see you, I cannot affirm that you are doing that.

[Witnesses are sworn in.]

Chair COONS. Thank you. We will now proceed with witness statements. Each of you has 5 minutes to make an opening statement to this Subcommittee. Ms. Bickert, please proceed with your testimony.

**STATEMENT OF MONIKA BICKERT, VICE PRESIDENT
FOR CONTENT POLICY, FACEBOOK,
MENLO PARK, CALIFORNIA**

Ms. BICKERT. Thank you. Chairman Coons, Ranking Member Sasse, and distinguished Members of the Subcommittee, thanks for the opportunity to be here with you today. I am Monika Bickert, and I lead content policy for Facebook.

Facebook uses algorithms for many of our product features, including enforcing our policies. However, when people refer to Facebook's algorithm, often they are referring to our content ranking algorithm that helps us order content for people's News Feed, so I will just dive into that one briefly.

Chairman Coons, as you pointed out, the algorithm ranks content because we have—when people come to Facebook, they have so much potential content they could see. The average Facebook user has thousands of eligible posts every day that she could see in her News Feed, and they are all there, but what we do is we try to save them the time of sorting through all of that to find what is most meaningful to them, by instead using a ranking algorithm that ranks each post and tries to put at the top the content the person will find the most meaningful. The algorithm looks at many signals, including things like how often the user typically comments on or likes content from this particular source, how recently that content was posted, and whether the content is in a format, such as a photo or a video, that that user tends to engage with. The process results in a News Feed that is unique to each person.

Naturally our users do not see the underlying computer code that makes up the algorithms, but we do publish information about how the ranking process works, and that includes describing the inputs that go into that ranking process. Also, we have a blog post that we put out whenever we have significant changes to how we are ranking content in the algorithm. Additionally, people can actually click on any post in their News Feed and then go to—toggle the menu and go down where it says “why am I seeing this post,” and they will see the factors and explanation for why the algorithm put that piece of content where it did in their News Feed. This helps

people understand what the algorithms are doing and why they are doing it.

I do want to underscore that people can opt out of this ranking algorithm. They can toggle over to a most recent News Feed, which basically means that all of that eligible content that you could see is simply ordered in reverse chronological order. They can also choose from a—an option that we call favorite speed, which basically allows you to select pages or accounts that are favorites of yours, and then those will be the only things that will be ranked in your News Feed. We recently released a feature that allows people to toggle among those different options.

As we work to bring more transparency to the algorithm and also give people more control over how it works for them, we also are working to improve the way that the ranking system itself works. We announced last week that part of that includes expanding our surveys to understand what is meaningful to people and what is most worth their time, and also making it easier for them to give us feedback on individual posts. That is feedback that we will take from them, and we will build into the ranking algorithms in a hope that, as we make this process better and better, people will leave Facebook feeling more inspired.

News Feed ranking is not the only thing that determines what people might see when they come to Facebook. We also have a set of community standards that says this is—there are certain categories of content that simply are not allowed on our service, and those are public standards that we have had for years, and we publish a quarterly report on how we are doing at finding that content and removing it. As the report shows, we have gotten better and better, made significant strides over the past years. If content is removed for violating those standards, then it does not appear in our News Feed at all. There are other types of clickbait—there are other types of content that do not violate the standards, but nevertheless people do not want to see them, like clickbait or borderline content, that the algorithms down rank.

The reality is, it is not in our interest financially or reputationally to push people toward increasingly extreme content. If we—if we do something like that to keep somebody on the site for a few extra minutes, but it makes them have a worse experience and be less likely to use our products, then that is self-defeating. Our long-term interest is to make sure that people want to value our products for years down the road. The algorithms are a key part of how we help people connect and share and how we fight harmful content and misinformation on our site, and we will continue to do more to help people understand how the systems work and how they can control their experience.

Thanks. I look forward to your questions.

[The prepared statement of Ms. Bickert appears as a submission for the record.]

Chair COONS. Thank you very much, Ms. Bickert. Ms. Veitch, would you please proceed with your testimony?

**STATEMENT OF ALEXANDRA VEITCH, DIRECTOR OF
GOVERNMENT AFFAIRS AND PUBLIC POLICY FOR THE
AMERICAS AND EMERGING MARKETS, YOUTUBE,
SAN BRUNO, CALIFORNIA**

Ms. VEITCH. Chairman Coons, Ranking Member Sasse, and distinguished Senators of the Subcommittee, thank you for inviting me to appear before you today. My name is Alexandra Veitch, and I am the director of government affairs and public policy for the Americas and emerging markets at YouTube. I appreciate the opportunity to explain how algorithms and machine learning support YouTube's mission to give everyone a voice and show them the world.

Through the adversity and uncertainty of the last year, YouTube has helped bring people together as we have stayed apart. More viewers than ever have come to YouTube to learn new skills, to understand the world more deeply, and to be delighted by stories that cannot be found elsewhere. YouTube's business relies on the trust of our users, our creators, and our advertisers. That is why responsibility is our number one priority.

Our approach is based on what we call the four R's. We remove content that violates our community guidelines, we raise authoritative voices, we reduce the spread of borderline content, and we reward trusted creators. Our written submission explains each pillar in detail, but I want to focus my comments today on how machine learning supports this responsibility work when it comes to recommendations.

Recommendations on YouTube help users discover content that they will enjoy and on key subjects. We want to recommend content to our users that is authoritative. Recommendations are based on a number of signals, including, if enabled, a user's watch and search history. We also consider factors like country and time of day, which help our system show relevant news consistent with our efforts to raise authoritative voices. We also give our users significant control over how their recommendations are personalized. Users can view, pause, edit, or clear their watch or search history at any time. We also give users the opportunity to provide direct feedback about recommendations so they can tell us if they are not useful.

We also believe we have a responsibility to limit recommendations of content that is not useful or may even be harmful. That is why in January 2019, we launched more than 30 changes to our recommendation systems to limit the spread of harmful misinformation and borderline content, which is content that comes close to, but does not cross the line of violating our community guidelines. As a result we saw a 70-percent drop in watch time of such content from non-subscribed recommendations in the U.S. that year. This borderline content is a fraction of one percent of what is watched on YouTube in the U.S., but we know that it is too much, and we are committed to reducing this number.

We know there is interest in the quality of the content we recommend to our users. Researchers around the world have found that YouTube's recommendation systems move users in the direction of popular and authoritative content. Our efforts to raise up content from authoritative sources and reduce recommendations—

excuse me—of borderline content and harmful misinformation outweigh other recommendation signals, even if the net result is decreased engagement. We are proud of our record here, but we also work continuously to improve.

Because responsibility and transparency go hand-in-hand, I would like to close with three recent transparency efforts we have undertaken to facilitate a better understanding of our platform. First, in May 2020, we collaborated with Google to launch the first threat analysis group bulletin. It regularly discloses actions that we have taken to combat coordinated influence operations from around the world. Second, in June 2020, we launched a website called How YouTube Works to answer frequently asked questions. It explains our products and policies in detail and provides information on critical topics, such as child safety, harmful content, misinformation, and copyright. Third, earlier this month, we added a new progress metric to our quarterly Community Guidelines Enforcement Report. Our violative view rate estimates the percentage of views on content that violates our policies. Last quarter, this number was .16 to .18 percent, meaning that out of every 10,000 views on YouTube, only 16 to 18 come from violative content. This is down by over 70 percent compared to the same quarter of 2017, thanks in large part to our investments in machine learning.

As we work to balance the open nature of our platform with our important work to be responsible, we appreciate the feedback we receive from policymakers. We will continue to do more. Thank you again for the opportunity to appear before you today. I look forward to your questions.

[The prepared statement of Ms. Veitch appears as a submission for the record.]

Chair COONS. Thank you, Ms. Veitch. Ms. Culbertson from Twitter, if you would now present your opening statement, your testimony, that would be wonderful. Thank you.

**STATEMENT OF LAUREN CULBERTSON, HEAD OF U.S.
PUBLIC POLICY, TWITTER, SAN FRANCISCO, CALIFORNIA**

Ms. CULBERTSON. Thank you, Chairman Coons, Ranking Member Sasse, and Members of the Subcommittee for the opportunity to testify on behalf of Twitter today on the role of algorithms and amplification of content.

Twitter's purpose is to serve the public conversation. In the early days, we were where you could go to share 140 character status updates. Our service has become the go-to place to see what is happening in the world and to have conversations about a wide range of topics, including current events, sports, entertainment, and politics. While much has changed since the company was founded 15 years ago, we believe our mission is more important than ever. While many of the challenges we grapple with today are not new, the creation and evolution of the online world have affected the scale and scope of these issues. Moreover, we must confront these issues amidst increasing global threats to free expression.

We believe that addressing the global challenges that internet services, like ours face requires a free and open internet. We are guided by the following principles as we seek to build trust with the people we serve. This includes increasing transparency, pro-

viding more consumer control and choice, and improving procedural fairness. Let me expand on the principle of consumer control and choice, as it is particularly relevant to today's discussion on algorithmic choice.

In 2018, we introduced a feature to give people on Twitter control over the algorithms that determine your home timeline. Through the sparkle icon you see on the top right corner of your screen, you can choose to see your tweets ranked or toggle to view tweets in reverse chronological order. When we implemented this, some suggested it would be bad for our business. We thought it was the right thing to do for our users, and it has been a core feature ever since.

Further in line with our commitment to choice and control, Twitter is funding BlueSky, an independent team of open-source architects, engineers, and designers, to develop an open and—open and decentralized standards for social media. It is our hope that BlueSky will eventually allow Twitter and other companies to contribute to and access open recommendation algorithms that promote healthy conversation and ultimately provide individuals greater choice. These standards could support innovation, making it easier for startups to address issues like abuse and harmful content at a lower cost. We recognize that this effort is complex, unprecedented, and will take time, but we are currently planning to provide the necessary exploratory resources to push this project forward. As we make investments to provide more transparency and choice, we have also launched our Responsible Machine Learning Initiative to conduct an in-depth analysis and studies to assess the existence of potential harms in the algorithms we use. We plan to implement our findings and share them through an open process to solicit feedback.

Finally, as policymakers and Members of Congress here debate internet regulation, I urge you to consider the ways algorithmic choice and machine learning make Twitter and other services a safer place for the public conversation. Technology is essential for rooting out harmful content, like terrorism and child sexual exploitation content. We also rely heavily on machine learning tools to surface potentially abusive or harmful content for human moderators to review. Simply put, we must ensure that regulations enable companies to tap technology to help solve some of the problems that technology itself poses.

In summary, we believe that moving toward more open systems will increase transparency, provide more consumer control and choice, and increase competition in our industry. This will ultimately lead to more innovation to solve today's and tomorrow's challenges. We appreciate the enormous privilege we have to host some of the most important conversations in the world. We are committed to working with a broad group of stakeholders to get this right for the future of the internet and for the future of our society.

Again, thank you for the opportunity to be here with you today.

[The prepared statement of Ms. Culbertson appears as a submission for the record.]

Chair COONS. Thank you, Ms. Culbertson. Mr. Tristan Harris of the Center for Humane Technology, if you would now please give your opening statement.

**STATEMENT OF TRISTAN HARRIS,
COFOUNDER AND PRESIDENT, CENTER FOR
HUMANE TECHNOLOGY, SAN FRANCISCO, CALIFORNIA**

Mr. HARRIS. Thank you, Senator Coons, Senate Sasse, and Chairman Durbin. It is an honor to be here with you today.

My background is I used to be a design ethicist at Google. That was before recently featuring in the film, “The Social Dilemma,” which many of you might have seen, which really had the insiders who understood how these technologies were built in the first place and have affected society. My friends in college were some of the people who ended up working at these companies in the very early days, including my friends Mike and Kevin, who actually started Instagram.

What we really are missing in this conversation is a focus on the business model and the intrinsic nature of what these platforms are about, not because they are evil, and none of the people who are here with us today are, you know, intentionally causing any harm. Neither do I believe that the tech companies who created these systems have intentionally wanted any of these harms to happen. We are now in a situation where if we do not diagnose the problem correctly, we are going to be in a bit of trouble.

While you are hearing, you know, from the folks here today about the dramatic reductions in harmful content, borderline content, hiring tens of thousands more content moderators, et cetera, it can sound very convincing. At the end of the day, a business model that preys on human attention means that we are worth more as human beings and as citizens of this country when we are addicted, outraged, polarized, narcissistic, and disinformed, because that means that the business model was successful at steering our attention, using automation. We are now sitting through the results of 10 years of this psychological deranging process that has warped our national communications and fragmented the Overton Window and the shared reality that we need as a Nation to coordinate to deal with our real problems, which are existential threats, like climate change, the rise of China, pandemic, education, and infrastructure.

Long as these companies profit by turning the American conversation into a cacophony, into a kind of Hobbesian war of all against all, because that is the business model, again, of not the advertising, but the model of everyone getting a chance to speak and have it go viral to millions of people. As long as that is the promise with personalization, we are each going to be steered into a different rabbit hole of reality, which Joan will do such a good job of talking about. If you care about or believe that masks work, you will see infinite evidence that masks work. If you click on a couple articles that say masks do not work and here is the, you know, stats in Florida showing that the data was different, you will see infinite evidence that masks do not work. Then we are pitted against each other with this sort of infinite virality where anything that somebody said can go viral.

Fundamentally, this is breaking many different aspects of the Nation's fundamental life organs. For children, increased cyber bullying leads to an increase in suicide. It takes momentary drama and it turns it into drama snowballs that drown out the effects of teachers and classrooms, who have to spend 2 hours on Monday morning clearing up all the drama that occurred on social media over the weekend. It can reverse huge progress that we have made in civil rights and not perpetuating racial stereotypes by increasing online harassment and rewarding the presentation of minorities in ways that are demeaning. It can increase—it can inhibit our progress on climate change because climate disinformation has gone viral on these platforms.

It can—it can—it can pose a threat to national security in the sense that if Russia or China try to fly a plane in United States, they would be shot down by our Department of Defense, but if they try to fly an information bomb in the United States, they are met by a white-gloved algorithm from one of these companies that says exactly which zip code would you like to target. It is the opposite of national security. What a cannon was to a castle, social media is to the nation state because it removes the power asymmetries of the millions and billions of dollars that we have spent on F-35s, on passport controls, and the Department of Homeland Security.

Once your society becomes virtual, all those protections go away. Most importantly, if we are not coordinated as a society, if we cannot even recognize each other as Americans, we are toast. That is the only thing that matters. If we do not have a truth that we can agree on, then we cannot actually change any—do anything on our existential threats. We are really sitting at a moment in history where we are transitioning into becoming a digital society, and we kind of already have a neural link brain implant for our society.

Right now we have two options. We have the Chinese brain implant, which leads to kind of an Orwellian control of thought, mass behavior modification, and surveillance, or we have the Western brain implant that is built on this business model that turns us into a performative culture. You know, there is the Orwellian dystopia or the Aldous Huxley and “Brave New World” dystopia in which we fall into a kind of development of amusing ourselves to death, constantly immersed in distractions and unable to focus on our real problems.

What I really encourage us to think about is someone is going to be controlling the 21st century. Will it be open societies or closed societies? Either we beat China at becoming China in a digital way, which we do not want to do, or we figure out how to be a digital open society that does not actually lose to that. That is our task. Either we figure it out, or the American experiment may be in question.

[The prepared statement of Mr. Harris appears as a submission for the record.]

Chair COONS. Thank you very much, Mr. Harris. Dr. Donovan, if you would now give your opening statement, please.

**STATEMENT OF JOAN DONOVAN, PH.D.,
RESEARCH DIRECTOR, SHORENSTEIN CENTER
ON MEDIA, POLITICS, AND PUBLIC POLICY, AND
LECTURER IN PUBLIC POLICY AT JOHN F. KENNEDY
SCHOOL OF GOVERNMENT AT HARVARD UNIVERSITY,
CAMBRIDGE, MASSACHUSETTS**

Dr. DONOVAN. Great. Thank you to the esteemed Members of the Subcommittee, Chairman Coons and Ranking Member, Senator Sasse, for inviting me, and thank you to your staff as well. I appreciate the opportunity to talk about how algorithms and amplification shape public discourse. I am Joan Donovan, the research director of the Shorenstein Center at Harvard Kennedy School, and I study the internet.

I want to remind everyone that the internet is a truly global technology requiring massive amounts of international labor. Whatever policy ends up coming from the U.S. will undoubtedly become the default settings for the rest of the world. I also want to begin by saying that I believe a public interest internet is possible, and I have to believe that in order to do the heinous job of researching hate, incitement, harassment, and disinformation on these social media products.

What a public interest internet means practically is crafting policy that draws together the best insights across many different professional sectors, matched with rigorous, independent research into how automation and amplification shape the quality of public life. We should begin by creating public interest obligations for social media timelines and news feeds, requiring companies to curate timely, local, relevant, and accurate information, as well as providing robust content moderation services and options.

Today, let us try to name the problem of misinformation at scale and its impacts. In the U.S., when we talk about politics, we are really talking about media about politics, and when those news and information flows get laced with strategic misinformation, then a simple search for something like “coronavirus origin” or “mail-in ballots,” can lead people down the rabbit hole of medical misinformation or political disinformation. In October 2020, I testified about misinformation at scale having similar harmful societal impacts as secondhand smoke, and it took a whole-of-society approach to address the burden of disease caused by secondhand smoke, and which led us to clear the air in workplaces, schools, and airports.

When I say “misinformation at scale,” I am not complaining that someone is wrong on the internet. What I am pointing to is the way that social media products amplify novel and outrageous statements to millions of people faster than timely, local, relevant, and accurate information can reach them. Post-2020, our society must assess the true cost of misinformation at scale and its deadly consequences. Disinformers, scammers, and grifters use social media to sell bogus products, amplify wedge issues, impersonate social movements, and push conspiracies. What I have learned over the last decade of studying the internet is that everything open will be exploited. Moreover, misinformation at scale is a feature of social media, not a bug.

What do I mean when I say that? For example, because of what I study, I often joke nervously that my computer thinks I am a

white supremacist. For researchers, going down the rabbit hole means getting pulled into an online subculture where key words, slang, values, and norms are unfamiliar, but, nevertheless, the content is plentiful. There are four aspects of the design of social media algorithms that can lead someone into the rabbit hole. Coincidentally, they are also four R's.

Repetition relates to seeing the same thing over and over on a single product, which, you know, likes, shares, retweets do that. Redundancy is seeing the same thing across different products; that is, you see the same thing on YouTube that you see on Twitter. It tends to produce a feeling that something is more true. Responsiveness is how social media and search engines always provide some answer, even if it is wrong, unlike other forms of media. Then last, reinforcement refers to the ways that algorithms work to connect people and content so that once you have searched for a slogan or a keyword, algorithms will reinforce these interests time and time again.

Nowhere, of course, is this more prevalent than on YouTube where any search for conspiracy or white supremacist content, using the preferred keywords of the ingroup, will surface numerous recommendations, and even offer up—offer up direct engagement with these communities and influencers. If you have recently searched for contentious content, like “Rittenhouse,” “QAnon,” “Proud Boys,” or “Antifa,” you are likely to enter a rabbit hole or extracting yourself from reinforcement algorithms ranging from the difficult to the impossible. The rabbit hole is best understood as an algorithmic economy where algorithms pattern the distribution of content in order to maximize growth, engagement, and revenue.

I have a few things that companies could implement if we want to talk about that later, but I think tackling a problem this big will require Federal oversight for the long-term. We did not build airports overnight, but tech companies are flying the plane with nowhere to land at this point. Of course, the cost of doing nothing is nothing short of democracy's end. Thank you.

[The prepared statement of Dr. Donovan appears as a submission for the record.]

Chair COONS. Thank you very much for your thoughtful testimony, to all of our witnesses. Given the limited number of Members, we may get several rounds of questioning, which is exciting to me.

I just want to say to Ms. Bickert, Ms. Culbertson, Ms. Veitch, your efforts to down rank borderline content to improve transparency and empower users are all positive steps, and we need to continue to find ways to preserve the positive benefits of algorithms in showing content to people that is meaningful to them, while addressing the very clear threats and challenges, the very real potential for the harmful impacts of algorithmic amplification. The questions I have today are meant to get a better understanding of how one might further build on your efforts and strike the right balance.

Some have proposed that social media platforms create virality circuit breakers—we are all familiar with the phrase “blowing up on the internet”—to detect content that is rapidly gaining widespread viewership so that humans can review whether it actually

complies with platform policies before it racks up tens or hundreds of millions of views. Professor Donovan, could you just briefly, concisely explain why this kind of mechanism might be particularly valuable?

Dr. DONOVAN. Yes, I think one of the things that we know now from decades of tracking flagging, especially in communities—conspiracist communities, hate communities, they only tend to flag things as a result of trying to get retribution on one another. They do not—they search for this content and they enjoy it, and so systems that are built in do not tend to work when it comes to particular kinds of strategic misinformation, especially hate or harassing content as well.

As a result, what you need to do as a corporation is really look for it. I know that there have been a couple of different instances recently where corporations have found and rooted out some really heinous stuff, but it obviously has to be part of the business process and the process of content moderation to seek out content that is, essentially, out of skew with signals from the past.

Chair COONS. Thank you.

Dr. DONOVAN. That is one of the ways that they could incorporate this.

Chair COONS. Thank you, Professor. Ms. Bickert, Facebook said last fall it was piloting this very concept. What did you find through this experience, and do you expect to further roll this out more broadly? Please explain briefly, if you might.

Ms. BICKERT. Senator, thank you for the question. We do look at virality of content as a signal in when we should assess proactively, as Dr. Donovan is suggesting, whether or not something in it violates our policies or should be referred to our fact checkers. The fact checkers, as you may know, are more than 80 independent fact-checking organizations that we work with. They can proactively rate content or we can send it to them. Either way, or—and that could be based on user reports, too. Either way, if they rate something “false,” then that is when we will put on that label saying this content is false, directing people to the fact check, and we will also reduce the distribution of that content in our News Feed.

Yes, we are seeing that those efforts are paying off. In fact, we see that when we put one of those informational labels on top of a piece of content, people are far less likely to actually click through and see the content than they would if we did not have that label.

Chair COONS. Ms. Bickert, I appreciated several steps Facebook announced it was taking just in advance of the *Derek Chauvin* verdict. One of these steps was limiting the spread of content. This is a quote from Facebook, “that systems predict is likely to violate our community standards in the areas of hate speech, graphic violence, and violence and incitement.” Facebook’s statement also noted the company had done this in other emergency situations in the past. My question for you is why Facebook would not always limit the rapid spread of content likely to violate these standards? Could you help us understand that?

Ms. BICKERT. Senator, yes. What we are doing, and I put that blog post out. What I meant by that was we use systems to

proactively identify when content is likely to violate or is maybe borderline. Often what that can help us do is send that to our reviewers and have them assess whether or not it violates. In extreme situations, because, of course, not all of that content will violate, you know. There will be some false positives in that.

There is a cost to, you know, for instance, taking action on that content with about half of it being real people who look at it. What we do is, generally, we use those measures to find content that we can send to reviewers. In situations where we know that there is extreme and finite, in terms of date, risk, such as an election in a country that is going through civil unrest or the situation in Minneapolis with the *Chauvin* trial, we will put in place a temporary measure where we will deemphasize content that the technology—that the algorithms say it is likely to violate.

Chair COONS. Let me ask a last question of the three social media representatives before I turn this to my Ranking Member. Ms. Bickert, Facebook has said, and I think you said in your opening statement, it is not in your long-term interest to pursue maximum engagement if it comes at the cost of spreading polarizing and sensationalized content, that it is not really, long term, in the financial interest of the company, let alone its reputational interest, to have algorithms that amplify harmful or divisive content. I agree with this. I am concerned about what the underlying incentives are at all three of your platforms for those who have to make decisions day in and day out about exactly how your companies operate.

The MIT Technology Review reported last month that pay incentives at Facebook for employees broadly are still tied to growth metrics and engagement metrics. If I am a Facebook employee who works on its News Feed, are the metrics the company has set up to measure my performance directly related simply to engagement and growth metrics, or is there some way that these broader, more positive social objectives are incorporated? If you could, all three, just answer briefly. Ms. Bickert, for Facebook, Ms. Veitch and Ms. Culbertson, do you provide pay incentives in terms of algorithms teams, directly or indirectly, based on engagement- and growth-related metrics? Thank you.

Ms. BICKERT. Senator, the engineers are not specifically goaled or given pay incentives simply to increase time on the site. The focus is really on making sure that the products are services that people find useful and will want to use for years to come.

Ms. CULBERTSON. Senator, for Twitter, a top priority for our company across our teams is to serve a healthy public conversation, and I would love to share with you our transcript from our latest analyst day, which is what we share with our investors and our advertisers. All of the concerns and priorities that we have talked about thus far today, you will see that what we are telling you is exactly what we tell our investors and our advertisers because they have the same concerns. We have no incentive to have a toxic or unhealthy conversation on the service.

Chair COONS. Thank you, ma'am.

Ms. VEITCH. Similarly, Senator. Responsibility is our number one priority, and when we set goals, we set those goals around what we define as responsible growth. We may set a goal to encourage

adoption of a feature, but also we want to take into account how that feature may be used or misused, and how we can ensure it is adopted responsibly.

Chair COONS. Mr. Harris, if you could just provide a brief comment on your understanding of the incentives of employees and how it aligns with responsible growth versus growth at all cost.

Mr. HARRIS. Yes. My understanding is, even to this day, I think there was a brief experimentation at Facebook with non-engagement-based performance incentives for social impact, but that those have largely gone away, and it is actually still a measure of engagement. This is things like not time onsite, but sessions, 7-day active users, growth, and that is still the focus. Everything else we are going to be talking about today, it is almost like having the heads of Exxon, BP, and Shell asking about what are you doing to responsibly stop climate change.

Again, their business model is to create a society that is addicted, outraged, polarized, performative, and disinformed. That is just the fundamentals of how it works. While they can try to skim the major harm off the top and do what they can—and we want to celebrate that, we really do—it is just fundamentally if they are trapped in something that they cannot change.

Chair COONS. Thank you all. Let me turn to my Ranking Member, Senator Sasse.

Senator SASSE. Thanks, Chris. My first question is actually building—based on exactly—pardon me—on where the Chairman just finished, and I really do think that constructive engagement in these Committees is better than people trolling for sound bites. I am not trying to get you all to fight, but the truth of the matter is this hearing would work a lot better if we were in the same room so we did not have to try to bring you all into dialogue. The last three answers from the social media companies and Mr. Harris' answers are just ultimately not reconcilable, I do not think.

I want to go back to—I will start with Ms. Bickert as well. Saying that you aspire to healthy engagement as opposed to just more quantity, I agree with Mr. Harris' line that you definitely aspire to skim the most destructive habits and practices off the top of digital addiction, but the business model is addiction, right? I mean, money is directly correlated to the amount of time that people spend on the site. I guess what would be useful for me is to hear each of the three of you say what you think is wrong with Mr. Harris' argument, because right now I think we are talking past each other.

I know that there is bad content and there is disinformation content that you all, well intentioned as your companies surely are, want to curtail. His argument is really more broadly about the business model, and the business model is addiction, is it not? Ms. Bickert, can we start with you? What is—what is Mr. Harris missing?

Ms. BICKERT. Senator, thanks for the question. You know, I will say—I will say two things that I hope will be helpful. One is, for us, the focus is always on the long term, and I will give one concrete example of that. In January 2018, we put out a post announcing that we were going to be prioritizing content from family and friends over, say, news content. It was called “meaningful social

interactions.” We suspected that it would lead to less time spent on the service and it did. It led to people spending tens of millions of fewer hours on Facebook every day. That was something that we did because we thought that, longer term, it was more important for people to see that sort of content because they would find it meaningful, and they would want to continue to use this site. It is not—it is a long-term picture.

The other thing I would say is, the teams that I work with, who include the engineers who are focused on safety issues, removing content, say, bullying content or hate speech, and the engineers who are focused on the way that we reduce, for instance, misinformation that has been labeled on the site, a key statistic for those engineers is prevalence. Violating content, that is their goal, and we put out public reports on their prevalence. You know, that is an example of how we are focused on the long term and making sure that we are stopping abuse and maintaining a healthy environment.

Senator SASSE. I want to be clear that I am not—I am not targeting the three of you because I—my opening statement is very sincere. I think that there is a danger in politics, in governance, where if you agree that there is a problem, then there must be a definitive regulatory solution that can come real fast and easy. On the other hand, if you are not persuaded there is a regulatory fix right away, then you have to deny there is a problem. I am sort of a heterodox tweener on this in that I do not have clarity about what the regulatory fixes would be, but I think society-wide, we should admit that there is a problem in the last 12 or 14 years as we have consumed more and more digital stuff that seems to be correlated with some benefits, but also some very real costs.

I do not think it is just your companies. I mean, there have been reports out of the New York Times about their own internal deliberations about how they would like to have more Americans engaging in healthy content, and they are just printing money right now over the course of the last 4 or 5 years, but engagement is much higher when they are angry. When the content is angry, it leads to more engagement.

I do not think any of you are really going to dispute that, but I would like to stay where I—where the question was 2 minutes ago which is, I would love it if, Ms. Culbertson, will you tell me what you think is wrong with Mr. Harris’s argument?

Ms. CULBERTSON. We are really focused on serving the public conversation, and that includes having controls in place so people can also control their experience. I think as we are talking about algorithms today, you know, Twitter really does one thing. We do tweets. We have a home timeline. As we are talking about algorithms, we have a ranking algorithm. That is designed to show you what might be most relevant to you. Then also if we are talking about screen time or how much time do you spend on a service, I think that is really relevant because I know as a user of Twitter myself, I rely on that so I can kind of see what happened in the day, what people are talking about, and then I kind of log off, move on with my day.

I think it is important to look at this in a nuanced view and recognize that algorithms can also be helpful in terms of cutting down on screen time or providing more valuable experience for people.

Senator SASSE. Sure, but the reality is the loop between the products that are being produced and the way, we as a narcissistic Senators, consume it is—maybe I will ask if this is the right question. Is it or is it not true that when somebody tweets something that is really anger-invoking and outrageous and it goes viral, but then 2 hours later, they realized they were wrong and they correct it, is the correction not usually, like, 3 percent of the traffic of the original outrageous, but false, thing? I mean, so it seems to me that what we know is that people are pretty good at short-term rage, and the product capitalizes on that, does it not?

Ms. CULBERTSON. I think when looking at Twitter, it is important to remember that it is an open, public conversation, and so everything that happens is in the open in the public, and, typically, you know, you will see these debates play out. You know, I am a firm believer that connection and connectivity is key to solving problems, and that is what our service does. Of course people have robust and spirited debates, but I think you have to look at the greater picture there.

Senator SASSE. We are basically at time, so I will not indulge the—get the Chairman to indulge me much longer. Ms. Veitch, do you have anything to say that you think Mr. Harris is wrong about? It would be useful, but right now we are not getting much direct engagement with that. He is making a big argument, and I think we are hearing responses that are only around the margins. Ms. Veitch, do you have any criticism of Mr. Harris' argument?

Ms. VEITCH. Yes. Thank you, Senator. I think in your opening statement, you call these nettlesome problems. I agree they are, but I would just make two quick points. First, misinformation is not in our interest. Our business relies on the trust of our users, but also our advertisers, who, on our platform, advertise on single pieces of content. We want to build these relationships for the long term. That is why we bake user choice, user control right into the product with things like timers and the opportunity to turn auto play off, take a break reminders of which we have sent over one billion. Again, those exist so we can build this relationship with our users for the long term.

Senator SASSE. Thank you. I do think a lot of those user preference tools to manage our level of engagement. Engagement bridging over to addiction is really an important innovation, so I applaud those tools.

Chair COONS. Thank you, Senator Sasse. Chairman Durbin?

Chair DURBIN. Thanks a lot, Chairman Coons, and it is a pleasure to be with you. First two disclaimers. I am a liberal arts lawyer, not nearly as tech savvy as I should be for this hearing. Point number two, my experience in Government, which has been over several decades, suggests that we are slow to recognize issues that are fast-breaking and have a very spotty record when it comes to responding to them in a thoughtful way. I hope this is an exception. If I could address Tristan Harris first. Aloha. Then may I ask you this question? I have been reading and trying to understand why the European Union is taking such an apparently bold and in-

novative approach to this subject and we are so slow to respond. Any thoughts?

Mr. HARRIS. I do not—in this country, obviously we value free speech above other values, and so that makes it, I think, harder to regulate an environment where the composition of what constitutes speech in our society is a Frankenstein monster that spins out blocks of attention virally to different people on a personalized basis and outrages them. I do not think that we have had a framework.

To your earlier point, one of the quotes we reference often is from E.O. Wilson: “The fundamental problem of humanity is we have Paleolithic emotions, medieval institutions, and then accelerating godlike technology.” It is not meant as an insult to any government body. It is just to say the challenge you pose, which is how do we deal with the first derivative of these issues, we are still talking about mostly conversations that we had, you know, 4 years ago about spread of misinformation, things like that. The rate and acceleration of new kinds of threats, new kinds of issues, the growth rate of that is going far—growing far faster than the growth rate of our capacity to mitigate or respond to those threats.

I was speaking with someone in the fact-checking network who said, you know, there are now 200 billion messages a day going through WhatsApp, 15 billion going through Facebook. They get about 100 fact checks per day. If you think about a bank being overleveraged and how much risk are they—how far over their skis are they, we have got about 200 billion-to-100 in terms of scale event of information that is running through a system without moderation. With—as Senator Sasse said, the decentralized incentive for yellow journalism is, if it bleeds, it leads. It wants to make each of us into yellow journalists because we each get more rewarded the more extreme things that we say in an inadvertent way.

In a subtle way, in this is very slow climate change and culture kind of way, just by 2 percent increasing the kind of outrageous sense of what you are saying, the sensationalism of what you are saying, you got a few more likes, you get a few more hits, and so you keep doing it. Suddenly, that heats up the global conversation and has us not even recognizing our fellow Americans as fellow Americans.

Chair DURBIN. I want to take you down a different path, if I can, for a moment as I try to read and absorb the European Union’s risk-based approach to this AI issue. They say two things they find unacceptable: use examples, manipulating human behavior. I think that is at the heart of it, as I understand the explanation. I have heard people from Facebook talk about making your Facebook experience more meaningful, and folks from YouTube and Twitter talking about healthy dialogue. The bottom line is, it appears that there—like it or not, there is a factor here where our human behavior is being affected by what we are seeing, what we are reading, what we are experiencing, and that seems to violate the basic premise of the EU regulation.

The second one in the extreme is this social credit scoring, which they use as an example, which apparently is rampant in places like China, and takes the manipulation, and analysis, and algorithms

to the point where they disqualify people from being able to get on a fast train in China because their social credit score does not merit it. Give me your thoughts on those two elements.

Mr. HARRIS. Yes, on the manipulation front, you know, that would disqualify just about all of the three companies that are sitting in front of you, including TikTok, by the way, which is not getting nearly as much attention and it is actually dominating children's minds on a daily basis. You know, I think if you have seen the film, "The Social Dilemma," we speak about my background in a class called the Stanford persuasive technology class and lab, and that these technologies are designed to be persuasive. When you hear Twitter talk about the healthy conversation, there is still a persuasive technology that manipulates human behavior. They are trying to do it in as healthy a way as they can. When you hear Facebook talk about meaningful social interactions, they are still creating this sort of digital addiction dopamine loop, getting you to invite your friends, create social obligations, dripping out notifications one at a time instead of in batches. They are—but they are doing it in "the most meaningful way" that they can.

They are trying to do the best they can, but to Senator Sasse's point earlier, it is almost like listening to a hostage in a hostage video. Nothing they are saying kind of makes much sense until you realize there is a gun off stage holding—you know, their business model—held to their heads, and it is causing them to say the things that they are saying. Again, these are really good people. They are just—we cannot talk about the actual underlying issue because the business model is based on this manipulation. Sorry. I think I ran out of time for your second part.

Chair DURBIN. In my last 5 seconds, I am plugging for the Clean Slate for Kids Online Act. It is a small, but very important, part of this conversation. Thank you, Mr. Chairman.

Chair COONS. Thank you, Chairman Durbin. Thank you for joining us today. Senator Hawley.

Senator HAWLEY. Thank you, Mr. Chairman, and thanks to the witnesses for being here. Mr. Harris, I would just like to come back to you, if I could, because I think you just said something in response to Senator Durbin that is very, very important. You talked about the business model of the companies, although the companies that are up before us today are, in fact, all of the dominant platforms. The point—your point I think is, and I would like you to elaborate on it, the business model of these companies is addiction. The business model of these companies is advertising, which takes place by getting more and more users to spend more and more time online, so that these companies can gather more and more personal information about them to sell them more and more stuff, right? It is an attention treadmill. It is an addiction economy. That is the design. That is by design. They did not wander into it. It was not an accident. It was not, oops, how did we end up here. They designed it this way. Addiction is the design.

You have written a lot about this. You produced, as you have referenced, a documentary about this. Can you just tell us more about this core business model of these dominant platforms?

Mr. HARRIS. Yes, thank you. I would actually say that there are so many of the ways that these platforms' work actually comes

from that original business model. I literally remember in the early days of Instagram when they were trying to figure out a way to get people to kind of come back, and they decided to borrow Twitter's follow model. The idea that, you know, you can follow any user. Why did they invent this follow model? Let us just ask it that way. Why are we following each other here and there? Because what it does is it creates this treadmill where every day or two you see you have got two more people who followed you, and that creates what is kind of like a viral bait, right? It is like a clickbait that gets you to come back into the service to see, oh, who is that person? What do they want to follow me for? They are preying on each other's social validation, and then they have this AI that sits on top that tries to predict, well, which users could we get to follow whom.

If you want to say the problem has gone away, Facebook right now—right now—is testing the increase of suggested users you may know—by the way, this is in the film, “The Social Dilemma”—when one of the AIs sort of twirls the mustache and says, could we invite them—could we get them to invite more friends. That is actually literally what they are doing right now, which is they are saying, here are some channels, here are some people you might want to follow, and they are very good at predicting that next person.

I have even done this myself because I have got a supercomputer pointed at my brain saying, yes, this is your old friend that you might actually know. It creates, again, this treadmill that is all about getting us to come back. It is really almost like a digital drug lord because if you are a child and you saw “The Social Dilemma,” you say, hey, I am going to uncheck—you know, pull out of these services for a while. Have you ever tried taking down your Facebook account or not using Instagram for a week? You will notice they dial up how aggressively, like a digital drug lord, to show you more notifications and more emails. They will send you more and more emails, each of the services will. They would not do that if their business model was not preying on addiction.

These are all techniques, again, from the kind of persuasive framework that all of the people that I came up with in the tech industry really learned, again, not because these people are evil. Just little bit by bit, you do what works, and then it keeps turning into this treadmill, and it sort of turns us all into attention vampires that want attention from other people.

Senator HAWLEY. Yes, “attention vampires” is a great term. The amount of control that this business model then gives these companies over our lives is absolutely unbelievable. There is that infamous experiment that Facebook ran on its users in 2014, half a million users, to see if it could depress them or change their moods by tweaking the algorithm that then would refer the content that they saw. Of course, the amazing and extremely scary thing is that they could. They could, in fact, directly influence their users' moods. They could, in fact, change how their users felt about the day, or felt about a particular story, or felt about a particular event by tweaking their algorithm, because they control—“they,” Facebook—control what their users see. They control the interactions. They increasingly control how much time people spend online. You know, and really these companies say that they are about

social media, but really they once were, right? I mean, they used to be social media networks.

Back in 2006 when Facebook first introduced the News Feed, there is this great post that Mark Zuckerberg wrote called “Calm Down. Breathe. We Hear You,” in which he assured users that the introduction of this new called the News Feed would be not a very big deal. He said, “We have been getting a lot of feedback about mini-feed and News Feed. We think they are great products, but we know that many of you are not immediate fans.” There is an understatement. “For those who are worried about privacy,” he went on, “nothing you do is being broadcast.” Right. “Rather, it is being shared with people who are about what you do, your friends.” Of course, what turns out to be true is it is not your friends who he is concerned about. It is advertisers, and who they are sharing this personal information with is advertisers who are in the business of trying to manipulate the user.

Let me just ask you this, Mr. Harris. These companies have been able to do this. They have been able to manipulate content. They have been able to push particular content to users. They have been able to try and interfere with their user’s own moods, in large part, because they get a special blanket immunity from this Government, from the Federal Government. Here is my question. Why should any platform that engages in algorithmic amplification or behavioral advertising, why should they get the Section 230 immunity? Why should we not just remove Section 230 immunity for any platform that engages in behavioral advertising or algorithmic amplification?

Mr. HARRIS. Section 230 is a—is a—is a difficult—it is a double trade, and so it is going to be a complex one to get into there. It is important to say that whether the companies want to or not, if they took their hand off the steering wheel, they are still manipulating people’s emotions. In fact, the more they take their hand off the steering wheel, the more outrageous, values-blind engagement, which means that, literally, the most outrageous stuff, the most child trafficking stuff, the most sexual pornographic stuff would rise to the top, and that would also be a form of manipulation. If you compare side-by-side how much restrictions we do in an IRB study in a psychology lab at a university where, if you are going to experiment on 14 people, you got to file an IRB review, Facebook, Twitter YouTube, TikTok, are, on a regular, daily basis, tinkering with the global brain implant of 3 billion people’s daily thoughts with no oversight.

I think what we need to do is actually compare side-by-side what are the regulations and protections, as you are talking about, that we apply in one domain and we are not applying in a different domain. I think the focus on content in 230 is problematic. I think it is more about the design and fundamental oversight about the way these platforms fundamentally operate.

Senator HAWLEY. I will just say in closing, Mr. Chairman, that I think Senator Durbin put a very good question, which is why are we so slow. Why has the United States been so slow to confront this kind of manipulation and these kinds of deleterious effects? I think part of the answer to that, frankly, is money, and we all know that is true. These companies spend enormous sums of

money trying to influence this body, trying to influence our regulators, trying to influence the Federal Government. It is time that this Congress did something about it to show who is really in charge. It is not them. It should be the people. It should be us. I am at the point I think we probably should repeal 230 completely, but we certainly have to take action to stop this kind of rampant manipulation for profit, which is what these companies do. Thank you, Mr. Chairman.

Chair COONS. Thank you, Senator Hawley. Senator Hirono is joining us next by Webex.

Senator HIRONO. Thank you, Mr. Chairman. I would like to ask Ms. Bickert to respond to three questions that I have for her very briefly. A recent report from the MIT Technology Review found that Facebook's ad delivery algorithm discriminated based on gender when surveying ads for jobs. There are numerous other examples of Facebook's system discriminating on the basis of age, gender, and race, and delivering housing, job, and financial services ads. Facebook's help center describes what factors it uses to target ads to users, including, "other information about you from your Facebook account, such as your age, your gender." Facebook also allows ads to be targeted based on things like zip codes that can be used as a proxy for race.

Question one. Are you concerned that Facebook's reliance on these factors in targeting ads results in discrimination? Can you give a "yes" or "no" answer, please, Ms. Bickert?

Ms. BICKERT. Senator, thank you. Making sure that the—that our ads are served in a fair way without discrimination is always a priority for us, and we do have policies in place to prevent discriminatory targeting. I am very happy to follow-up with you in the interest of time on some of the specifics of those policies.

Senator HIRONO. How does Facebook ensure that it does not violate, for example, civil rights laws when targeting ads for housing, employment, and financial services?

Ms. BICKERT. Senator, thank you. We have policies around when people can use certain targeting criteria. For instance, we do not allow some of the more sensitive targeting criteria that you have mentioned. We do not allow that for certain types of advertisements, such as financial services advertisements or housing advertisements. I can follow-up with you on some of those specifics on how we ensure that the ads are served in a fair way.

Senator HIRONO. The—you have addressed the concerns raised by the recent report from the MIT Technology Review that found discrimination on the basis of gender, for example, that you have addressed those concerns raised by this report?

Ms. BICKERT. Senator, I am not familiar with that specific report, so I cannot comment on that. I can tell you in my years of being in this job, making sure that we do not have discriminatory ads has been a priority for us. We have worked on that for years, made a number of improvements, and I can follow up with those details.

Senator HIRONO. Yes, I hope you can follow-up after you have read the MIT Technology Review, and I do not think they are the only ones who raise those concerns. Third question. When Facebook has been sued for discrimination by its ad targeting algorithm, it has often hidden behind Section 230. Earlier this year, I joined

Senators Warner and Klobuchar in introducing the Safe Check Act, which would remove Section 230 immunity for violations of civil rights laws. We are not talking about total removal of Section 230, but as referencing civil rights laws. Do you agree that Facebook should not be immune under Section 230 when it discriminates when delivering ads?

Ms. BICKERT. Senator, thank you for the question. I agree that there should be regulation to hold social media companies accountable. I think that there is there is a lot to consider when crafting that regulation. I put out a white paper on how I think regulation could work, and I look forward to having those conversations, and I know our team does as well with your office and other offices.

Senator HIRONO. For Mr. Donovan and Mr.—Dr. Donovan and Mr. Harris, do you think that regulation is appropriate to prevent Facebook ads from discriminating, or should we just eliminate the immunity from lawsuits for civil rights laws' discrimination? Briefly, Dr. Donovan and Mr. Harris.

Dr. DONOVAN. Yes, this is Dr. Donovan. I think that we do need to have some carve-out related to civil rights violations, especially those that would require oversight. One of the main problems here that we are addressing is that when automation is matched with amplification—that is, there is no review of these ads—we do not know who we are doing business with. Not only does Facebook not know who they are doing business with directly, but—and we see a bunch of different shell games with some of the stuff that has been implemented related to advertisers disclosing who they really are. Overwhelmingly, over the pandemic, we have seen all kinds of scams, grifts, and hoaxes that violate people's civil rights, and so we do need this.

Senator HIRONO. I appreciate that, that you think there should be a carve-out. Mr. Harris, do you think there should be a carve-out for civil rights violations?

Mr. HARRIS. I am not familiar the exact way the legislation is written, but that is—I would be sort of in support of that. I think the important thing to recognize here, as Joan was just saying, is that the companies make money by not having human oversight discernment.

Senator HIRONO. Thank you. I—

Mr. HARRIS. The premise is that more of this—yes.

Senator HIRONO. I am dying to ask this one last question, Mr. Chairman. Apple released a software update this week that gives users control over whether apps are able to track them when they use other apps and surf the web. Giving users greater control over their online privacy strikes me as a positive thing. This is for Mr. Harris. What do you think the impact of this change will be, both on the issue of misinformation and division we are talking about today, and the system of surveillance capitalism that companies, like Facebook and Google, rely on more broadly?

Mr. HARRIS. Yes. I applaud Apple for making this small step in—I think of it almost like a carbon tax on micro-targeted advertising. If we are left with this sort of extractive business model that treats us as the product and not the customer, then removing micro-targeted advertising, sort of the hyper-personalization because apps can track you across applications. Think of it as going more close

to the 1970s model of billboards, which are de-personalized, as opposed to this micro-targeted model. That is not completely true because advertisers can still micro-target you from within the Facebook ecosystem, so it is not going to affect things that much, and it will not address problems like misinformation or polarization, which will continue. It just is almost like a subtle carbon tax on the advertising business model.

Senator HIRONO. Thank you. Thanks, Mr. Chairman.

Chair COONS. Thank you very much, Senator Hirono. Senator Grassley?

Senator GRASSLEY. Thank you, Mr. Chairman. I use Twitter regularly. Facebook, Twitter, and YouTube are also popular social media platforms where users disseminate their views and opinions for billions of users around the world. Just here in the United States, in 2019, an estimated 72 percent of Americans use at least one social media site. People can make their voices heard, share their opinions, and interact. Increasingly, however, these Big Tech companies are deciding what we can and cannot say and infringing on Americans' freedom of speech.

I constantly hear from Iowans about their concerns, with control that Big Tech has over the discourse in this country as well as the biases that these platforms have against conservative voices in Middle America. I have heard numerous stories about posts being deleted, businesses removed, and creditors silenced. Many times, this happens without warning and very little, if any, due process. These platforms have monopoly powers with very few competitors and are not constrained by market forces, and consumers have no alternative. Big Tech is also immune from liability under Section 230. This immunity, combined with monopoly, allows them to censor, block, and ban whatever they want. We must look at the power and control that a handful of companies have over speech and their silencing voices with which they disagree.

My question is to Ms. Bickert, Culbertson, and Mr. Veitch—Ms. Veitch. When you decide to remove current content from platforms, do you believe that you do that consistent with First Amendment free speech principles, such as viewpoint neutrality? If you believe that you are doing that, then why is it that conservative voices are consistently the ones who are being censored?

Ms. BICKERT. Senator, thank you for the question. We are a platform for ideas across the political spectrum. I do believe that we enforce our policies without regard to political affiliation. I do hear questions from both sides of the aisle, if you will, about whether or not we are fair in our content policy enforcement, but I can tell you that we enforce our policies without regard to political ideology.

Ms. VEITCH. Senator, I also appreciate the question here. We want YouTube to be a place where diversity of viewpoints are heard. We do have public-facing community guidelines that govern what is allowed on our platform and what is not. We do enforce these consistently without regard for political viewpoint. You did mention due process, so I wanted to call out that when content is removed from a creator, a creator does receive an email explaining that and is given an opportunity to directly appeal. We make public the data around our appeals. In the last quarter of 2020, we did have 223,000 appeals and 83,000 reinstatements, showing we do

not always get this right, but we certainly want to apply our policies evenly.

Ms. CULBERTSON. As for Twitter, and, Senator, thank you for the question, and we love to see your tweets on Twitter. You are one of my favorite follows. As you probably appreciate, Twitter would not be Twitter if everyone had the same viewpoints, and we welcome diverse perspectives. It is what makes our service Twitter. We have rules in place. We enforce them impartially. I know people have concerns and they believe that companies like ours should be more transparent. That way—that is why we have put forth three core solutions which we think would go a long way to addressing some of these concerns. The first is increased transparency, the second is more user control and choice over algorithms, and the third is enhanced due process. If we do make a mistake, that users have the ability to appeal and have their decision reviewed against our terms one more time.

Senator GRASSLEY. Yes. There are countless examples of material being removed by a platform stating that it is misinformation, but it is actually just viewpoints that liberals might disagree with. What are your platforms doing to ensure that they are not using pretextual reasons to censor differing opinions? Then that is my last question.

Ms. CULBERTSON. I am happy to take this one, and Twitter has taken a very narrowly scoped focused on misinformation at this time. We have three categories that govern our policies. The first is synthetic and manipulated media, the second is civic integrity, and the third is COVID-19 misinformation. We are piloting a program called Bird Watch that would crowd source annotations to potential misinformation. This is something to address all forms of misinformation, and it would also bring more voices in to help us with that work.

Ms. VEITCH. Senator, we do have robust community guidelines on YouTube. Those exist to keep people safe. To your point, it is important to note that those community guidelines are public facing and can be reviewed by any of our users.

Senator GRASSLEY. I guess, Mr. Chairman, nobody—the third person did not want to comment, so you can—I will give up my time. Go ahead.

Chair COONS. Thank you, Senator Grassley. I appreciate that. Senator Klobuchar.

Senator KLOBUCHAR. Hard act to follow there. Okay. Thank you, Senator Grassley, for your interest in this issue. Mr. Harris, you and I were on a panel together in March, and good to see you again. Could you explain more about how companies' market power exacerbates problems of disinformation, extremist contact, and bias?

Mr. HARRIS. Yes. It is great to see you again, too. Thank you. Thank you for the question, Senator. You know, if there is anyone with an alternative model to the current problems that plague us in misinformation, disinformation, and morality, can they succeed in the marketplace? There is something in, you know, the literature called Metcalfe's Law, right, where the power of a network grows exponentially with the number of participants. Really what we have between social media platforms is a race to Metcalfe. Once

you have a dominant platform, it is very hard for there to be an alternative.

Market concentration means that even if there are alternatives that are trying to do any—and solve any of the problems we are talking about today differently, they are going to get bought up by the existing platforms. If you are a venture capitalist, the only way you are going to fund an existing company is by knowing that there is an exit pathway, and we kind of all learned the lesson as all the sort of competing platforms and things that have come out have just been acquired by the existing companies. We also—

Senator KLOBUCHAR. Yes. Could I—I think that point just cannot be lost because there are regulations we can put in place—that is one way to do it—and you can do both things at once. If you have a company that buys out everyone from under them—in the words of Mr. Zuckerberg, they would rather buy than compete—and buys companies like Instagram and/or WhatsApp, we are never going to know if they could have developed the bells and whistles to help us with misinformation because there is no competition. Do you want to comment more on that, Mr. Harris?

Mr. HARRIS. Yes. I mean, just as you said, there—if WhatsApp were to remain independent, and let us say we are living in some alternative reality where now WhatsApp was separate, and we solve these problems and WhatsApp decided they are going to spend billions more dollars on content moderation because they want to actually be the platform that people can trust. They cannot make that choice because Facebook bought them, and now they are sort of integrated in how much they are working on these problems, and it is a race to sweep the garbage under someone else's rug.

What we have seen, unfortunately, is instead of collaboration between all these platforms, in some cases, we have seen, hey, look how bad their problems are because we do not want to pay attention to ours, not, again, because they are evil. It is just game theory happening between the companies. It really does—

Senator KLOBUCHAR. Okay. Thank you. Dr. Donovan, in your research, you have looked at medical misinformation at scale and the role of social media platforms. Could you please comment on how the sheer size of a few powerful platforms affects the problems that we should be addressing?

Dr. DONOVAN. Yes, thank you, Senator Klobuchar, and I really look forward to reading your book, "Antitrust." The problem of medical misinformation, of course, is one that was exacerbated by the pandemic, but anti-vaccination activists have a long history of using social media in order to attack the public understanding of science. During the pandemic, of course, the way in which the tech companies have turned to medical misinformation is really—it is like putting a band-aid on an open wound. Right now what we need is a comprehensive plan for ensuring that people have access to timely, local, relevant, and accurate information, like public interest obligations, but instead what we have is a very slapdash approach to, you know, whatever the breaking news event is of the day. I do think that the size of the platform and the way in which medical misinformation scales much more quickly than any inter-

vention is probably the most pressing public health issue of our time.

Senator KLOBUCHAR. Okay. Thank you. Ms. Bickert, a recent poll found that nearly 1 in 4 Americans said they will not get the coronavirus vaccine. Meanwhile, a recent report from the Center for Countering Digital Hate identified 12 specific content producers as the original source of an estimated 65 percent of coronavirus disinformation online. Recently, Senator Luján and I, after he conducted a hearing, sent a letter to Jack Dorsey and Mark Zuckerberg calling on them to remove these individuals from the platforms. Do you agree that more action needs to be taken? What is your response to our letter? I guess I would start with you, Ms. Bickert and then go to you, Ms. Culbertson.

Ms. BICKERT. Senator, thank you, and thank you for the letter as well. I know that we have assessed that content and removed those accounts that were violating, and I can—I can follow-up more with you on the specific details of that.

More broadly, and I think this is a really important issue, we know that we have to get it right when it comes to misinformation around COVID. One of our goals is to help 50 million people get vaccinated. We are doing that both proactively through partnerships with local and national health authorities, making sure that we are directing people to authoritative health information, including where they can get vaccinated. We have now directed—we have connected more than 2 billion people with those authoritative health resources. We also, since the very beginning, have been partnering with the CDC to remove content that contradicts CDC guidance that could lead to an increased risk that people could contract or spread COVID. That includes removing over 12 million pieces of safety-related COVID-19 misinformation.

Senator KLOBUCHAR. Okay. In my role on the Commerce Committee, of course, Senator Cantwell is leading a bill on privacy. Do you agree that consumers should have the ability to access their data and control how it is used, including what data is used in social media company algorithms? Do you give customers that ability now for both content and advertising algorithms, Ms. Bickert?

Ms. BICKERT. Senator, thank you. We do give people a number of controls. That includes everything from the ability to download your own information, remove it, control who can see your posts, see what type of—you can opt out of our algorithm. You can see who can see your content at any time, and you can change those—

Senator KLOBUCHAR. Is the company then supportive of our bill on privacy Senator Cantwell and I—

Ms. BICKERT. I would have to have our U.S. public policy team follow up with you on the specifics of that.

Senator KLOBUCHAR. Okay.

All right. Okay. Thank you. I appreciate that. I am going to—I will ask the—I can see Senator Coons over his mask raising his eyebrows at me. That is his sign enough is enough.

Chair COONS. No, no, the Chairman welcomes additional questions from the celebrated author of an outstanding book I need to—

[Laughter.]

Senator KLOBUCHAR. Okay. I was just going to ask—I will ask one more question then of Ms. Culbertson from Twitter, just the original question that I had asked Ms. Bickert about the “Disinformation Dozen,” as we call them, the accounts online. Of course, some of these issues that I have had—I am talking about with market power, is not as applicable to Twitter, which I appreciate, but—as a competitive—a competitive platform. Could you at least answer the question here about this disinformation dozen?

Ms. CULBERTSON. Certainly. Thank you for the question. We have and are continuing to review this particular group of individuals against our policies, and we have taken enforcement action on several of these individuals. Our team will be following up this week with all the details around that. Also, I just wanted to note that while we are competitors, we are partners to address a lot of really harmful content—content issues. We have collaborated on COVID. We work together on terrorism, child sexual exploitation, opioids. I take issue with the premise that was mentioned earlier. There is collaboration across industry to address some of the most harmful content on the internet.

We also invest heavily in our partnerships with experts, especially around COVID. We worked very closely with the CDC, HHS, the White House, to not only enforce against our rules, but to also ensure that people try to have access to credible information on our service.

Senator KLOBUCHAR. Were you saying you take issue with something that I had said or was it something—

Ms. CULBERTSON. No, no, No, Senator. No, Senator. One of the other panelists—

Senator KLOBUCHAR. Oh, okay.

Ms. CULBERTSON [continuing]. Suggested that we have a competitive edge to compete on addressing these harms where we actually collaborate in a lot of these areas.

Senator KLOBUCHAR. Okay. Thank you very much. I appreciate it.

Ms. CULBERTSON. Thank you.

Chair DURBIN [presiding]. We now go to Senator Kennedy, remote. Can you hear us, Senator Kennedy?

Senator KENNEDY. I can hear you, Mr. Chairman. Can you hear me?

Chair DURBIN. Yes, I can. The time is yours. Take it away.

Senator KENNEDY. Thank you. It seems to me that in the guise of giving consumers what they want, a lot of our social media platforms first use surveillance to identify a person’s hot buttons, and then they use algorithms to show that person stuff that pushes those hot buttons. This is called, as you know, optimizing for engagement. The social media platform wants a person to visit its platform early and often. That is how it makes more money advertising. In any event, when that person that we are talking about, as a result of those algorithms, gets all revved up with no place to go, he posts something outrageous. Not every time, but quite frequently, and that is why you can still find kindness in America, but you have to go offline to do it.

Mr. Harris, I would like a straight answer from you. Would you—I have a bill—others have a similar bill—a bill to say that—

that would say that Section 230 immunity will no longer apply to a social media platform that optimizes for engagement. Would you—if you were a Senator, would you vote for it?

Mr. HARRIS. I would have to see the way that the bill is written, Senator.

Senator KENNEDY. Do not do—do not do that to me, Mr. Harris. Give me a straight answer. We all want to read the bills. Would you vote for it or not?

Mr. HARRIS. I would—I would be in support of a bill that had technology companies not measure as their primary mode of success any of the engagement metrics—time spent, clicks shared, et cetera.

Senator KENNEDY. That is swell, but if the bill said—I do not like to waste time in these hearings. If the bill said no Section 230 immunity if you optimize for engagement, would you vote for it? If you do not want to—if you do not want to answer, just tell me.

Mr. HARRIS. It sounds like a very interesting directional proposal. I just—I would have to know the details, but I am sorry for not being more clear.

Senator KENNEDY. You are being very clear. You are dodging the answer. Ms.—Dr. Donovan, would you vote for it?

Dr. DONOVAN. Yes. When it comes to bills, the reason why I am in research is so I do not have to make those decisions. I would say that when we are talking about what these companies optimize for and the way in which it is optimized—

Senator KENNEDY. Doc. Doc. Doc. Doc.

Dr. DONOVAN. Yep. Please.

Senator KENNEDY. Doc. Doc, would you vote? Would you vote for the bill?

Dr. DONOVAN. I would—I would vote for some form of bill that required oversight of these algorithmic systems.

Senator KENNEDY. All right. I mean, we have these hearings and I appreciate them, but we never get down to it. Everybody just wants—you know, we all talk. I am as guilty as anyone else. At some point you got to get down to it, and if you say—and that is where I am coming from. I am not trying to be rude. I am just trying to get an answer out of you. You have both been very critical of what we have today. I am, too. I am looking for solutions. I am not just looking to—for us all to show how intelligent we are—

Dr. DONOVAN. I think that one of the things—

Senator KENNEDY [continuing]. Or not.

Dr. DONOVAN [continuing]. That we could address, Senator—

Senator KENNEDY. I appreciate it, Doc. I am going to run—I am going to run out of time. Let me ask—I am thinking about introducing a bill—in fact, we are working on it—to take the principles of the general data protection regulation in the EU. I never thought I would do something like this, but take the principles and the general data protection regulation in the EU and have that—those principles apply here in the United States. Ms. Bickert, would you support that bill?

Ms. BICKERT. Senator, I focus on content, but there are people at our company, and we can have someone follow-up on that.

Senator KENNEDY. That is a dodge. Ms. Culbertson, would you vote for it?

Ms. CULBERTSON. We certainly comply with GDPR. There are some tensions with the First Amendment in the U.S., but we would welcome a longer conversation about this. Generally, yes.

Senator KENNEDY. Yes?

Ms. CULBERTSON. Yes, Senator.

Senator KENNEDY. Oh, God bless you. God bless you. Thank you for an answer. Ms.—I am sorry, but I am mispronouncing your name. “Vytch?” “Veetch?”

Ms. VEITCH. Senator, it is “Veetch,” yes.

Senator KENNEDY. I am sorry, Ms. Veitch. I apologize. Will you vote—would you—if you were a Senator, would you vote for it?

Ms. VEITCH. Senator, I am not an expert on GDPR. I can tell you on privacy, what we want to do is give our users security—

Senator KENNEDY. I know. I know. You want—you want privacy, but your whole model is built around finding out everything you can about me, other than my DNA, and you may it, for all I know. I am not trying to be rude, but I cannot tell you the number of these hearings I have been to, and there—I learn something every time. When we get down to it, what are we going to do about it? Nobody wants to answer, and you are supposed to be our experts. I would strongly encourage you to come to these hearings with positions, firm positions, on behalf of yourselves, or on behalf of your companies, that you are ready to take. Do not just word whip us. We are trying to solve a problem here. What—

Chair DURBIN. Senator Kennedy.

Senator KENNEDY. Yes, sir.

Chair DURBIN. I have to ask you for a “yes” or “no” answer. Do you realize you have gone over time?

Senator KENNEDY. I realize that—yes, and I realize everybody else has gone over time.

[Laughter.]

Chair DURBIN. Take another minute, and then please wrap it up.

Senator KENNEDY. I am done.

Chair DURBIN. Thank you, sir. Senator Ossoff, remote?

Senator OSSOFF. Thank you, Mr. Chairman. Thank you to the panel. Ms. Bickert, much of the public discussion is focused on Facebook’s moderation practices, but there is a compelling argument that the real problem is not the quality of your moderation policies or the nature of the algorithm, but the underlying business model, your scale, and your power. While you clearly have an obligation to remove certain content, for example, incitement to violence or hate speech, I am not at all enthusiastic about huge multinational tech companies becoming the arbiters of legitimate speech and expression, especially when the decisions about what you may boost or suppress algorithmically are often made in secret and under heavy pressure from politicians, and advertisers, and public opinion.

On the subject of your scale and your power, I would like to ask, does Facebook anticipate that it will embark on further acquisitions of competitor services in light of the suit that you are already facing from the FTC and a number of State attorneys general alleging that your acquisitions of Instagram and WhatsApp constituted anticompetitive activity?

Ms. BICKERT. Senator, thank you for the question. Of course I cannot comment on any litigation. I can tell you, because I am responsible for our content policies and a lot of what we do around moderation, that we do take very seriously both the balance between expression and safety, but also the need for transparency. So with, for instance, our algorithm, over the past few years, we have put out a number of blog posts and other communications where we have actually given the inputs for what goes into the ranking algorithm. We have explained any significant ranking changes. We have introduced this tool where on any post on Facebook, you can click on it and go under “why am I seeing this post,” and it will tell you why that is appearing in your News Feed where it is. Then, significantly, we have made it more visible how you can opt out of that News Feed ranking algorithm.

If people just want to see their content in reverse chronological—reverse chronological order—excuse me—they can—

Senator OSSOFF. Ms. Bickert, yes, respectfully, and I greatly appreciate your response, and I heard some of these points earlier in the hearing, and I am not asking you to comment on any specific litigation. To be clear, my point is actually that everything you just said about improving the quality of your moderation practices, disclosing some of the decisions underlying the algorithm, are not the root issue. The root issue is that Facebook has too much power, and one company perhaps should not be such a massive gatekeeper that determines what ideas prosper and what ideas do not. That is why the question that I asked was, does Facebook anticipate that it will embark on any further acquisitions of competitor services.

Ms. BICKERT. Senator, this is—acquisitions is really not my area at all. I am focused on content. I can tell you, though, from where I sit, from my perspective, it is a highly competitive space, and I know that not only from, you know, being the—being an executive working on content at Facebook, but also being the parent of two teenage daughters, both of whom use social media, and there are a lot of services out there that people use. Nevertheless, I do think it is really important that we recognize that these content modernation rules are really important, and we have to be very transparent about what they are so people can make informed choices about whether or not they want to use our services.

Senator OSSOFF. Thank you, Ms. Bickert. Ms. Bickert, Apple’s recent iOS update will require apps to seek additional explicit authorization from users in order for those apps, presumably some of your products included, to continue tracking users across the internet. Tracking cookies and other technologies allow Facebook and other entities to monitor virtually all of their users’ web browsing activity. I want to commend Apple for taking this step and ask whether you will take significant steps in the short term to reduce your tracking—your ubiquitous tracking of your users’ web activity, location data, the technology that they use, and whether you will consider extending the feature that allows the removal of personal data from Facebook, to include the removal of personal data not just from Facebook, but from any entities to whom Facebook sold such data, and including in your contracts with those to whom you

sell data, a provision that they must delete all data that they have purchased from Facebook at the command of the user.

Again, it is two questions. Will you follow Apple's lead in ceasing tracking of users across the web, and will you include in contracts with those to whom you sell data a provision requiring them to permanently delete and verify the deletion of all data you have sold to them about any user who activates the Facebook feature to remove their data from Facebook? Thank you so much.

Ms. BICKERT. Senator, thank you for the question. First, let me be really clear. We do not sell user data. That is not the way our advertising works. The way that it works is an advertiser selects from among different targeting criteria, and then we deliver that ad to a relevant audience. We can follow-up with more details on how that works.

With respect to controls, I know we have introduced controls around people's off-Facebook experience. I am not an expert in that area. There are those in the company who are, so I can get that information and follow-up with you.

Senator OSSOFF. Thank you, Ms. Bickert, and thank you, Mr. Chairman.

Chair DURBIN. Thank you, Senator. Senator Blackburn, are you available by remote?

Senator BLACKBURN. Yes, I am. Thank you, Mr. Chairman. I appreciate the witnesses and the hearing today, and I think all the witnesses are hearing that Americans are pretty much fed up with the arrogance of Big Tech. You are seeing it from all sides, and certainly Twitter's CEO, Jack Dorsey, was—had his contempt for Congress on full display in the House Energy and Commerce Committee hearing last—I think it was last month. He tweeted out a poll on possible answers to the questions, basically treating the hearing as a joke. Ms. Culbertson, do you agree it is unacceptable for Twitter's CEO to tweet while he is testifying before Congress? Yes or no.

Ms. CULBERTSON. Certainly, he is the CEO and creator of Twitter, and he likes to tweet, and that is the way he communicates.

Senator BLACKBURN. Okay. I asked for a "yes" and "no," but I will say I am pleased you are looking and appearing more presentable than your CEO in his testimonies before us. When he behaves disrespectfully in a congressional hearing and before the American people, he embarrasses Twitter. It is just such proof of how out of touch Big Tech is with the rest of the country. Big Tech is, in my opinion, destroying news, free speech, competition, original content. It is responsible also for much of our children's minds. This is something that bothers me as a mom and a grandmom, the power of Facebook and YouTube's algorithms to manipulate social media addiction. We are even reading that it is among babies, toddlers, kids, tweens, and teens, and this is something that should terrify each of us.

YouTube deploys algorithms to breed this addiction, clickbait in children, and they do it because it pays well. Our children's brains are being trashed so if you Silicon Valley CEOs can pocket billions of dollars in ad revenue. YouTube algorithms create an unpoliced automated reward system. Videos with little educational content are amplified to unsuspecting toddlers and kids and to their

unsuspecting parents. Senator Thune mentioned that we are reintroducing the bipartisan Filter Bubble Transparency Act to force Big Tech to disclose if their secret algorithms are manipulating customers.

Ms. Veitch, YouTube has a history of exploiting children to harvest and profit off of their viewing history. Is it not true YouTube has illegally collected data on kids under age 13 in violation of COPA, and marketed that data to companies? Ms. Veitch.

Ms. VEITCH. Thanks for the question, Senator. I am familiar with COPA that you are referring to. That was a novel interpretation of COPA. We worked directly with the FTC to reach an agreement about how we treat made-for-kids content on YouTube main. We do—

Senator BLACKBURN. Okay. Ms. Veitch, Yes, you reached an—a settlement in 2019. You were fined a record \$170 million. Do you recall that?

Ms. VEITCH. Yes, ma'am.

Senator BLACKBURN. Okay. The FTC order does not require YouTube to police the channels that deceive by mis-designating their content. However, Commissioner Slaughter said YouTube should have to take the extra step of creating an algorithmic classifier to better police YouTube content for kids. I know your engineers are capable of designating and designing algorithms for all sorts of purposes, good and evil. Let me ask you this. Is the YouTube engineering team capable of designing an algorithm that can identify a designated and child-directed content and turn off behavioral advertising?

Ms. VEITCH. Senator, they are capable of that, and they have done that. We do require creators to designate their content as made for kids or not, but we also run classifiers, as you mentioned, to check that system and to determine what content is appropriate to be made for children and serve to children. We also, just to be clear, Senator, do not allow personalized advertising on made-for-kids content.

Senator BLACKBURN. Are you prioritizing profit over children?

Ms. VEITCH. No, Senator. Child safety on our platform is our top priority. We build our product with parental controls baked right in, things like timers—

Senator BLACKBURN. The FTC is prioritizing children and taking steps to safeguard them. Under the settlement, you promised to stop illegally marketing targeted ads to children. Videos now have been labeled as made for kids, as you just mentioned. So, and made-for-kid videos will no longer include a comment section or in screens that allow viewers to subscribe to children. Are you allowing this behavioral advertising to be turned off?

Ms. VEITCH. Yes, Senator. We do not serve personalized advertisements on made-for-kids content.

Senator BLACKBURN. Okay. I am over my time. Ms. Bickert, I have a question for you I will submit for the record. Thank you, Mr. Chairman.

Chair COONS [presiding]. Thank you, Senatr Blackburn. Senator Blumenthal.

Senator BLUMENTHAL. Thank you, Mr. Chairman. Thank you to all of our witnesses for being part of this hearing and to the Chair-

man for holding it. It is a very, very critically important topic and hearing, and I apologize that I am late coming here because I was chairing a Subcommittee of Commerce on Consumer Protection dealing with COVID scams.

I am very proud that last week, the United States Senate approved the bipartisan Jabara-Heyer NO HATE Act, which I led alongside Senator Moran. We have known for a long time that hate crimes are on the rise. They are exploding in this very polarized and vitriolic time. Viral videos of individual crimes posted by—on Facebook, Twitter, and YouTube, no matter how horrifying or stomach-turning, really tell only part of the story. The NO HATE Act will improve hate crime reporting because so many of them are invisible and unreported, and it will expand assistance and resources for victims of hate crimes as well as for law enforcement, and hopefully will enable us to understand the full scope of the problem so that we can take more effective action against hate crimes.

We know that the tech platforms play a role in hate crimes and hate speech online and off. The Anti-Defamation League recently found that as many as one in three Americans experience hate crimes and harassment online. Following the ADL's concern—very concerning findings, I teamed up with Representative Raskin to request a Government Accountability Office study specifically on the prevalence of online hate crimes and hate speech in the United States. During the 2020 election, Facebook spoke about the break-the-glass measures it was taking to “dial down” the hate, incitements to violence, and misinformation on its platform. Last week, Ms. Bickert, you wrote a blog post about turning the dial down on hate speech, graphic violence, violence and incitement, as the country was anticipating the verdict in the *Chauvin* trial.

If Facebook does, in fact, have a dial for hateful content, can the company dial it down now? Why does not it dial it down already? To all of the representatives who are here today from YouTube, Twitter, as well as Facebook, can you commit to providing access data to independent researchers to help us better understand and address the scourge of hate and harassment online?

Ms. BICKERT. Senator, thank you, and let me start by saying I completely agree that the rise of hate speech and hate crimes is very concerning and needs to be a priority for us and is a priority for us. I will just point to one quick example, which is we have started publishing the prevalence in our quarterly reports that we put out, our community standards enforcement reports. We now publish the prevalence of hate speech, which means we go through with a fine-tooth comb and see what we missed for a significant—statistically significant subset of content. The prevalence of hate speech on our service is very low—less than a 10th of 1 percent—but it is something that we are really focused on finding. I am happy to say that more than 95 percent of the content that we removed for hate speech violations we find ourselves before anybody reports it to us, so we are making strides.

To respond to your point about why we do not—the measures we took around the *Chauvin* trial and the election, why we do not always do that, let me—let me sort of give you an example of the cost of those measures because they have benefits, but they have costs.

In the—in the run-up to the election, for instance, we took some very aggressive measures to reduce the distribution of content that might be violating our policies. We did that with the *Chauvin* trial as well. Those measures are not perfect, and so there will be content that actually does not violate our policies that was flagged by our technology that really should not be reduced.

When we take those measures, we are mindful of the cost. It is always this balance between trying to stop abuse and trying to make sure that we are providing space for freedom of expression and being very fair. We take those measures where there is a risk of false positives only when there is an additional risk of abuse.

Senator BLUMENTHAL. Thank you. My time has expired. Thank you very much. Thanks, Mr. Chairman.

Chair COONS. Thank you so much, Senator Blumenthal. We are going to have a second round of questioning that may be participated in only by the Ranking Member and myself given that there are votes actively ongoing on the floor. Let me thank our five witnesses again and the many Members of this Subcommittee who have come to question.

Ms. Veitch, I understand that 70 percent of the views on YouTube come by—are driven by its recommendation algorithm. With 2 billion users worldwide and over 1 billion hours of video watched each and every day, that makes your recommendation algorithm incredibly powerful. Members of the public can see how many times any video has been viewed, but members of the public cannot see how many times that video has been recommended, though I understand YouTube does collect this information and gives it to content providers. If a video ends up getting taken down by YouTube for violating its content policies, we have no way of knowing how many times it was recommended by your algorithm before it was ultimately removed. Could YouTube commit today to providing more transparency about your recommendation algorithm and its impacts?

Ms. VEITCH. Thanks for this question, Senator. Just generally speaking, if content violates our policies, we want to remove it as quickly as possible. As you will see in our Public Community Guidelines Enforcement Report, of the 9.3 million videos we removed in Quarter 4 of 2020, more than 70 percent were removed before they had 10 views. I think you have brought up an interesting idea, and we are always looking to expand transparency when it comes to our platform. One way we have done this recently is by making public what we call our violative view rate. It is the percentage of views on our platform that violate our community guidelines. Last quarter they were between .16 and—

Chair COONS. Ms. Veitch? Ms. Veitch, if I might, I just want to know if you are willing to release the data. I believe you are already collecting about how many times videos that violate your content standards have been recommended by your recommendation algorithm.

Ms. VEITCH. Thank you, Senator. I cannot commit to really saying that today, but it is an interesting idea. We want to be more transparent, so let us work with you on that.

Chair COONS. Thank you. I look forward to getting an answer as soon as is reasonably possible. Ms. Bickert, several publications

have reported that significant portions of misinformation and polarizing content on Facebook comes from readily identifiable, hyper-active users, or super inviters, who generate a lot of activity on your platform. Dr. Donovan can you comment briefly on how these hyper-active users create problems, and then, Ms. Bickert, I want to ask about whether or not Facebook intends to tackle this challenge. Dr. Donovan.

Dr. DONOVAN. Yes, you are referring to the BuzzFeed article that reported on an internal memo from Facebook that showed that there is a power law at play where it skews to highly—you know, misinformation tends to be most potent when you have a densely networked and highly coordinated small group of people working essentially around the clock to try to get their groups stocked with the public. What has been interesting about reading the document internal to Facebook is that, even as they tried to counter super inviters, their own internal systems and teams were not able to overcome that coordinated small network. There is a lot that the company needs to do to address adversarial movements, and in this case was—they were looking at the formation of Stop the Steal groups and the Patriot Party.

Chair COONS. Thank you, Dr. Donovan. Ms. Bickert, the Wall Street Journal reported last year that Facebook considered seriously, but ultimately declined, to take measures that would put limits on these users' activities. There was a proposal reportedly called Sparing Sharing, which would have reduced the spread of content that was disproportionately favored by these so-called hyper-active users. Could you speak to how Facebook is intending to approach this issue?

Ms. BICKERT. Yes, Senator, and let me say we did actually put a restriction, a limit on the number of invites that an individual user could send out in a day to a group during the election period. I want to speak also to the point that Dr. Donovan raised, and I completely agree. There are—there are networks of bad actors, who are particularly sophisticated, who try to target—use social media to achieve their objectives. Understanding the way that those networks work has been something that we have really been focused on in the past few years, building a team under Nathaniel Gleicher, who has got expertise in this area and I know knows Dr. Donovan as well.

In terms of identifying these sophisticated actors, who are often engaged in shell games and, you know, other attempts to sort of obfuscate what they are doing using inauthentic identities, we have gotten far better at that. We have removed more than 100 such networks since the beginning of 2017. We are public about it when we do. We publish the results of those, and we have also gotten better generally at identifying fake accounts. We removed more than 1 million fake accounts at the time—up or near the time of upload every day now.

Chair COONS. Thank you. I look forward to delving into this further with you and with other folks at Facebook. Let me just ask two more—maybe three more, Mr. Ranking Member. A quick just structural question. I know it is common for employees at major tech companies to be required to sign nondisclosure agreements as a condition of employment. When I was in the private sector, that

was a common practice in the businesses that I knew about and practiced for. Ms. Bickert, Ms. Culbertson, Ms. Veitch, do each of your companies generally require your employees to sign NDAs? It strikes me as a “yes” or “no” question.

Ms. BICKERT. I will go first, Senator. I do not know the answer, but I will follow-up—I will have the team follow-up with you.

Ms. VEITCH. Senator, I want to be careful here because I am not a lawyer—an employment lawyer, but I do believe that we have standard agreements to protect proprietary information with our employees.

Ms. CULBERTSON. I would want to come back to you with the answer, but, of course, we have certain provisions in place to make sure people are not sharing private data they might be handling. I would just say generally, the Twitter-y spirit among our employees is to share their perspectives. You will oftentimes see our employees tweeting about our different products and services.

Chair COONS. Thank you. In general, my concern is that if a former employee from one of your companies wants to question or criticize the company or its decision making, that they might risk facing legal action. Mr. Harris, Dr. Donovan, I would welcome some more input from you following this hearing on that dynamic, and whether or not NDAs actually prevent some of the most relevant information about algorithms from getting out to the general public.

Two last questions, if I might, one on transparency. I appreciate—I appreciate the information that has been shared today about how algorithms work at a high level. Many independent researchers have said it is critical to know the details, the dials and knobs of algorithms to understand how components that drive decisions are weighted, so how much a metric, like meaningful social interaction, is actually correlated with growth and engagement, which, as Mr. Harris has repeatedly asserted, and as I fundamentally believe, the business model of social media requires you to accelerate. Given the immense impact of the knobs and dials of your algorithms in potentially both positive and negative ways, I think greater transparency about those matters, about how your algorithms actually work and about how you make decisions about your algorithms, is critical.

Ms. Bickert, Ms. Veitch, Ms. Culbertson, could you speak to whether your companies are considering the release of more details about this kind of information or other types of enhanced transparency measures or audits about the impact of your companies’ algorithms moving forward?

Ms. CULBERTSON. I am happy to start. We are constantly thinking about how we can be more transparent about any actions that we take or our systems in place, including our algorithms. That is why we are investing in our responsible machine learning initiative. We would be happy to provide more details in the interest of time.

We have an interdisciplinary group at Twitter looking at our algorithms, our machine learning, studying these—our machine learning. We will also be sharing some of our findings with the open—with the public so we can be open throughout this process.

Then just more broadly, we totally agree that we should be more transparent. We should also provide more consumer control and choice. We are also committed to improving procedural fairness. To those first two points, we have invested in this independent project called BlueSky, which is aimed at creating open protocols, which would essentially potentially create more controls for the people who use our services as well as transparency.

Chair COONS. Thank you. My last comment will be just this one. Mr. Harris spoke forcefully and pointedly about how the business model of social media's attention harvesting, and that after a decade of the positive and negative impacts of social media, which has accelerated to be one of the most important forces in our society today, that we have more than not seen the toxic impacts of division and disinformation. Mr. Harris has asserted that your entire business model is based on dividing society, and that as we transition into a digitized society in the 21st century, in order for Western open democratic societies to survive, we have to develop model humane standards for how social media works. It is my hope, I will share with my Ranking Member, that the next time we convene, it might be to consider what sorts of steps are possible, necessary, or appropriate to make that progress that Mr. Harris speaks about.

To my Ranking Member, Senator Sasse.

Senator SASSE. Thank you, Chairman Coons, and, again, thank you to all five of you for being here. I do—I do want to put another question to Mr. Harris, but before I start that second round of questioning, I would like to just briefly address colleagues on both sides of the aisle because both Republican and Democratic colleagues today have said a number of things that presumed more precision about the problem than we have actually identified here, and then sort of picked up the most ready tool, usually the 230 discussion.

I am—I think I am a lot more skeptical than maybe most on this Committee to push to a regulatory solution at this stage. I think, in particular, some of the conversations about Section 230 have been well off point to the actual topic at hand today. I think much of the zeal to regulate is driven by short-term partisan agendas, and I think it would be more useful for us to stick closer to the topic that the Chairman identified from this hearing.

I also think it is important for Members of Congress to constantly remind ourselves that we are bound by First Amendment constraints in our job, and many—a number of the lines of questioning, again, on both the right and left sides of this panel today talked as if the First Amendment is sort of, you know, this marginal topic that we do not have to be obsessively concerned about, and yet we need to draw a distinction between the First Amendment and the true public square as regulated by the powers of the government, and the fact that the companies we are talking about—Amy Klobuchar has raised some important topics about scale and antitrust issues, but the companies we are talking about are private companies. I do think there are a number of First Amendment, public-private distinctions that we should—we should be attending to a little more closely than maybe we did today.

Mr. Harris, can you tell us what discussions you have seen or been a part of, either inside the extant companies or, you know, at

the VCPE environment, potential different business models besides an ad revenue-centric business model? Can you just give us a kind of blue sky on that question?

Mr. HARRIS. Yes, fantastic question. Thank you, Senator. I mean, obviously there are subscription models. There are public interest models, more like Wikipedia. I want to make an additional distinction, which is not just the funding model, but it is the design model. The engagement and advertising model works because of the design that says, user-generated content, we are all the unpaid journalists. Previously, you had to pay a journalist at Fox News and New York Times \$100,000 a year to write content to get people to look at it, and that is the cost of attention production. What if you could harvest each of us as useful idiots to take our 5 minutes of moral outrage, and then use that to generate attention production for free? We are the unpaid labor, like, for the attention economy that is sort of duped into sharing information with each other, which reduces the costs for all these technology companies.

Then on the editorial side, instead of paying an editor at a New York Times, at a Fox News, at a whatever, \$100,000 a year, \$200,000 a year, we actually have algorithms, which we also do not have to pay, to randomly sort that to people. This happens in a values-blind process, which means that, in general, you get harm showing up in all of the blind spots. Suddenly Joan wakes up and says, hey, there is this problem, there is this problem, this problem. The companies will respond and say, okay, fine, we will take the whack-a-mole stick and we will deal with those three problems.

In general, values blindness destroys our democracy faster than people like, you know, Renee, and, Joan, and so many of our friends in this community are essentially raising the alarms about it, and that is fundamentally the kind of core design model more so than the funding model. We could have public interest technology that is funded for public interest. We could tax these companies to put into a regenerative fund. There are a whole bunch of models we could do. One in energy just like—you know, energy companies have this perverse incentive where they make more money the more energy you use, so theoretically, leave the lights on, leave the faucets on, we make more money. They do not—they do not do that because, instead, they have a model that is regulated, so that after a certain amount, they double charge you, triple charge you to disincentivize your energy use. Then that, instead of going into the private business model of the company, the balance sheets, it gets put into a regenerative fund to increase the transition to solar.

Imagine that technology companies, which today profit from, you know, an infinite amount of engagement, only made money from a small portion of that, let us say some small amount of time, and the rest basically with tax to put into a regenerative public interest fund, and this funded things like the Fourth Estate, fact checkers, researchers, public interest, technologists, things like that because what we really need to do is, as we said, organize a comprehensive shift to more humane technology, so a digital open society can compete with digital closed societies.

Senator SASSE. Very helpful. There is a—if we had—if we had more time, I was going to ask some questions about whether or not

you think there are—given your role as an ethicist, whether or not there are debates inside the company about what the optimal user time on a site is every given day, and is there a distinction between, you know, a fully consenting, -assenting 49-year-old, like myself, and how those platforms think about it for a 13-year-old and a 17-year-old as well. I know that Chris and I both need to go and vote, so I will just echo the thanks to all five of you for the fulsome discussion today and to be continued.

Chair COONS. Thank you, Senator Sasse. Let me conclude by thanking all five of our witnesses for appearing today and to my 11 colleagues who have appeared and engaged in robust questioning. I appreciate, in particular, the willingness of witnesses from Facebook, Twitter, and YouTube to answer some direct and difficult questions about their platforms and their business models. I am encouraged to see that these are topics that are broadly of interest and where, I believe, there could be a broadly bipartisan solution.

None of us wants to live in a society that, as a price of remaining open and free, is hopelessly politically divided or where our kids are hooked on our phones—their phones and being delivered a torrent of reprehensible material. I also am conscious of the fact that we do not want to needlessly constrain some of the most innovative, fastest-growing businesses in the West. Striking that balance is going to require more conversation, and I look forward to continuing to work with Ranking Member Sasse on these matters, whether by roundtable or additional hearing, and whether by seeking voluntary reforms, regulation, or legislation. That includes exploring how best to align incentives, both within companies and with the rest of our society, to ensure greater transparency and user choice, and I think we have to approach these challenging and complex issues with both humility and urgency. The stakes demand nothing less.

Members of the Committee may submit questions for the record for the witnesses. They are due by 5 p.m., 1 week from today, on May 4th.

With that, this hearing is adjourned.

[Whereupon, at 12:18 p.m., the hearing was adjourned.]

[Additional material submitted for the record follows.]

**HEARING BEFORE
THE UNITED STATES SENATE JUDICIARY COMMITTEE
SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW**

April 27, 2021

Testimony of Monika Bickert
Vice President for Content Policy, Facebook

I. Introduction

Chairman Coons, Ranking Member Sasse, and distinguished members of the Subcommittee, thank you for the opportunity to appear before you today. My name is Monika Bickert, and I am the Vice President of Content Policy at Facebook. Prior to assuming my current role, I served as lead security counsel for Facebook, working on issues ranging from children's safety to cybersecurity. And before that, I was a criminal prosecutor with the Department of Justice for eleven years in Chicago and Washington, DC, where I prosecuted federal crimes, including public corruption and gang violence.

Facebook is a community of more than two billion people, spanning countries, cultures, and languages across the globe. Every day, members of our community use Facebook to connect and share with the people they care about. These personal interactions are at the core of our mission to give people the power to build community and bring the world closer together.

It is important to us that people find content that is meaningful to them on our platform, and our algorithms help them do just that. We also understand that people have questions about how these algorithmic systems work. I look forward to discussing today the ways in which Facebook is already working to provide greater transparency and the additional steps we are taking to put people even more firmly in charge of the content they see.

II. Algorithmic Ranking on Facebook

Facebook uses algorithms for many of our product features, including to enable our search function and to help enforce our policies. But when people refer to Facebook's "algorithm," they are often talking about the content ranking algorithms that we use to order a person's News Feed.

The average person has thousands of posts in her News Feed each day. This includes the photos, videos, and articles shared by the friends and family she chooses to connect to on the platform, the Pages she chooses to follow, and the Groups she chooses to join. Most people don't have time to look at all of this content every day, so we use a process called ranking to sort this

content and put the things we think you will find most meaningful closest to the top of your News Feed. This ranking process is personalized and is driven by your choices and actions.

To make sure you don't miss meaningful content from your friends and family, our systems consider thousands of signals, including, for example, who posted the content; when it was posted; whether it's a photo, video, or link; and how popular it is on the platform. The algorithms use these signals to predict how likely content is to be relevant and meaningful to you: for example, how likely you might be to like it or find that viewing it was worth your time. The goal is to make sure you see what you find most meaningful—not to keep you on the service for a particular length of time.

Notably, in 2018, we changed the way we approached News Feed rankings to focus not only on serving people the most relevant content, but also on helping them have more meaningful social interactions—primarily by doing more to prioritize content from friends, family, and Groups they are part of. We recognized that this shift would lead to people spending less time on Facebook, because Pages—where media entities, sports teams, politicians, and celebrities, among others, tend to have a presence—generally post more engaging (though less personally meaningful) content than a user's personal friends or family. The prediction proved correct; the change led to a decrease of 50 million hours' worth of time spent on Facebook per day, and we saw a loss of billions of dollars in the company's market cap. But we view this change as a success because it improved the experience of our users, and we think building good experiences is good for the business in the long term.

III. Increasing Transparency and Control

This sifting and ranking process results in a News Feed that is unique to each person. Naturally, users don't see the computer code that makes up the algorithm, but we do share information about how the ranking process works, including publishing blog posts that explain the ranking process and announce any significant changes.

Of course, not everyone is going to read our blogs about how the systems work, so we're also doing more to communicate directly to people in our products. For some time, people on Facebook have been able to click "Why Am I Seeing This?" on any ad they see to learn why that ad was placed in their News Feed, and they're also able to change their advertising preferences. This real-time transparency and control approach has helped improve the Facebook experience for many people. Starting in 2019, we launched a similar "Why Am I Seeing This?" tool to help people understand why a particular post showed up where it did in their News Feed. To access it, people simply need to click on the post itself, then click on "Why am I seeing this post?," and they will see information about why that post appears where it does. This tool also provides easy access to their News Feed Preferences, so they can adjust the composition of their News Feed.

We have increased the control that people have over their News Feed so that they know they are firmly in charge of their experience. For instance, we recently launched a suite of product changes to help people more easily identify and engage with the friends and Pages they care most about. And we're placing a new emphasis not just on creating such tools, but on ensuring that they're easy to find and to use.

A new product called Favorites, which improves on our previous See First control, allows people to select manually the friends and Pages that are the most meaningful to them. Posts from people or Pages that the user selects will then be shown higher in that user's News Feed and marked with a star. A person can even choose to see a feed of only the content that comes from those Favorite sources in a new version of News Feed called the Favorites feed.

Facebook users can also choose to reject the personalized ranking algorithm altogether and instead view their News Feed chronologically, meaning that their News Feed simply shows them the most recent posts from their eligible sources of content in reverse chronological order.

So that people can seamlessly transition among standard News Feed, Favorites feed, and the chronological Most Recent feed, Facebook now provides a bar on the site where users can select which version of News Feed they want to see.

As we work to enhance transparency and control, we're also continuously improving the way our ranking systems work so that people see what's most meaningful to them. Just last week, we announced that we are expanding our work to survey people about what's most meaningful to them and worth their time. These efforts include new approaches to take into account whether people find a post inspirational, whether they are interested in seeing content on a particular topic, or whether certain content leaves people feeling negative. We are also making it easier to give feedback directly on an individual post. We believe that continuing to invest in new ways to learn more about what people want (and don't want) to see in News Feed will help improve the ranking process and the user experience. We'll continue to incorporate this feedback into our News Feed ranking process in the hopes that Facebook can leave people feeling more inspired, connected, and informed.

IV. Working to Combat Harmful Content and Misinformation

Of course, News Feed ranking isn't the only factor that goes into what a person might see on Facebook. There are certain types of content we simply don't allow on our services. Our content policies, which we call our Community Standards, have been developed over many years with ongoing input from experts and researchers all over the world. We work hard to enforce those standards to help keep our community safe and secure, and we employ both technology and

human review teams to do so. We publish quarterly reports on our work, and we've made significant progress identifying and removing content that violates our standards.

We recognize that not everyone agrees with every line in our Community Standards. In fact, there is no perfect way to draw the lines on what is acceptable speech; people simply do not agree on what is appropriate for discourse. We also recognize that many people think private companies shouldn't be making so many big decisions about what content is acceptable. We agree that it would be better if these decisions were made according to frameworks agreed to by democratically accountable lawmakers. But in the absence of such laws, there are decisions that need to be made in real time.

Last year, Facebook established the Oversight Board to make an independent, final call on some of these difficult decisions. It is an external body of experts, and its decisions are binding—they can't be overruled by Mark Zuckerberg or anyone else at Facebook. Indeed, the Board has already overturned a number of Facebook's decisions, and we have adhered to the Board's determinations. The Board itself is made up of experts and civic leaders from around the world with a wide range of backgrounds and perspectives, and they began issuing decisions and recommendations earlier this year.

If content is removed for violating our Community Standards, it does not appear in News Feed at all. Separately, there are types of content that might not violate Facebook's Community Standards and are unlikely to contribute to a risk of actual harm but are still unwelcome to users, and so the ranking process reduces their distribution. For example, our algorithms actively reduce the distribution of things like clickbait (headlines that are misleading or exaggerated), highly sensational health claims (like those promoting "miracle cures"), and engagement bait (posts that explicitly seek to get users to engage with them). Facebook also reduces distribution for posts deemed false by one of the more than 80 independent fact-checking organizations that evaluate the accuracy of content on Facebook and Instagram. So overall, how likely a post is to be relevant and meaningful to you acts as a positive in the ranking process, and indicators that the post may be unwelcome (although non-violating) act as a negative. The posts with the highest scores after that are placed closest to the top of your Feed.

Facebook's approach goes beyond addressing sensational and misleading content post by post. When Pages and Groups repeatedly post misinformation, Facebook reduces their overall distribution. If Groups or Pages repeatedly violate our Community Standards, we restrict or remove them.

The reality is that it's not in Facebook's interest—financially or reputationally—to push users towards increasingly extreme content. The company's long-term growth will be best served if people continue to use and value its products for years to come. If we prioritized trying to keep a

person online for a few extra minutes, but in doing so made that person unhappy or angry and less likely to return in the future, it would be self-defeating. Furthermore, the vast majority of Facebook's revenue comes from advertising. Advertisers don't want their brands and products displayed next to extreme or hateful content—they've always been very clear about that. Even though troubling content is a very small proportion of the total content people see on our services (hate speech is viewed 7 or 8 times for every 10,000 views of content on Facebook), Facebook's long-term financial self-interest is to continue to reduce it so that advertisers and users have a good experience and continue to use our services.

V. Conclusion

Facebook's algorithms are a key part of how we help people connect and share, and how we fight harmful content and misinformation on our platform. We will continue to do more to help people understand how our systems work and how they can control them. This is an area where we are investing heavily, and we are committed to continuing to improve.

Thank you, and I look forward to your questions.

**United States Senate Committee on the Judiciary
Subcommittee on Privacy, Technology, and the Law**

**Testimony of Lauren Culbertson
Head of U.S. Public Policy
Twitter, Inc.**

April 27, 2021

Chairman Coons, Ranking Member Sasse, and Members of the Subcommittee:

Thank you for the opportunity to appear before you today to provide testimony on behalf of Twitter at today's hearing, "Algorithms and Amplification: How Social Media Platforms' Design Choices Shape Our Discourse and Our Minds."

Twitter's purpose is to serve the public conversation. While in 2006, this meant providing a platform for people to share 140-character status updates, our service has evolved to become the go-to place for people to see what's happening in the world, share opinions and observations, and engage in conversations on topics as diverse as sports, popular culture, and politics.

While technology has changed significantly since we were founded 15 years ago, our mission has not. We remain committed to giving people the power to create and share ideas and information instantly with the world.

Many of the questions we grapple with today are not new, but the rise and evolution of the online world have magnified the scale and scope of these challenges. As a global company that values free expression, we find ourselves navigating these issues amidst increasing threats to free speech from governments around the world. We strive to give people a voice while respecting applicable law and staying true to our core principles.

We use technology every day in our efforts to automatically improve outcomes and experiences for people on Twitter. We do that, in part, through algorithms. For example, our machine learning tools help identify potentially abusive or harmful content, including content that violates Twitter's Rules, to human moderators for review. In fact, we now take enforcement action on more than half of the abusive Tweets that violate our rules before they're even reported. We think this is critical, as we don't think the burden to identify and report such content should be on those who are the subject of abusive content.

As members of Congress and other policymakers debate the future of Internet regulation, they should closely consider the ways technology, algorithms, and machine learning make Twitter a safer place for the public conversation and enhance the global experience with the Internet at large.

We've invested significantly in our systems and have made strides to promote healthy conversations. However, we believe that as we look to the future, we need to ensure that all our efforts are centered on trust. Our content moderation efforts or the deployment of machine learning can be successful only if people trust us. That's why we think it is critical that we focus on being more open and decentralized. That means we must prioritize and build into our business increased transparency, consumer choice, and competition. In my testimony, I will highlight how we are innovating and experimenting in this area through (1) expanded algorithmic choice; (2) the Twitter Responsible Machine Learning initiative; (3) the Birdwatch initiative; and (4) the Bluesky project.

Expanded Algorithmic Choice

At Twitter, we want to provide a useful, relevant experience to all people using our service. With hundreds of millions of Tweets every day on the service, we have invested heavily in building systems that organize content to show individuals the most relevant information for that individual first. With over 192 million people using Twitter each day in dozens of languages and countless cultural contexts, we rely upon machine learning algorithms to help us organize content by relevance.

We believe that people should have meaningful control over key algorithms that affect their experience online. In 2018, we redesigned the home Timeline, the main feature of our service, to allow people to control whether they see a ranked timeline, or a reverse chronological order ranking of the Tweets from accounts or topics they follow. This "sparkle icon" improvement has allowed people using our service to directly experience how algorithms shape what they see and has allowed for greater transparency into the technology we use to rank Tweets. This is a good start. And, we believe this points to an exciting, market-driven approach that provides individuals greater control over the algorithms that affect their experience on our service.

Responsible Machine Learning Initiative

We are committed to gaining and sharing a deeper understanding of the practical implications of our algorithms. Earlier this month, we launched our “Responsible Machine Learning” initiative, a multi-pronged effort designed to research the impact of our machine learning decisions, promote equity, and address potential unintentional harms. Responsible use of technology includes studying the effects that the technology can have over time. Sometimes, a system designed to improve people’s online experiences could begin to behave differently than was intended in the real world. We want to make sure we are studying such developments and using them to build better products.

This initiative is industry-leading and the very first step and investment into a journey of evaluating our algorithms and working through ways we can apply those findings to make Twitter and our entire industry better. We will apply what we learn to our work going forward, and we plan to share our findings and solicit feedback from the public. While we are hopeful about the ways this may improve our service, our overarching goal is increasing transparency and contributing positively to the field of technology ethics at large.

Birdwatch

We’re exploring the power of decentralization to combat misinformation across the board through Birdwatch — a pilot program that allows people who use our service to apply crowdsourced annotations to Tweets that are possibly false or misleading. We know that when it comes to adding context, not everyone trusts tech companies — or any singular institution — to determine what context to add and when. Our hope is that Birdwatch will expand the range of voices involved in tackling misinformation as well as streamline the real-time feedback people already add to Tweets. We are working to ensure that a broad range of voices participate in the Birdwatch pilot so we can build a better product that meets the needs of diverse communities. We hope that engaging the broader community through initiatives like Birdwatch will help mitigate current deficits in trust.

We are committed to making the Birdwatch site as transparent as possible. All data contributed to Birdwatch will be publicly available and downloadable. As we develop algorithms that power Birdwatch — such as reputation and consensus systems — we intend to publish that code publicly in the Birdwatch Guide.

Bluesky

Twitter is funding Bluesky, an independent team of open source architects, engineers, and designers, to develop open and decentralized standards for social media. It is our hope that Bluesky will eventually allow Twitter and other companies to contribute to and access open recommendation algorithms that promote healthy conversation and ultimately provide individuals greater choice. These standards could support innovation, making it easier for startups to address issues like abuse and hate speech at a lower cost. We recognize that this effort is complex, unprecedented, and will take time but we currently plan to provide the necessary exploratory resources to push this project forward.

Conclusion

We appreciate the enormous privilege to host some of the most important conversations happening at any given time — from real-time updates on Supreme Court rulings to information-sharing about COVID-19 vaccine clinical trials. We are proud of the open service we have built and the steps we take each day to ensure a safe venue for diverse voices and vibrant debate. Moving forward, we believe that more open and decentralized systems will increase transparency, provide more consumer control and choice, and increase competition across our industry. Our hope is that such a system will lead to the necessary innovation to meet today's needs and solve tomorrow's challenges. Most importantly, it will build trust.

Thank you again for the opportunity to share Twitter's perspective with the Subcommittee and the public.

STATEMENT OF

JOAN DONOVAN, PHD
RESEARCH DIRECTOR AT HARVARD KENNEDY SCHOOL'S
SHORENSTEIN CENTER ON MEDIA, POLITICS AND PUBLIC POLICY

HEARING ON “ALGORITHMS AND AMPLIFICATION: HOW SOCIAL
MEDIA PLATFORMS’ DESIGN CHOICES SHAPE OUR DISCOURSE AND
OUR MINDS”

BEFORE THE SENATE COMMITTEE ON THE JUDICIARY
SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW

APRIL 27, 2021

From “Get Big Fast” to “Move Fast and Break things” and Back Again.

Before there was “move fast and break things,” there was another animating ethic of the tech industry: “Get big fast!” This philosophy has proven to be good for the industry, but bad for the world. Over the last decade, social networking (connecting people to people) morphed into social media (connecting people to people *and* to content), which resulted in exponential profits and growth. Most people don’t know the difference between social networking and social media, but this transition was the key to products like Facebook, Twitter, and YouTube dominating global markets in mass communication. In short, *networks are the wealth of society*. Networks are where the rich and powerful derive their importance and high status, hence saying “he or she is connected” when referencing someone you do not want to mess with. When social media is the vector of attack against our democracy and public health, a small group of highly motivated and connected actors can manipulate public understanding of any issue simply by using these products they are as designed.

How social media companies got big fast was a combination of lax consumer regulation, eschewing risks, buying out the competition where possible, and a focus on scale that made for poor security decisions. Beyond connecting people and content, products like Facebook, YouTube, and Twitter rely on other companies and individuals to provide them with more data, increasing the scale in this massive and sprawling data infrastructure across the web. Mapping, tracking, and aggregating people’s social networks made social media a viable business because companies could sell data derived from interactions or monetize those relationships as other products, such as advertising, targeted posts, and promoted messages. Social media data should be legally defined at some point, but for now, I am referring to information about people, how they behave online, interactions with people and content, and location tagging.

But, it wasn't enough just to collect and sort data on the product: targeted advertising and data services only become useful when paired with other kinds of data. For example, in Nov. 2012, when looking at different models for monetizing Facebook, Zuckerberg wrote in a company email that allowing developers access to data without having these companies share their data with Facebook would be "good for the world, but bad for us."¹ This is because Facebook knew, even back then, that their products could threaten privacy on a scale society had never reckoned with before. Now, these social media products that favor runaway scale and openness threaten not only individual rights, but also the future of democracy and public health.

By leveraging people's networks and content at the same time, a business model emerged where key performance indicators included:

- (1) growth of daily and monthly active users,
- (2) increasing engagement metrics, and
- (3) advertising revenue.

The last decade has been marked by these companies expanding exponentially on all of these indicators. In a *PC Mag* article from 2011 about the best mobile apps, Facebook and Twitter were both ranked lower than an app that turns your camera into a flashlight.² In 2011, Twitter had approximately 100 million users, Facebook had 845 million, and YouTube had 800k.³ By 2020, Twitter reports 353 million active users, Facebook reports 2.7 billion active users, and 2.29 billion for YouTube.⁴ Advertising revenue continues to grow across all of these products, where Google (\$146 billion) and Facebook (\$84 billion) dominate.⁵

Using accounts as a key performance indicator drove a shadow industry of growth hacking, which eventually was integrated directly into the products—allowing a

¹ Solon, Olivia, and Cyrus Farivar. 2019. "Thousands of Leaked Facebook Documents Show Mark Zuckerberg as 'Master of Leverage' in Plan to Trade User Data." NBC News. <https://www.nbcnews.com/tech/social-media/mark-zuckerberg-leveraged-facebook-user-data-fight-rivals-help-friends-n994706>

² Segan, Sasha. 2011. "The Top 10 Free BlackBerry Apps 2011." PC Mag. <https://www.pcmag.com/news/the-top-10-free-blackberry-apps-2011>.

³ Wasserman, Todd. 2012. "Twitter Says It Has 140 Million Users." Mashable. <https://mashable.com/2012/03/21/twitter-has-140-million-users/>.

Associated Press. 2012. "Number of Active Users at Facebook over the Years." <https://finance.yahoo.com/news/number-active-users-facebook-over-years-214600186--finance.html?guccounter=1>.

⁴ Statista. 2021. "Most Used Social Media 2021." Statista. Accessed April 26, 2021. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.

⁵ H. Tankovska, "Facebook's advertising revenue worldwide from 2009 to 2020." <https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/#:~:text=In%202020%2C%20Facebook%20generated%20close,of%20the%20social%20network's%20revenue>.

Joseph Johnson, "Advertising revenue of Google from 2001 to 2020," February 5, 2021, Statista, <https://www.statista.com/statistics/266249/advertising-revenue-of-google/#:~:text=In%202020%2C%20Google's%20ad%20revenue%20amounted%20to%20146.92%20billion%20US%20dollars>.

massive and known vulnerability of sock puppets, or fake accounts, to persist.⁶ For those who understood how to manipulate this vulnerability, increasing engagement meant delivering more novel and outrageous content, which is why false news, harassment, and defamation thrive on social media. For social media companies, decisions about profit drive innovation, not higher principles like access to truth, justice, or democracy. As a result, these products are not only a parasite on our social networks feeding off every click, like, and share, but they also cannot optimize for the public interest. It did not have to be this way. Back in 2011, mobile was developing quickly and there were many ways in which social media could have been designed to foster community safety and to maximize privacy. Instead, the drive to maximize the number of users, engagement, and revenue led us here.

Most crucially the entire internet infrastructure needs an overhaul, so that companies are not able to siphon data and leverage it to maximize an advantage over consumers.⁷ But, users are not necessarily the customers, advertisers are.⁸ The structure of online advertising pipelines systematically advantages these companies at the expense of several industries, most importantly journalism. By becoming the gateway to news audiences, top social media companies hoard advertising revenue that belongs to those who create engaging content for display on their products, most notably journalists.

When criticized about the squeeze their products have placed on journalism, Facebook and Google will cite their various news initiatives. But, these initiatives pick and choose partners and then channel journalists labor directly back into their products. Facebook's fact-checking program, for instance, partners with several reputable news outlets, but labelling has done little to disincentivize fake news. Moreover, fact-checking is ad-hoc and will never rival supporting independent investigative journalism, a bedrock of a strong democracy. Instead, this initiative expands Facebook's ever-growing web of influence over news as it becomes increasingly more difficult to criticize the corporation for fearing of losing resources.

Nevertheless, as journalism wanes, social media serves **misinformation-at-scale** to hundreds of millions of daily active users instantaneously, especially odious when misinformation is promoted in trends and recommendations. In October 2020, I testified about conspiracies and misinformation having similar harmful societal impacts as secondhand smoke.⁹ Post-2020, we see misinformation-at-scale's deadly effects in the US. Scammers and grifters use social media to sell bogus products and push conspiracies—including monetizing the pandemic in grotesque ways to sell fake cures or to scaremonger.¹⁰ Going into the pandemic, anti-vaccination activists had a huge

⁶ Donovan, Joan, and Brian Friedberg. n.d. "Source Hacking: Media Manipulation in Practice." Data & Society (blog). Accessed January 5, 2020. <https://datasociety.net/output/source-hacking-media-manipulation-in-practice/>.

⁷ Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

⁸ Anthony Nadler, and Joan Donovan. 2018. "Weaponizing the Digital Influence Machine." Data & Society Research Institute. October 17, 2018. <https://datasociety.net/library/weaponizing-the-digital-influence-machine/>.

⁹ Donovan, Joan. 2020. "Thank You for Posting: Smoking's Lessons for Regulating Social Media | MIT Technology Review." <https://www.technologyreview.com/2020/10/05/1009231/social-media-facebook-tobacco-secondhand-smoke/>.

¹⁰ Donovan, Joan. 2021. "Without Leadership on Vaccine Rollout, Scams Are Inevitable | MIT Technology Review." <https://www.technologyreview.com/2021/01/06/1015813/florida-eventbrite-vaccine-scam-inevitable/>.

advantage over public health officials, where anti-vaccination activists were able to leverage already dense and sprawling networks across social media products. As a result, they attached their strikes to breaking news cycles by attacking public confidence in science. There was nothing public health officials could do to stop the torrent of misinformation drowning doctors and hospitals, as evidenced by the reporting of Brandy Zadrozny and Ben Collins at *NBC News*.¹¹ The same situation holds for other public servants, like election officials, who continue to bear the costs of election disinformation and are leaving their jobs because its managing misinformation-at-scale is unsustainable.¹²

For journalists, researchers, and everyone trying to mitigate misinformation, the experience is like trying to put your hands up against a growing ocean swell as it washes over you. Journalists, universities, public service, and our healthcare professionals take on the true costs of misinformation-at-scale, which isn't an existential statement. There are millions of resources lost to mitigating misinformation-at-scale, where the cost of doing nothing is even worse. For example, take the blatant lie that the vaccines have microchips. To counter it, journalists traded off covering other stories, while public health professionals continue to explain that there are no microchips in the vaccine.¹³

The only way to fix a problem like motivated misinformers involves platforms enforcing existing policies, researchers and journalists working together as tech watchdogs, and policymakers opening the way for a public interest internet. Regulators should introduce public interest obligations to social media newsfeeds and timelines so that timely, local, relevant, and accurate information reaches the masses-at-scale. Together, we must make a public interest internet a whole-of-society priority.

Going Down the Rabbit Hole

What I have learned studying the internet over the last decade is simple: *everything open will be exploited*. There is nothing particularly new about misinformation and conspiracies circulating. After all, there is no communication without misinformation. However, over the last decade the design of social media itself created favorable conditions for reaching millions instantaneously while also incorporating financial and political incentives for conducting massive media manipulation campaigns. The most dangerous aspects of these products come to light when we analyze who gains an advantage when openness meets scale.

¹¹ Zadrozny, Brandy. 2020. "They're on the Misinformation Front Line. Here's What They Predict about the Virus." NBC News. <https://www.nbcnews.com/tech/social-media/these-disinformation-researchers-saw-coronavirus-infodemic-coming-n1206911>.

Zadrozny, Brandy, and Ben Collins et al. 2020. "Whitmer Conspiracy Allegations Tied to 'boogaloo' Movement." NBC News. <https://www.nbcnews.com/tech/tech-news/whitmer-conspiracy-allegations-tied-boogaloo-movement-n1242670>.

Zadrozny, Brandy. 2020. "Anti-Vaccination Movement Gets Traction in Unlikely Source: Local News." NBC News. <https://www.nbcnews.com/tech/tech-news/anti-vaccination-groups-target-local-media-after-social-media-crackdowns-n1251485>.

¹² CNN, Fredreka Schouten and Kelly Mena. n.d. "High-Profile Elections Officials Leave Posts after a Tumultuous 2020." CNN. Accessed April 26, 2021. <https://www.cnn.com/2021/02/19/politics/election-officials-lose-and-leave-jobs/index.html>.

¹³ Donovan, Joan, Brian Friedberg, Gabrielle Lim, Nicole Leaver, Jennifer Nilsen, and Emily Dreyfuss. 2021. "Mitigating Medical Misinformation: A Whole-of-Society Approach to Countering Spam, Scams, and Hoaxes," 52.

I often joke nervously that “my computer thinks I’m a white supremacist.” One only needs to look at my homepage on YouTube to illustrate this point. On the homepage, YouTube clearly displays your interests and makes recommendations. Daily, it recommends me content from a white supremacist who they have already banned, yet recent videos of his livestreams are continuously recommended. I first learned of the pandemic in January 2020 from a conspiracist and nationalist YouTuber, who was excited by shutting down the borders to stop the “Wu Flu.” I had spent countless hours down the rabbit hole with this YouTuber before, who that night in January 2020 spent over three hours extoling his xenophobic views. Racialized disinformation continues to be a critical source of political partisanship in the US because it is so easy to manipulate engagement on race and racism—and it’s profitable.¹⁴

While some debate the existence of “the rabbit hole” on social media, our research team at Shorenstein has been looking deeper at this phenomenon. Going down the rabbit hole means getting pulled into an online community or subculture, where the slang, values, norms, and practices are unfamiliar, but nevertheless engrossing. There are four aspects of the design of social media that lead someone down the rabbit hole. They are:

- (1) repetition relates to seeing the same thing over and over on a single product,
- (2) redundancy is seeing the same thing across different products,
- (3) responsiveness is how social media and search engines always provide some answer unlike other forms of media, and
- (4) reinforcement is the ways that algorithms work to connect people and content so that once you’ve searched for a slogan or keyword, algorithms will reinforce these interests.

Nowhere is this more prevalent than on YouTube, where any search for conspiracy or white supremacist content, using their preferred keywords of the in-group, will surface numerous recommendations. It’s a misconception that these online echo chambers or filter bubbles are hyper-personalized and conclusively shape individual behavior in a specific direction. Instead, what algorithms tend to do is group people with homogeneous characteristics into buckets, who are served similar content in batches. From 9/11 conspiracies, to the “vaccines cause autism” meme, to QAnon, some conspiracist communities have been thriving on social media for decades. But, it is a misnomer, albeit a popular one, to imagine social media as an attention economy, where individual users are making independent choices of where to spend their time.

It’s more correct to call the rabbit hole an “algorithmic economy,” where algorithms pattern the distribution of content based on signals from millions of people according to generic profiles in buckets, coupled with algorithmic grouping in batches. On its surface, the design is not insidious: the buckets and batches are related to generic interests. For example, if you’re a baseball fan and YouTube knows you want more

¹⁴ Freelon, Deen, Michael Bossetta, Chris Wells, Josephine Lukito, Yiping Xia, and Kirsten Adams. 2020. “Black Trolls Matter: Racial and Ideological Asymmetries in Social Media Disinformation.” *Social Science Computer Review*, April, 0894439320914853. <https://doi.org/10.1177/0894439320914853>.

sports content, that's a great service. But if you've searched for more contentious content, like QAnon, Proud Boys, or Antifa recently, you are likely to enter a rabbit hole, where extracting yourself from reinforcement algorithms ranges from difficult to impossible. While customers, such as advertisers, have lobbied these social media companies for better ad placement, users are not able to easily swap out interests or stop targeted recommendations altogether.

Getting Out of the Rabbit Hole

My last point is about the past five years of social media shaping our public discourse. Social media provides a different opportunity for the enemies of democracy to sow chaos and plan violent attacks. It's fourth generation warfare, where it is difficult to tell the difference between citizens and combatants. The reason why Russia impersonated US social movements in 2016 was expressly because movements elicit lots of engagement, where participants see sharing content and network-making as political acts. That kind of political participation was challenging for city governance during the 2011 Occupy Movement, but that moment—a decade ago—should have taught Facebook, YouTube, and Twitter more about the range of effects their products could have on society. Now we see these products used by authoritarians who leverage a mix of authentic political participation paired with false accounts and fake engagement to win elections.¹⁵

Cobbled together across products, our new media ecosystem is the networked terrain for a hybrid information war that ultimately enables dangerous groups to organize violent events—like the nationalists, militias, white supremacists, conspiracists, anti-vaccination groups, and others who collaborated under the banner of Stop The Steal in order to breach the Capitol. Last week, a *Buzzfeed* article included a leaked internal Facebook memo on the exponential growth of “Stop the Steal” groups on their platform. The report clearly illustrated that groups exposing violent and hateful content can grow very fast on across the product. Even when Facebook removes groups, it does not stop the individuals running them from trying again. Adaption by media manipulators is a core focus of our research at the Shorenstein Center.¹⁶ Facebook found that their own tools allowed Stop the Steal organizers to leverage openness and scale to grow faster than Facebook's own internal teams could counter.

In short, even when aware of the risks of their product to democracy, Facebook's interventions do little to contain exposure of misinformation-at-scale to the general public. When determined to stop the spread of misinformation, Facebook could not counter it with their internal policies. Misinformation-at-scale is a feature of Facebook's own design and is not easily rooted out. Because Facebook defines the problem of misinformation-at-scale as one of coordinated inauthentic behavior, they were woefully unprepared handle the threats posed by their own products. They were unprepared in

¹⁵ Ong, Jonathan, and Jason Vincent Cabañes. 2018. “Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines.” *Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines*, January. <https://doi.org/10.7275/2cq4-5396>.

¹⁶ Donovan, Joan, Emily Dreyfuss, Brian Friedberg, and Gabrielle Lim. n.d. “A Blueprint for Documenting and Debunking Misinformation Campaigns.” *Nieman Reports*. Accessed November 8, 2020. <https://niemanreports.org/articles/a-blueprint-for-documenting-and-debunking-misinformation-campaigns/>.

2016 and have since then been unable to handle the new ways that motivated misinformers use their products.

What began in 2016 with false accounts and fake engagement inflaming and amplifying societal wedge issues slowly transformed overtime into a coordinated attack on US democracy and public health. The biggest problem facing our nation is misinformation-at-scale, where technology companies must put community safety and privacy at the core of their business model, ensure that advertising technology is utilized responsibly, and quickly act on groups coordinating disinformation, hate, harassment, and incitement across the media ecosystem. A problem this big will require Federal oversight.

But I am hopeful that another future is possible, if tech companies, regulators, researchers, and advocacy begin to work together to build a public interest internet modeled on the principles that the public has a right to access accurate information on demand.¹⁷ The cost of doing nothing is democracy's end.

¹⁷ Technology and Social Change Research Project. 2021. "Submission to the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression." Media Manipulation Casebook. February 17, 2021. <https://mediamanipulation.org/research/submission-un-special-rapporteur-promotion-and-protection-right-freedom-opinion-and>.



United States Senate Committee on the Judiciary
Subcommittee on Privacy, Technology, and the Law
*Algorithms and Amplification: How Social Media Platforms' Design Choices
Shape Our Discourse and Our Minds*
April 27, 2021

Written Statement of Tristan Harris, President and Co-Founder
Center for Humane Technology
[Humanetech.com](https://www.humanetech.com)

Thank you Senator Coons and Senator Sasse.

I was featured in the Netflix documentary, *The Social Dilemma*, which has now been seen by more than an estimated 100 million people in 190 countries and 30 languages. The film burst into the public conversation because it confirmed what so many people knew and felt already: that the business model behind the social media platforms that have rewired human civilization with addiction, mental health problems, alienation, extremism, polarization, and breakdown of truth. Now the world wants to see real change.

Why did the film take off? Not because it spoke to a few nuisances of technology, but because insiders who were involved spoke clearly about why tech's deranging influence was existential for democracy. Quoting the film: "*If we can't agree on what's true, then we can't navigate out of any of our problems.*"

In the Cold War, the United States invested heavily in *continuity of government*. Faced with the threat of a nuclear attack, we spent millions on underground bases and emergency plans to ensure a continuity of U.S. government decision-making to maintain our capacity to respond to adversaries. But today, an invisible disruption to the continuity of U.S. government has already happened underneath our noses. Not by nuclear missile or by sea, but through the slow, diffuse process by which social media made money from pitting our own citizens and Congressional representatives into an online Hobbesian war of "all against all"—making agreement or good faith impossible, shattering our shared reality and effectively disabling our societal O.O.D.A. loops (Observe, orient, decide and act). The gears have jammed.

Meanwhile, we face genuine existential threats that require urgent attention: from the rise of China, to a climate crisis, nuclear proliferation, vulnerable infrastructure, to dangerous inequality. Today's tech platforms disable our capacity to address these urgent problems.

That is why we must reset our criteria for success. Instead of evaluating whether my fellow Facebook, Twitter and YouTube panelists have improved their *content policies* or *hired more content moderators*, we should ask what would collectively constitute a “humane” Western digital democratic infrastructure that would strengthen our capacity to meet these threats. Instead of shortening attention spans, distracting us, competing for addiction and outrage... they would compete from the bottom-up to deepen and cultivate our best traits, sustained thinking and concentration, better critical thinking, facilitating easier ways to understand each other and identify solutions built on common ground. We should be interested in structural reforms for tech platforms’ incentives that would *comprehensively strengthen* rather than *disable* our capacity to respond to these existential threats, especially in competition with China.

My fellow panelists from technology companies will say:

- We catch XX% more hate speech, self-harm and harmful content using A.I.
- We took down XX billions of fake accounts, up from YY% last year.
- We have Content Oversight Boards and Trust & Safety Councils.
- We spend \$X million more on Trust & Safety in 2021 than we made in revenue in an entire year.

But none of this is adequate to the challenge stated above, when the entire model is predicated on dividing society. It’s like Exxon talking about the number of trees they have planted, while their extractive business model hasn’t changed.

As *The Social Dilemma* explains, the problem is their attention-harvesting business model. The narrower and more personalized our feeds, the fatter their bank accounts, and the more degraded the capacity of the American brain. The more money they make, the less capacity America has to define itself as America, reversing the United States inspiring and unifying motto of *E Pluribus Unum* or “out of many, one” into its opposite, “out of one, many.”

We are raising entire generations of young people who will have come up under these exaggerated prejudices, division, mental health problems, and an inability to determine what’s true. They walk around as a bag of cues and triggers that can be ignited. If this continues, we will see more shootings, more destabilization, more children with ADHD, more suicides and depression— deficits that are cultivated and exploited by these platforms.

We should aim for nothing less than a comprehensive shift to a humane, clean “Western digital infrastructure” worth wanting. We are collectively in the middle of a major transition from 20th century analog societies to 21st century “digitized” societies. Today we are offered two dystopian choices: either to install a Chinese “Orwellian” brain implant into society with authoritarian controls, censorship and mass behavior modification. Or we can install the U.S./Western “Huxleyan” societal brain implant that saturates us in distractions, outrage, trivia and amusing ourselves to death.

Let’s use today’s hearing to encourage a 3rd way, to have the government’s help in incentivizing Digital Open Societies worth wanting, that outcompete Digital Closed Societies.

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

**Written Testimony of Alexandra N. Veitch
Director, YouTube Government Affairs & Public Policy, Americas**

**Senate Judiciary Committee
Subcommittee on Privacy, Technology & the Law
“Algorithms and Amplification: How Social Media Platforms’ Design Choices Shape Our
Discourse and Our Minds”
April 27, 2021**

Introduction

Chairman Coons, Ranking Member Sasse, and distinguished members of the subcommittee:

Thank you for the opportunity to appear before you today. My name is Alexandra Veitch, and I am the Director of Public Policy for the Americas at YouTube. As part of my role, I lead a team that advises the company on public policy issues around online, user-generated content.

At YouTube, we believe that the Internet has been a force for creativity, learning, and access to information. Supporting this free flow of ideas is at the heart of our mission to give everyone a voice and show them the world. We have built and continue to improve YouTube to empower users to access, create, and share information like never before; this has enabled billions to benefit from a bigger, broader understanding of the world. In addition, our platform has created economic opportunities for small businesses across the country and around the world, and we have provided artists, creators, and journalists a platform to share their work. Over the last three years, we’ve paid more than \$30 billion to creators, artists, and media companies around the world¹. And according to an Oxford Economics report², YouTube’s creative ecosystem supported the equivalent of 345,000 full time jobs in 2019 in the United States.

¹ <https://blog.youtube/inside-youtube/letter-from-susan-our-2021-priorities/>

²

<https://blog.oxfordeconomics.com/content/from-opportunity-to-impact-assessing-the-economic-societal-and-cultural-benefits-of-youtube-in-the-us>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

Over the years, we have seen more and more people come to YouTube to share their experiences and understand their world more deeply. This is especially true when it comes to learning new skills, participating in civic engagement, and developing informed opinions about current events. With so many users around the world looking to YouTube for information, we have a responsibility to provide a quality experience and support an informed citizenry. Over the past several years, responsibility has been a critical area of investment across our company. We have focused extensively on developing policies and building product solutions to live up to this responsibility while preserving the opportunity of an open platform. And we work continuously to identify areas where we can do more. In my testimony today, I will (1) explain how we think about algorithms, (2) discuss our approach to responsibility and how technology supports this work, (3) highlight our efforts to provide more transparency and visibility into how YouTube works, and (4) illustrate how our products protect users.

How YouTube thinks about algorithms

YouTube is a multi-faceted video-sharing platform enjoyed by billions of consumers and creators. The popularity of our service unlocks business opportunities for creators and helps businesses grow their reach. With this volume of users and economic engine for creators comes a significant responsibility to protect those users. With more than 500 hours of video uploaded to YouTube per minute, enabling users to easily find content they are looking for and will enjoy while protecting them from harmful content simply would not be possible without the help of technology.

Because of the importance of algorithms in the YouTube user experience, we welcome the opportunity to clarify our approach to this topic. In computer science terms, an algorithm is a set of instructions that direct a computer to carry out a specific task. An algorithm can be simple—asking a computer to calculate the sum of two numbers—or extremely complex, such as machine learning algorithms that consistently refine their ability to accomplish the goal for which they were programmed. An algorithm can manage a few inputs or nearly limitless inputs, and they can do one thing or perform a number of functions at once. Nearly everything that people do today on their devices is made possible by algorithms.

YouTube uses machine learning techniques to manage and moderate content on YouTube. YouTube's machine learning systems sort through the massive volume of content to find the most relevant and useful results for a user's search query, to identify opportunities to elevate authoritative news, and to provide a user with additional context via an information panel if appropriate. We also rely on machine learning technology to help identify patterns in content that may violate our Community Guidelines or videos that may contain borderline

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

content—content that comes close to violating our Community Guideline but doesn't quite cross the line. These systems scan content on our platform 24/7, enabling us to review hundreds of thousands of hours of video in a fraction of the time it would take a person to do the same. For example, more than 94% of the content we removed between October and December of 2020 was first flagged by our technology. This underscores just how critical machine learning is for content moderation.

Another area where we use machine learning is for recommendations. Recommendations on YouTube help users discover videos they may enjoy, and they help creator content reach new viewers and grow their audience across the platform. We share recommendations on YouTube's homepage and in the "Up next" section to suggest videos a user may want to watch after they finish their current video. Our recommendation systems take into account many signals, including a user's YouTube watch and search history (subject to a user's privacy settings) and channels to which a user has subscribed. We also consider a user's context—such as country and time of day—which, for example, helps our systems show locally relevant news, consistent with our effort to raise authoritative voices. Our systems also take into account engagement signals about the video itself—for example, whether others who clicked on the same video watched it to completion or clicked away shortly after starting to view the video. It is important to note that, where applicable, these signals are overruled by the other signals relating to our efforts to raise up content from authoritative sources and reduce recommendations of borderline content and harmful misinformation—even if it decreases engagement.

We also empower our users by giving them significant control over personalized recommendations, both in terms of individual videos as well as the way that watch and search history may inform recommendations. Users control what data is used to personalize recommendations by deleting or pausing activity history controls. Signed out users can pause and clear their watch history, while signed in users can also view, pause, and edit watch history at any time through the YouTube history settings. Clearing watch history means that a user will not be recommended videos based on content they previously viewed. Users can also clear their search history, remove individual search entries from search suggestions, or pause search history using the YouTube History settings.

In-product controls enable users to remove recommended content—including videos and channels—from their Home pages and Watch Next. Signed in users can also delete YouTube search and watch history through the Google My Account settings, set parameters to automatically delete activity data in specified time intervals, and stop saving activity data

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

entirely. We also ask users directly about their experiences with videos using surveys that appear on the YouTube homepage and elsewhere throughout the app, and we use this direct feedback to fine-tune and improve our systems for all users.

YouTube's Pillars of Responsibility: the 4Rs

Responsibility is our number one priority at YouTube. Some speculate that we hesitate to address problematic content because it benefits our business; this is simply false. Failure to consistently take sufficient action to address harmful content not only threatens the safety of our users and creators, it also threatens the safety of our advertising partners' brands. Our business depends on the trust of our users and our advertisers. This is why we have made significant investments over the past few years in teams and systems that protect YouTube's users, partners, and business. Our approach towards responsibility involves 4 "Rs" of responsibility, described in detail below.

REMOVE VIOLATIVE CONTENT: Our Community Guidelines provide clear, public-facing, guidance on content that is not allowed on the platform. These include policies against spam, deceptive practices, scams, hate, harassment, and identity misrepresentation and impersonation. We remove content that violates our policies as quickly as possible, and removed videos represent a fraction of a percent of total views on YouTube. We work continuously to shrink this even further through improved detection and enforcement, relying on a combination of technology and people.

We are dedicated to providing access to information and freedom of expression, but YouTube has always had clear and robust content policies. For example, we have never allowed pornography, incitement to violence, or content that would harm children. Harmful content on our platform makes YouTube less open, not more, by creating a space where creators and users may not feel safe to share. That's why our policy development team systematically reviews and updates all of our policies to ensure that they are current, keep our community safe, and preserve openness. They frequently consult outside experts and YouTube creators during the process, and consider regional differences to ensure proposed changes can be applied fairly and consistently around the world.

Our COVID-19 Medical Misinformation Policy³ represents one such example of YouTube working closely with experts. Over the course of the last year, we have worked with and relied on information from health authorities from around the world to develop a robust policy

³ <https://support.google.com/youtube/answer/9891785>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

anchored in verifiably false claims tied to real world harm. Our policy addresses false and harmful claims about certain treatments and public health measures, as well as misinformation about COVID-19 vaccines. Each claim we prohibit has been vetted as verifiably false by the consensus of global health authorities, including the CDC. This policy has evolved alongside misinformation trends about the pandemic, and we have invested significant time and resources into carefully developing the policy and creating the necessary training and tools required to enforce it.

We adapted this policy over time to address the challenges of this pandemic. We began to remove content for COVID-19 misinformation in March 2020, under provisions of our policy prohibiting Harmful and Dangerous content. But as the pandemic progressed, we developed a fulsome and separate COVID-19 misinformation policy. In October 2020, we further expanded the policy to include vaccine misinformation about COVID vaccines. Since March 2020, we have vigorously enforced our COVID-19 misinformation policy to protect our users, removing 900,000 videos worldwide. And in the fourth quarter of 2020, we removed more than 30,000 videos for violating the vaccine provisions of our COVID-19 misinformation policy.

Once we have implemented a policy, we rely on a combination of people and technology to enforce it. Machine learning plays a critical role in content moderation on YouTube, and we deploy it in two key ways: proactively identify and flag harmful content, and automatically remove content that is very similar to what has been previously removed. In both cases, data inputs are used to train the systems to identify patterns in content—both the rich media content in videos, as well as textual content like metadata and comments—and our systems then use those patterns to make predictions about new examples to match. Machine learning is well-suited to detect patterns, which also helps us to find content similar to other content we have already removed, even before it is ever viewed. We sometimes use hashes (or “digital fingerprints”) to catch copies of known violative content before they are even made available to view. The systems then automatically remove content only where there is high confidence of policy violation—e.g., spam—and flag the rest for human review.

Machine learning is critical to keeping our users safe. In the fourth quarter of 2020, of the 9.3 million videos removed for violating our Community Guidelines, 94% of those videos were first flagged by machine detection. But the human review piece here is critical as well: machines are effective for scale and volume but are not able to analyze and evaluate context, whereas human reviewers allow us to evaluate context and consider nuance when enforcing our policies. Once our machine learning systems flag a potentially violative video, reviewers then remove videos that are violative while non-violative videos remain live. These decisions are in

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

turn used as inputs to improve the accuracy of our technology so that we are constantly updating and improving the system's ability to identify potentially violative content. In addition, when we introduce a new policy or alter an existing one, it takes our systems a bit of time to catch up and begin to detect relevant content. As we explained when we updated our hate speech policy, our enforcement of new policies improves quarter over quarter.

But as with any system, particularly operating at scale like we do, we sometimes make mistakes, which is why creators can appeal removal decisions. Creators are notified when their video is removed, and we provide a link with simple steps to appeal the decision. If a creator chooses to submit an appeal, it goes to human review, and the decision is either upheld or reversed. And we are transparent about our appeals process. As reported in our most recent Transparency Report, in Q4 2020, creators appealed a total of just over 223,000 videos. Of those, more than 83,000 were reinstated.

We also recently added a new metric to the YouTube Community Guidelines Enforcement report⁴ known as Violative View Rate (VVR). This metric is an estimate of the proportion of video views that violate our Community Guidelines in a given quarter (excluding spam). Our data science teams have spent more than two years refining this metric, which we consider to be our North Star in measuring the effectiveness of our efforts to fight and reduce abuse on YouTube. In Q4 of 2020, YouTube's VVR was 0.16-0.18%, meaning that out of every 10,000 views on YouTube, 16-18 come from violative content. We have added historical data for this metric to our Transparency Report, showing that, since Q4 of 2017, we have seen a 70% drop in VVR. This reduction is due in large part to our investments in machine learning to identify potentially violative content. Going forward, we will update this metric quarterly alongside our regular data updates.

RAISE UP AUTHORITATIVE VOICES: YouTube is a source for news and information for people around the world—whether about events unfolding in local communities or more existential global issues like climate change. Not all queries are the same. For topics like music or entertainment, relevance, newness, and popularity are most helpful to understand what people are interested in. But for subjects such as news, science, and historical events, where accuracy and authoritativeness are key, the quality of information and context are paramount.

Our search and recommendations systems are designed to raise up authoritative voices in response to user queries, especially those that are “news-y” or related to topics prone to

⁴ <https://transparencyreport.google.com/youtube-policy>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

misinformation. In 2017, we started to prioritize authoritative voices such as local and national news sources for information queries in search results and “watch next” recommendation panels. This work continued with the addition of a short preview of text-based news articles in search results on YouTube, along with a reminder that breaking and developing news can rapidly change. And in 2018, we introduced Top News and Breaking News sections to highlight quality journalism⁵. Our work here is far from done, but we have seen significant progress in our efforts to raise authoritative voices on YouTube. Globally, authoritative news watchtime grew by more than 85% from the first half of 2019 to the first half of 2020, with a 75% increase in watchtime of news in the first 3 months of 2020 alone. And during the 2020 U.S. elections, the most popular videos about the election came from authoritative news organizations. On average 88% of the videos in top 10 search results related to elections came from authoritative news sources.

Authoritativeness is also important for topics prone to misinformation, such as videos about COVID vaccines. In these cases, we aim to surface videos from authoritative news publishers, public health authorities, and medical experts. Millions of search queries are getting this treatment today and we continue to expand to more topics and countries. In addition, in April 2020, we expanded our fact-checking panels in YouTube search results to the U.S., providing fresh context in situations where a news cycle has faded but where unfounded claims and uncertainty about facts are common. These panels highlight relevant, third-party fact-checked articles above search results for relevant queries, so that our viewers can make their own informed decision about claims made in the news.

We also recognize that there may be occasions when it is helpful to provide viewers with additional context about the content they are watching. To that end, we have a variety of information panels that provide context on content relating to topics and news prone to misinformation, as well as the publishers themselves. For example, a user viewing a video about climate change—regardless of the point of view presented in the content—will see an information panel providing more information about climate change with a link to the relevant Wikipedia article.

Information panels provide critical context as well as point users to reliable sources of authoritative information. For example, when a user in the U.S. watches a video about COVID-19, we display an information panel that points to the CDC’s official resource for information about COVID-19 and the Google search results page with health information from

⁵ <https://blog.youtube/news-and-events/building-better-news-experience-on>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

the CDC and local statistics and guidance. Beginning last week, when a U.S. user watches a video about COVID-19 vaccines, we show a panel that points to the CDC's online resource for vaccine information, with an additional link to the Google search results page with local information about vaccination. To date, our COVID-19 information panels have received more than 400 billion views.

For the U.S. 2020 election, we provided a range of new information panels in addition to our existing panels to provide additional context around election-related search results and video content. For example, when viewers searched for specific queries related to voter registration on YouTube, they were shown an information panel at the top of the page that linked to Google's "How to register to vote" feature for their state. When a viewer searched for 2020 presidential or federal candidates on YouTube, we surfaced an information panel with information about that candidate—including party affiliation, office, and when available, the official YouTube channel of the candidate—above search results. We also provided an election results info panel, the content of which was flexible in order to keep pace with new developments and key milestones along the road to inauguration. These panels were collectively shown more than 8 billion times.⁶

REDUCE THE SPREAD OF BORDERLINE CONTENT:

While we have strong and comprehensive policies in place that set the rules for what we don't allow on YouTube, we also recognize that there's content that may be problematic but doesn't violate our policies. Content that comes close to violating our Community Guidelines but does not cross the line—what we call "borderline content"—is just a fraction of 1 percent of what is watched on YouTube in the United States. We use machine learning to reduce the recommendations of this type of content, including potentially harmful misinformation.

In January 2019, we announced changes to our recommendations systems to limit the spread of this type of content. These changes resulted in a 70 percent drop in watchtime on non-subscribed recommended content in the U.S. that year. We saw a drop in watchtime of borderline content coming from recommendations in other markets as well. While algorithmic changes take time to ramp up and consumption of borderline content can go up and down, our goal is to have views of non-subscribed, recommended borderline content below 0.5%. We seek to drive this number to zero, but no system is perfect; in fact, measures intended to take this number lower can have unintended, negative consequences, leading legitimate speech to

⁶

<https://kstatic.googleusercontent.com/files/2dfe4c0bde90ab01c566d50ec70c8d74bc60f34ac5a409c28c5d5b1bac8aae93ee1187562a0e12f672d32df8d5cd8fa8fe52c95ee96293bc92b0ca7f9301d0b3>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

not be recommended. As such, our goal is to stay below the 0.5% threshold, and we strive to continually improve over time.

This content is ever-evolving and it is challenging to determine what content may fall into this category. This is why we rely on external evaluators located around the world to provide critical input on the quality of a video, and these evaluators are trained with public guidelines⁷. Each evaluated video receives up to nine different opinions and some critical areas require certified experts. For example, medical doctors provide guidance on the validity of videos about specific medical treatments to limit the spread of medical misinformation. Based on the consensus input from the evaluators, we use well-tested machine learning systems to build models, which in turn review hundreds of thousands of hours of videos each day to find and limit the spread of borderline content. These models continue to improve in order to more effectively identify and reduce recommendations of borderline content.

Our efforts here have been publicly validated in several ways. Researchers in the United States and around the world who have studied YouTube have acknowledged that YouTube's recommendation algorithm actively discourages viewers from visiting radicalizing or extremist content. According to a 2020 study⁸ conducted by an Australian data scientist and a researcher at the University of California, Berkeley's School of Information: "...data suggest that YouTube's recommendation algorithm actively discourages viewers from visiting radicalizing or extremist content. Our study thus suggests that YouTube's recommendation algorithm fails to promote inflammatory or radicalized content, as previously claimed by several outlets." In 2018, a study⁹ from Pew Research also found that, on average, our recommendations point users to popular videos, which videos tend to be gaming, vloggers, and music rather than conspiracy theories or other types of misinformation—which again account only for a fraction of the content on YouTube.

REWARD TRUSTED CREATORS: In our mission to create and sustain an open, global platform, YouTube has also expanded economic opportunity for small businesses, artists, creators, journalists, rightsholders, and more. For many, sharing video content on YouTube is

⁷ <https://support.google.com/youtube/answer/9230586>

⁸ <https://journals.uic.edu/ojs/index.php/fm/article/view/10419/9404>

⁹

<https://www.pewresearch.org/internet/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lesson/>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

not just a hobby, but a business. Globally, the number of creators earning five figures annually increased more than 40% from December 2018 to December 2019.

Chris Bosio, owner of the Tampa barbershop Headlines, is one such creator¹⁰. When initial business was slow, Chris used his free time to teach the other barbers in his shop new techniques. Impressed by his down-to-earth, easy-to-understand lessons, his business partners convinced him to upload a tutorial to YouTube to bring attention to the shop. The video was a hit, so Chris kept creating videos, and before long he saw his subscribers turn into clients. Within a couple of months, Headlines went from five clients a week to 800, many of them mentioning that they had watched Chris's YouTube videos before coming into the shop.

As Chris's channel grew, his subscribers began to ask him to make his own shaving accessories—a revenue stream Chris hadn't considered yet. He learned how to make a shaving gel from YouTube tutorials and soon launched a line called Tomb45. YouTube is the main way Chris promotes the line, which creates a constant sales funnel for the company. With the support of his subscribers, Tomb45 sold over 70% of its inventory the first day. Today, Tomb45 sells 10,000 products a month in 15 countries. His YouTube business was a lifeline during the COVID-19 shutdown, when barbershops were forced to close. Without it, Chris isn't sure his business could have survived.

Today, millions of channels from over 90 different countries earn revenue from their videos by participating in our YouTube Partner Program (YPP). Through YPP, creators earn revenue generated from advertising that is shown to viewers before or during a video. This revenue from ads is shared between YouTube and the creator, with the creator receiving a majority share—thus empowering creators to directly profit from their work.

But generating revenue on YouTube is a privilege, reserved for creators who meet specific eligibility requirements. To be eligible for YPP, a creator must have more than 1,000 subscribers, 4,000 watch hours in the last 12 months, and a track record of adhering to our Community Guidelines. YPP creators must also adhere to Google Ads policies, and any videos that are monetized must also meet an even higher bar by adhering to YouTube's Advertising-Friendly Guidelines¹¹ for content. These guidelines outline what content may not be monetized and what content warrants limited monetization. If creator content violates any of our Community Guidelines, that content will be removed from YouTube. We also enforce our Advertising-Friendly Guidelines by limiting or blocking ads on videos in accordance with those

¹⁰ <https://blog.youtube/creator-and-artist-stories/chris-bossio-tomb45/>

¹¹ <https://support.google.com/youtube/answer/6162278>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

guidelines. Creators who repeatedly violate any of our rules may be suspended from YPP for 90 days and need to apply again in order to rejoin YPP.

The 4Rs and Misinformation

We confront new challenges of balancing openness with responsibility every day. This is especially true when combating harmful misinformation on our platform. We invest a great deal of resources in research, policy development, technology, and experimentation to inform our approaches and improve our effectiveness in addressing misinformation on our platform. We continuously review our policies to evaluate whether the lines are in the right place; we prominently raise authoritative sources to the top of search results and make authoritative information readily available via a range of information panels; we regularly update our recommendation systems to hone our reduction of borderline content, including harmful misinformation; and we disincentivize creators seeking to profit off of misinformation by blocking ads on their content and suspending repeat offenders from our monetization program. We believe our 4R approach to responsibility provides a powerful and effective range of tools to combat harmful misinformation online, but we know there is more we can do. We commit to continuing to improve in our efforts to combat harmful misinformation on our platform.

Prioritizing transparency and accountability

At YouTube, we believe transparency is essential to earning and sustaining the trust of our users and our business partners. As a part of Google, we have led the way for the industry in terms of reporting on content removal at the request of governments and according to our own Community Guidelines, as well as information about government requests for information about users. We continue to expand our initiatives and the information we share, and we have rolled out three major resources over the last 12 months that underscore our commitment to transparency.

First, in May 2020, we collaborated with Google to launch the first Threat Analysis Group (TAG) Bulletin¹². The Bulletin—published on the TAG blog every quarter—discloses removal actions that Google and YouTube have taken to combat coordinated influence operations in a given quarter. Our hope is that this bulletin helps others who are also working to track these groups, including researchers working in this space, and that our information sharing can help confirm findings from security firms and other industry experts.

¹² <https://blog.google/threat-analysis-group>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

Second, in June 2020, we launched a website called How YouTube Works¹³, which was designed to answer the questions we most often receive about our responsibility efforts and to explain our products and policies in detail. How YouTube Works addresses some of the important questions we face every day about our platform, and provides information on topics such as child safety, harmful content, misinformation, and copyright. The site also covers timely issues as they arise, like our COVID-19 response, and our work to support election integrity. Within the site, we explain how we apply our responsibility principles—which work alongside our commitment to users’ security—to manage challenging content and business issues.

Third, YouTube publishes quarterly data in our Community Guidelines enforcement report¹⁴. This report provides public data about the number of videos we remove from our platform for each of our policy verticals (except spam) as well as additional information about channel removals, views before removals, appeals and reinstatements, and human and machine flagging. And as noted above, just this month, we updated our report to include the Violative View Rate to reflect how effectively we identify and take action on violative content.

These are important steps but we know we are being called to do more so that we can be held accountable for the decisions we make—algorithmic or otherwise. We appreciate the feedback that we have received from Members of Congress on our efforts to date and look forward to continuing to examine additional steps that could be taken to build upon our transparency efforts. We will continue to expand the information we share through our transparency report, cross-industry initiatives, blog posts, public disclosures, and other mechanisms like tools for researchers. Our goal is to achieve transparency and accountability by providing meaningful information while protecting our platform.

Child safety and digital wellbeing

As discussed above, responsibility is our number one priority at YouTube—and nowhere is this more important than when it comes to protecting kids. We continue to make significant investments in the policies, products and practices to help us do this. From our earliest days, YouTube has required our users to be at least 13 years old, and we terminate accounts belonging to people under 13 when they are discovered. In 2015, we created YouTube Kids, an app just for kids that we created to provide a safe destination to explore their interests while providing parental controls. With availability in over 100 countries, now over 35 million viewers

¹³ <https://www.youtube.com/howyoutubeworks/>

¹⁴ <https://transparencyreport.google.com/youtube-policy/removals>

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

use YouTube Kids every week. We continue to expand product availability, add new features, and offer several parental tools.

We have also heard from parents and older children that tweens have different needs that were not being fully met by our products. As children grow up, they have insatiable curiosity and a need to gain independence and find new ways to learn, create, and belong. Over the last year, we have worked with parents and child development experts across the globe in areas related to child safety, child development, and digital literacy. This collaboration informed a recently announced new supervised experience for parents on our main YouTube platform with three content settings for parents to choose from. The YouTube supervised experience looks much like YouTube's flagship app and website, but with adjustments to the features children can use and ads protections. For example, comments and live chat are disabled, as well as the ability to upload content and make purchases. Additionally, automatic reminders will appear for breaks and bedtime, which they can adjust to reinforce healthy screen-time habits.

In addition to these specially designed products, our YouTube main app treats personal information from anyone watching children's content on the platform as coming from a child, regardless of the age of the user. This means that on videos made for kids—whether explicitly designated as such by the creator or identified as child-directed by our content classification systems—we limit data collection and use, and as a result, we restrict or disable some product features. For example, we do not serve personalized ads on this content on our main YouTube platform or support features such as comments, live chat, notification bell, stories, and save to playlist. To be clear, we have never allowed personalized advertising on YouTube Kids.

Conclusion

Technology that uses algorithms is critical to our day-to-day operations, both in terms of basic user-facing functionality as well as content management and moderation at scale. But so, too, is input from people, whether by evaluating context, providing feedback on the quality of videos, or controlling how and when they choose to use YouTube. Just as machine learning systems are constantly taking new inputs to hone their pattern detection and efficacy, we work continuously to address new threats and identify ways to improve our systems and our processes. Responsibility is and will continue to be our number one priority—our business depends on it.

We look forward to continuing to engage and discuss areas where we share priorities, and how we can join together to support research to identify novel approaches to problems that threaten

YouTube, LLC
901 Cherry Avenue
San Bruno, CA 94066
USA



Phone +1 650 253.0000
www.youtube.com

both our users and your constituents, as well as thinking about how media literacy efforts can help users develop skills to build resiliency against misinformation.

Thank you for the opportunity to discuss our work with you today.

Monika Bickert – Algorithms and Amplification: How Social Media Platforms’ Design Choices Shape Our Discourse and Our Minds
Questions for the Record
 Submitted May 4, 2021

QUESTIONS FROM SENATOR COONS

1. As you testified, it is not in Facebook’s long-term interest to pursue short-term engagement or to push users toward increasingly extreme content. Some have suggested, however, that the metrics Facebook uses to set goals and measure product teams’ performance may create structural incentives within the company potentially at odds with this long-term interest.¹ During last Tuesday’s hearing, Tristan Harris testified that although Facebook does not measure teams’ performance based on time spent on the site, it does currently rely on metrics like numbers of sessions and seven-day active users growth. He also said that Facebook abandoned a brief experiment in setting goals for teams based on other, non-engagement-based metrics for social impact.
 - a. Please enumerate the metrics that Facebook uses to set goals for its product teams—particularly the News Feed Team—for purposes of performance and pay.
 - b. Is Mr. Harris correct about the metrics that Facebook uses?
 - c. Please explain why Facebook chose the metrics that it uses.
 - d. Is Mr. Harris correct that Facebook experimented with non-engagement-based metrics? If so, please explain what was tried and why this was discontinued.
 - e. Does Facebook include integrity metrics in its performance and pay assessments for any of its teams, including the Integrity Team?
 - f. To the extent there are tensions between topline goal metrics and integrity metrics in assessing business units or product features, how do you resolve them, and who makes those decisions?

2. The *MIT Technology Review* reported in March that Facebook does not have a team or initiative specifically focused on how its content ranking systems affect the spread of misinformation.² Facebook responded that this responsibility is distributed among teams with the relevant expertise.
 - a. Please elaborate on this response, and specifically explain which teams share in this responsibility and how they coordinate.
 - b. While Facebook has invested in its integrity teams, some have raised concern about the scope of integrity issues that these teams are charged with addressing and whether they have adequate staffing and resources, particularly if other teams are more oriented around growth. Please respond to this concern.

3. Reporting has indicated that one challenge for Facebook is the issue of “hyperactive users” or “super-invites” who generate a lot of activity and often promote either low-

¹ Karen Hao, “How Facebook got addicted to spreading misinformation,” *MIT Technology Review* (Mar. 11, 2021), <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/>.

² *Id.*

value or violating content.³ This reporting has raised the concern that Facebook may be reluctant to address these users' behavior because of the impacts that doing so would have on engagement and growth.

- a. Please provide information regarding the concentration of violating content among such "hyperactive users."
 - b. Please describe Facebook's efforts to address this issue, and in particular, please respond to the concerns raised in *BuzzFeed News* about "super-invites" associated with the "Stop the Steal" campaign and in the *Wall Street Journal* about the "Sparing Sharing" proposal that Facebook considered.
4. In Facebook's responses to my questions for the record from November last year, Facebook confirmed that it was building a "parallel viral content review system to flag posts that may be going viral—no matter what type of content they are—as an additional safety net." Facebook also pointed to a Crisis Assessment Dashboard to correlate spikes in hate speech or voter interference content in near real time. I appreciate Facebook's innovations in this space.
 - a. Please elaborate on the current status of these efforts, and in particular whether they are being implemented at scale.
 - b. Who decides where to set the threshold for virality in this system, and what factors are considered in making that decision?
 - c. Has Facebook found the parallel viral content review system to be effective? Please explain.
 5. I appreciate Facebook's recognition of the importance of estimating the prevalence of violating content and how this is a key statistic for engineers focused on misinformation. Although Facebook reports a single global prevalence statistic for hate speech on a quarterly basis, that statistic is not broken out into more granular categories.
 - a. Is Facebook able to estimate the prevalence of violating content on a more granular basis for internal purposes (e.g., as to the types of violations, the particular content sources, countries)?
 - b. To what extent are Facebook's estimates of prevalence factoring into company decision making on new products and features? Please explain how so.
 6. I appreciate your testimony that Facebook is committed to increasing transparency about your algorithmic recommendation engines. Many independent researchers have said it is essential to know not only what "signals" are considered but also how those signals are weighted; otherwise, it is not apparent which signals are actually driving the ranking. Reporting has also indicated that Facebook has developed certain higher-level "knobs

³ See Craig Silverman, Ryan Mac & Jane Lytvynenko, "Facebook Knows It Was Used to Help Incite the Capitol Insurrection," *BuzzFeed News* (Apr. 22, 2021), <https://www.buzzfeednews.com/article/craigsilverman/facebook-failed-stop-the-steal-insurrection>; Jeff Horwitz & Deepa Seetharaman, "Facebook Executives Shut Down Efforts to Make the Site Less Divisive" (May 26, 2020), <https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>.

and dials” such as “news ecosystem quality” that Facebook can adjust to modify the rankings of the News Feed, as it did in advance of the 2020 election.⁴

- a. Please provide more information about this news ecosystem quality metric, how it is calculated, and the effect that changing its weighting has on the News Feed.
 - b. What are the primary ways in which Facebook adjusted its weighting of factors in the News Feed algorithm in advance of events like the 2020 election and the Derek Chauvin verdict?
 - c. Will Facebook provide greater transparency into how it weighs different inputs for its News Feed rankings, whether through reporting, granting access to independent researchers, or conducting an audit?
7. You testified that Facebook has prioritized “meaningful social interactions” over engagement or time spent on the platform.
- a. How does Facebook measure “meaningful social interaction”?
 - b. Has Facebook analyzed the correlation of this metric with engagement or growth?
 - c. What is the practical impact of building products to prioritize meaningful social interactions?
8. During last Tuesday’s hearing, I asked whether Facebook generally requires employees to sign non-disclosure agreements (NDAs). You said you did not know the answer but would follow up. Can you provide an answer here?
- a. If Facebook does require employees to sign NDAs, please briefly describe the scope of those agreements and the types of information that employees and former employees are barred from disclosing.
 - b. Do the NDAs cover more than trade secrets, private user data, and/or technical information that employees learn while employed by Facebook?
 - c. Do the NDAs or any other contractual agreements bar employees from criticizing or disparaging Facebook after they leave the company?
 - d. If an employee disagrees with a policy decision made by Facebook during the employee’s period of employment, would the employee’s NDA prohibit him or her from discussing that disagreement publicly after leaving the company?
 - e. Do Facebook’s NDAs contain binding arbitration clauses—meaning employees must submit to arbitration in the event of any legal dispute under the terms of the agreement?
 - f. Has Facebook ever threatened legal action against an employee (current or former) based on potential violations of that employee’s NDA?
9. According to public reporting, some Facebook employees alleged that Facebook made special exceptions to its “repeat offender” policies to avoid taking measures against certain conservative pages, including Breitbart, because Facebook feared accusations of anti-conservative bias.⁵

⁴ See Kevin Roose, Mike Isaac & Sheera Frenkel, “Facebook Struggles to Balance Civility and Growth,” *New York Times* (Nov. 24, 2020), <https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html>.

⁵ Olivia Solon, “Sensitive to Claims of Bias, Facebook Relaxed Misinformation Rules for Conservative Pages,” NBC News (Aug. 7, 2020), <https://www.nbcnews.com/tech/tech-news/sensitive-claims-bias-facebook-relaxed->

- g. According to Facebook’s description, “Pages and websites that repeatedly share false information will see their distribution reduced and their ability to advertise removed.”⁶ Can you provide any more details about this repeat offender policy—such as how many times a page or user must share false information before Facebook reduces the distribution of content from that page or user?
- h. NBC News stated in its reporting that “The company operates on a ‘strike’ basis, meaning a page can post inaccurate information and receive a one-strike warning before the platform takes action. Two strikes in 90 days places an account into ‘repeat offender’ status.”⁷ Is that true?
- i. If employees and independent fact-checkers responsible for enforcing Facebook’s content policies determine that a page or user is a “repeat offender,” is that decision subject to review by anyone outside of Facebook’s content policy or integrity teams?
- j. Is public perception ever taken into consideration when determining whether or how to enforce Facebook’s repeat offender policies? If so, how?
- k. Are potential political repercussions ever taken into consideration when determining whether or how to enforce Facebook’s repeat offender policies? If so, how?
- l. Are potential business consequences—such as potential dissatisfaction from advertisers—ever taken into consideration when determining whether or how to enforce Facebook’s repeat offender policies? If so, how?

[misinformation-rules-conservative-pages-n1236182](#); Craig Silverman & Ryan Mac, “Facebook Fired an Employee Who Collected Evidence of Right-Wing Pages Getting Preferential Treatment,” BuzzFeed News (Aug. 6, 2020), <https://www.buzzfeednews.com/article/craigsilverman/facebook-zuckerberg-what-if-trump-disputes-election-results>.

⁶ Facebook, “How is Facebook addressing false information through independent fact-checkers?”, <https://www.facebook.com/help/1952307158131536>.

⁷ Solon, *supra*.

**Questions for the Record for Monika Bickert
From Senator Mazie Hirono**

1. During the hearing, I asked you how Facebook ensures it does not violate civil rights laws by discriminating on the basis of characteristics such as age, gender, and race when it targets ads for housing, employment, and financial services. Without equivocation, you testified that Facebook does not allow the use of “some of the more sensitive targeting criteria” when targeting ads for “certain types of advertisements, such as financial services advertisements or housing advertisements.” Two days after the hearing, *The Markup*¹ published an article in which it identified 91 unique ads promoting financial services products that explicitly discriminated on the basis of age. And recent litigation has shown that Facebook routinely serves non-credit financial services ads that are targeted based on age and gender.

Do you stand by your testimony that Facebook does not allow the use of “some of the more sensitive targeting criteria” when targeting ads for “certain types of advertisements, such as financial services advertisements or housing advertisements?” If not, please explain what in your testimony was inaccurate. If you contend your testimony was accurate, what is your explanation for the advertisements identified in the article in *The Markup* and pending litigation such as *Liapes v. Facebook, Inc.*?

2. Facebook’s Help Center describes what factors it uses to target ads to users, including “[o]ther information about you from your Facebook account” such as “your age [and] your gender.” Facebook also allows ads to be targeted based on things like zip code that can be used as a proxy for race.

Does Facebook allow advertisers to use factors such as age, gender, and zip code to target ads for housing, employment, credit, and other financial services (such as insurance, bank accounts, and debit cards)? If it does, how does Facebook ensure it does not violate civil rights laws when targeting these ads to users?

3. In 2019, Facebook entered into a settlement with the National Fair Housing Alliance and the Communications Workers of America in which it agreed to implement changes to its ad platform to prevent advertisers from targeting housing, employment, and credit ads based on age, gender, zip code (a proxy for race), and other factors.
 - a. **What changes did Facebook make to its ad platform to ensure factors such as age, gender, and zip code could not be used to target ads for housing, employment, and credit?**
 - b. **What additional steps has Facebook taken to ensure compliance with the settlement agreement?**

¹ <https://themarkup.org/citizen-browser/2021/04/29/credit-card-ads-were-targeted-by-age-violating-facebooks-anti-discrimination-policy>

- c. **Do the ads identified by *The Markup* comply with the settlement agreement? If you contend they do, please explain why. If they do not, how did they get through Facebook’s system?**
- d. **Does Facebook allow advertisers to use age, gender, and zip code to target non-credit financial services ads (e.g., insurance, bank accounts, debit cards)? If it does, why does Facebook consider it appropriate or desirable to allow discrimination for these types of non-credit financial services?**
4. Even when advertisers do not expressly target ads for housing, employment, and financial services on the basis of age, gender, and race, Facebook’s ad delivery algorithm may still serve housing, employment, and financial services ads in a discriminatory way. For example, during the hearing, I referenced an article in *MIT Technology Review*² that found “Facebook’s ad-delivery system shows different job ads to women and men even though the jobs require the same qualifications.”
- a. **Does Facebook’s ad delivery algorithm rely on protected statuses such as age, gender, and race or well-known proxies for protected statuses to determine which users will receive ads on Facebook, including ads for employment, housing, and financial services?**
- b. **What steps does Facebook take to ensure its ad delivery algorithm does not result in a disparate impact based on protected statuses such as age, gender, and race?**
- c. **How often, if ever, does Facebook audit housing, employment, and financial services ads served by its ad platform to ensure Facebook’s ad delivery algorithm does not result in a disparate impact based on protected statuses such as age, gender, and race? If Facebook has performed such audits, please provide the results, including any audits or reports on audits that Facebook has provided to any federal or state agencies, including the Department of Justice or the Department of Housing and Urban Development. If Facebook has not performed such audits, will it commit to doing so and sharing the results with this Subcommittee?**
- d. **Has Facebook considered any methods for reducing or eliminating the extent to which its ad delivery algorithm results in a disparate impact based on protected statuses such as age, gender, and race, including in any discussions with federal, state, or local agencies? If so, what are those methods.**
5. Last fall, Jack Dorsey testified before the full Senate Judiciary Committee that “people should have choices about the key algorithms that affect their experience online.” Following the hearing, I asked Mark Zuckerberg in a Question for the Record whether he agreed with Mr. Dorsey’s statement and if he would commit Facebook to allowing users to choose from a selection of curation algorithms. He failed to answer the question and instead described how Facebook frequently makes changes to the algorithms that drive Facebooks News Feed ranking “in an effort to improve people’s experience on Facebook.”

² <https://www.technologyreview.com/2021/04/09/1022217/facebook-ad-algorithm-sex-discrimination/>

- a. **Will Facebook commit to allowing users and/or advertisers to choose the algorithm that ranks and filters the content they see on the platform, including a choice of algorithms developed by third parties? If not, why not?**
 - b. **Will Facebook commit to giving advertisers the ability to deliver their ads with an algorithm that does not rely on protected statuses or proxies for them to target their ads, including ads for employment, housing and financial services?**
6. Facebook's independent Oversight Board recently began accepting appeals of content left up on the platform.

Does Facebook commit to expanding the Oversight Board's jurisdiction to include reviews of Facebook's algorithm for ranking and filtering content? If so, when will the Oversight Board be given this authority? If not, why not?

7. Facebook has claimed that it has Section 230 immunity when Facebook publishes ads that exclude people based on age, gender, and race and when it relies on protected statuses in its own ad delivery algorithm to decide which users receive ads, including ads for employment, housing, and financial services.

Why does Facebook believe it should receive Section 230 immunity for structuring its own platform in a discriminatory way and for taking its own actions to discriminate through its ad delivery algorithm even in cases where advertisers do not want Facebook to engage in such discrimination on their behalf?

Senator Ben Sasse
Questions for the Record
Senate Judiciary Committee Privacy, Technology, and the Law Subcommittee
Hearing: “Algorithms and Amplification: How Social Media Platforms’ Design Choices
Shape Our Discourse and Our Minds”
April 27, 2021

For Monika Bickert:

1. How does your firm define the ideal user experience? Does your answer differ between adult and children?
2. From your firm’s perspective, what is the ideal amount of time for a user to spend on the platform over a specified period of time? How much is too much?
3. If an individual pulled up her account on your platforms, could somebody at your company walk her through the exact calculation behind why the first post ended up where it is? And if not, what are the limitations, and how close could you get to a full explanation?
4. If a family member asked you for advice on whether and under what conditions the family member should allow her child to use your firm’s social media platform, what advice would you offer? How would it change based on age?
5. Do you think that there is anything fundamentally wrong with the business mode of your firm along the lines of the criticisms outlined by Dr. Donovan and Mr. Harris?
6. If somebody like Dr. Donovan who spends her time researching online extremists has her feed filled with extremist content, should we consider that to be a job well done in optimizing her user experience or something that should worry us?
7. If a user’s behavior indicates that he or she consistently chooses to focus on content that makes them angry or anxious or depressed, what is the responsibility of the platform in terms of curating the content that it places in front of that user?
8. Do you measure whether content that stimulates a reaction of anger elicits more engagement from users? If so, what are the results, and how are those results used to shape the platform?
9. In the hearing, I noted that political debate about regulating social media platforms often fails to respect the distinctions between the First Amendment limitations of government actors and the greater freedom enjoyed by private actors. While your firm is not a government entity, we should not overlook the history of cooperation between social media platforms and federal, state, and local governments in some content decisions. Many of these efforts, such as limiting access to terrorist propaganda, have long earned bipartisan support. Nevertheless, learning more about the role of government actors in

your content decisions would be a significant and helpful step in informing the ongoing debate. Please give an overview of your partnerships with federal, state, and local governments in making all aspects of content decisions (e.g., developing standards, removing content that violates these standards, deprioritizing or demonetizing content, flagging content as potentially inaccurate or misleading or providing links to authoritative resources, promoting certain kinds of content).

10. At the federal level, to what extent is this liaising with the federal government done by political appointees instead of career officials? Does contact occur at the agency level or the White House level? Please characterize the nature of this contact.
11. How often do you receive communications from legislators urging more aggressive content moderation?
12. Has your firm ever been the subject of threats of negative consequences by federal officials for refusing to take certain content moderation steps?
13. When the changes announced as part of the “Meaningful Social Interactions” initiative were rolled out, how much did traffic to Facebook decrease?
14. What is Facebook’s current thinking in regards to the potential for creating an ad-free, subscription-based version of the platform?

Lauren Culbertson – Algorithms and Amplification: How Social Media Platforms’ Design Choices Shape Our Discourse and Our Minds
Questions for the Record
Submitted May 4, 2021

QUESTIONS FROM SENATOR COONS

1. I appreciate your testimony that Twitter is focused on long-term usage of its products rather than prioritizing short-term engagement. Some have raised concern, however, about whether the goals and metrics that social media platforms use to measure employee performance may create structural incentives at odds with this long-term interest.
 - a. Please enumerate the metrics that Twitter uses to set goals for its product teams—particularly the timeline team—for purposes of performance and pay.
 - b. Please explain why Twitter chose those metrics.
 - c. Does Twitter take any integrity-related metrics into account for its performance and pay assessments for any of its teams, including its Trust and Safety Team?

2. Some have raised concerns about whether social media companies are internally organized to ensure that trust and safety teams have adequate voice, staffing, and resources in product development.
 - a. What teams are responsible for ensuring that the Timeline algorithm does not promote violating content? Please elaborate on how this process is structured and coordinated internally.
 - b. To the extent there are tensions between topline goal metrics and integrity metrics in assessing business units or product features, how do you resolve them, and who makes those decisions?

3. In a response to questions for the record I posed last November, Facebook said that it has been building a “parallel viral content review system to flag posts that may be going viral—no matter what type of content it is” and prioritize these posts for review. The idea is that potential virality should be a signal that content would be particularly dangerous if it violates platform policies, and thus warrants a trigger for human review.
 - a. Does Twitter incorporate potential virality in reviewing content?
 - b. Please explain the factors Twitter considers when determining which content to review.

4. Both Facebook and YouTube now report metrics on prevalence of violating content—that is, rather than reporting only on what the platforms are taking down, they are estimating what is not being caught and being viewed. Twitter has said, “We also recognize the importance of measuring prevalence of certain content on Twitter, and we have begun a multi-year initiative to enable us to provide more consistent transparency on these issues.”¹ I appreciate Twitter’s recognition of the importance of this metric. Please provide more information about the status of this initiative and why you expect it to take multiple years.

¹ Twitter, Insights from the 17th Twitter Transparency Report, https://blog.twitter.com/en_us/topics/company/2020/ttr-17.html.

5. I appreciate your testimony that Twitter is committed to increasing transparency about your algorithmic recommendation engines. Many independent researchers have said it is essential to know not only what “signals” are considered but also how those signals are weighted; otherwise, it is not apparent which signals are actually driving the ranking.
 - a. Please provide more information about the inputs, signals and weightings used in Twitter’s Timeline ranking algorithm.
 - b. Will Twitter provide greater transparency going forward into how it weighs different inputs for its Timeline rankings, whether through reporting, granting access to independent researchers, or conducting an audit?

6. Some have distinguished between sorting “connected content” (i.e., the tweets from individuals that a user has chosen to follow) and promoting “recommended content” (i.e., content that is not directly tweeted or liked by a user’s followers).
 - a. Please enumerate all instances in which Twitter currently recommends content not directly connected to people the user has chosen to follow (e.g., recommended “topics”).
 - b. What measures, if any, does Twitter take to ensure that it is not recommending content that violates its policies?
 - c. Will Twitter provide transparency on how often it recommends content that later proves to have violated Twitter policies?

7. During last Tuesday’s hearing, I asked whether Twitter generally requires employees to sign non-disclosure agreements (NDAs). You said that although Twitter has provisions in place to prevent employees from sharing “private data,” Twitter’s employees frequently share their perspectives about the company’s products and services.
 - a. Please briefly describe the scope of the NDAs that Twitter employees are required to sign, including the types of information that employees and former employees are barred from disclosing.
 - b. Do the NDAs cover more than trade secrets, private user data, and/or technical information that employees learn while employed by Twitter?
 - c. Do the NDAs or any other contractual agreements bar employees from criticizing or disparaging Twitter after they leave the company?
 - d. If an employee disagrees with a policy decision made by Twitter during the employee’s period of employment, would the employee’s NDA prohibit him or her from discussing that disagreement publicly after leaving the company?
 - e. Do Twitter’s NDAs contain binding arbitration clauses—meaning employees must submit to arbitration in the event of any legal dispute under the terms of the agreement?
 - f. Has Twitter ever threatened legal action against an employee (current or former) based on potential violations of that employee’s NDA?

8. In general, does Twitter ever take public perception into consideration when determining whether or how to enforce Twitter’s content policies? If so, how?
 - a. Are potential political repercussions ever taken into consideration when determining whether or how to enforce Twitter’s content policies? If so, how?

- b. Are potential business consequences—such as potential dissatisfaction from advertisers—ever taken into consideration when determining whether or how to enforce Twitter’s content policies? If so, how?
- c. According to Twitter’s “public-interest exception” policy, “in rare instances, [Twitter] may choose to leave up a Tweet from an elected or government official that would otherwise be taken down. Instead we will place it behind a notice providing context about the rule violation that allows people to click through to see the Tweet. . . . These actions are meant to limit the Tweet’s reach while maintaining the public’s ability to view and discuss it.”² Please describe the decision-making process that Twitter uses in determining when to apply this policy, and the primary factors that it takes into account.

² Twitter, “About public-interest exceptions on Twitter,” <https://help.twitter.com/en/rules-and-policies/public-interest>.

Senator Ben Sasse
Questions for the Record
Senate Judiciary Committee Privacy, Technology, and the Law Subcommittee
Hearing: “Algorithms and Amplification: How Social Media Platforms’ Design Choices
Shape Our Discourse and Our Minds”
April 27, 2021

For Lauren Culbertson:

1. How does your firm define the ideal user experience? Does your answer differ between adult and children?
2. From your firm’s perspective, what is the ideal amount of time for a user to spend on the platform over a specified period of time? How much is too much?
3. If an individual pulled up her account on your platforms, could somebody at your company walk her through the exact calculation behind why the first post ended up where it is? And if not, what are the limitations, and how close could you get to a full explanation?
4. If a family member asked you for advice on whether and under what conditions the family member should allow her child to use your firm’s social media platform, what advice would you offer? How would it change based on age?
5. Do you think that there is anything fundamentally wrong with the business mode of your firm along the lines of the criticisms outlined by Dr. Donovan and Mr. Harris?
6. If somebody like Dr. Donovan who spends her time researching online extremists has her feed filled with extremist content, should we consider that to be a job well done in optimizing her user experience or something that should worry us?
7. If a user’s behavior indicates that he or she consistently chooses to focus on content that makes them angry or anxious or depressed, what is the responsibility of the platform in terms of curating the content that it places in front of that user?
8. Do you measure whether content that stimulates a reaction of anger elicits more engagement from users? If so, what are the results, and how are those results used to shape the platform?
9. In the hearing, I noted that political debate about regulating social media platforms often fails to respect the distinctions between the First Amendment limitations of government actors and the greater freedom enjoyed by private actors. While your firm is not a government entity, we should not overlook the history of cooperation between social media platforms and federal, state, and local governments in some content decisions. Many of these efforts, such as limiting access to terrorist propaganda, have long earned bipartisan support. Nevertheless, learning more about the role of government actors in

your content decisions would be a significant and helpful step in informing the ongoing debate. Please give an overview of your partnerships with federal, state, and local governments in making all aspects of content decisions (e.g., developing standards, removing content that violates these standards, deprioritizing or demonetizing content, flagging content as potentially inaccurate or misleading or providing links to authoritative resources, promoting certain kinds of content).

10. At the federal level, to what extent is this liaising with the federal government done by political appointees instead of career officials? Does contact occur at the agency level or the White House level? Please characterize the nature of this contact.
11. How often do you receive communications from legislators urging more aggressive content moderation?
12. Has your firm ever been the subject of threats of negative consequences by federal officials for refusing to take certain content moderation steps?
13. What is Twitter's current thinking in regards to the potential for creating an ad-free, subscription-based version of the platform?
14. What percentage of users enable to chronological ordering function on Twitter?

Dr. Joan Donovan – Algorithms and Amplification: How Social Media Platforms’ Design Choices Shape Our Discourse and Our Minds”
Questions for the Record
Submitted May 4, 2021

QUESTIONS FROM SENATOR COONS

1. You have written in your testimony and elsewhere about the ability of coordinated groups to conduct misinformation campaigns by manipulating tech platforms’ algorithms and other features. You have also written that “mitigation efforts” undertaken by these platforms “often come too late.”¹
 - a. In your view, what are the key early warning signals that should alert social media companies to the existence of a coordinated misinformation campaign?
 - b. Some platforms do not disclose complete information about their policies for sanctioning users who violate their terms of service on the grounds that disclosing this information could allow bad actors to game their systems. In your view, are there circumstances in which greater transparency regarding enforcement policies could have that effect?
2. You have spoken about the need for social media platforms to curate content in a way that proactively amplifies reliable and accurate sources of information.
 - a. In your view, what are the kinds of incentives that could drive platforms toward such a model?
 - b. Please provide any further detail you would like about the ways in which a curation-based model could operate within a social media platform.

¹ Brian Friedberg, Emily Dreyfuss, Gabrielle Lim & Joan Donovan, “A Blueprint for Documenting and Debunking Misinformation Campaigns,” Nieman Reports (Oct. 20, 2020), <https://niemanreports.org/articles/a-blueprint-for-documenting-and-debunking-misinformation-campaigns/>.

Senator Ben Sasse
Questions for the Record
Senate Judiciary Committee Privacy, Technology, and the Law Subcommittee
Hearing: “Algorithms and Amplification: How Social Media Platforms’ Design Choices
Shape Our Discourse and Our Minds”
April 27, 2021

For Dr. Joan Donovan:

1. Is there anything offered in the testimony of Ms. Bickert, Ms. Culbertson, Ms. Veitch, or even Mr. Harris with which you disagree?

Tristan Harris – Algorithms and Amplification: How Social Media Platforms’ Design Choices Shape Our Discourse and Our Minds
Questions for the Record
Submitted May 4, 2021

QUESTIONS FROM SENATOR COONS

1. You have stated in your testimony and elsewhere that social media companies’ efforts to address harmful content on their platforms will inevitably fall short because their business models rely upon engaging their users’ attention in order to sell advertising.
 - a. In your view, what other incentives ought to drive these companies’ business models?
 - b. Are there market-based solutions, in your view, that could drive social media platforms to adopt incentives that more positively impact our society?
 - c. Are there any regulatory solutions, in your view, that could drive social media platforms to adopt incentives that more positively impact our society?

2. I strongly agree with you that the United States and our allies need to work toward a “Western digital infrastructure” that preserves the values of democracy and an open internet. In your view, what key principles would form the basis for such a structure?

Senator Ben Sasse
Questions for the Record
Senate Judiciary Committee Privacy, Technology, and the Law Subcommittee
Hearing: “Algorithms and Amplification: How Social Media Platforms’ Design Choices
Shape Our Discourse and Our Minds”
April 27, 2021

For Tristan Harris:

1. In the hearing, you offered a glimpse of what other potential business models could be for social media platforms instead of one based primarily on maximizing ad revenue through maximized user engagement. Please use this opportunity to sketch out in further detail what you think different potentially feasible models might be, what the costs and benefits are to each approach, and the rough level of government intervention required to achieve the change in business model.
2. With a subscription-based business model, should we worry that social media platforms will still have an incentive to maximize user engagement in order to ensure that customers think the subscription fee is worth the price?
3. Which are you more worried about: nefarious actors exploiting legitimate tools to propagate misinformation purposefully or misinformation being propagated in the service of selling ads to legitimate advertisers?
4. When we’re having these conversations, it’s hard not to worry about a dystopian vision of the future, either driven by chaos or authoritarianism. Why shouldn’t we be modern-day Luddites? Should we be techno-optimists?
5. How would we educate individuals and families to be smart consumers of social media?

**Responses to Questions for the Record
U.S. Senate Committee on the Judiciary
Subcommittee on Competition Policy, Antitrust, and Consumer Rights
“Algorithms and Amplification: How Social Media Platforms’ Design Choices Shape Our
Discourse and Our Minds”
April 27, 2021**

**Alexandra Veitch
Director Of Government Affairs And Public Policy For The Americas And Emerging
Markets, YouTube**

Written Questions Submitted by Sen. Sasse to Alexandra Veitch:

Question 1. How does your firm define the ideal user experience? Does your answer differ between adult and children?

Because the answers to these questions are related, we have grouped together the responses to Question Numbers 1 and 4.

At YouTube, we believe that the Internet has been a force for creativity, learning, and access to information. Supporting this free flow of ideas is at the heart of our mission to give everyone a voice and show them the world. We have built and continue to improve YouTube to empower users to access, create, and share information like never before; this has enabled billions to benefit from a bigger, broader understanding of the world. Over the years, we have seen more and more people come to YouTube to share their experiences and understand their world more deeply. This is especially true when it comes to learning new skills, participating in civic engagement, and developing informed opinions about current events.

YouTube is committed to supporting users of all ages to develop healthy relationships with technology and use our services in ways that support their mental health and wellbeing. We work to understand both the opportunities and the challenges our users face, including through research and engagement with experts. And we build on this by developing features in our products and design policies that help our users — particularly children and families — form healthier digital habits.

YouTube has always had policies that govern how people may use the service, including restrictions on the types of content that they may post. These policies are designed and regularly updated to make YouTube a safer and more welcoming place for users, creators, and advertisers. They reflect years of experience and investment, and an ongoing conversation between YouTube and its users, creators, and advertisers.

For many years, we've worked with child development specialists to provide guidance to YouTube creators on how to create high-quality content for children. As a result of extensive consultations on children's media, digital learning, and the study of good citizenship, we established a set of quality principles (available at <https://support.google.com/youtube/answer/10774223>) to help guide our kids and family creator ecosystem. These principles include content that demonstrates or encourages respect and healthy habits, such as being a good friend, content that is thought-provoking or imaginative, such as arts and crafts activities, and content that celebrates and encourages diversity and inclusion. We also use these principles to determine which high-quality content we raise up in our recommendations on YouTube. This means that when you're watching content for kids and families on YouTube, we aim to recommend videos that are age-appropriate, educational, and inspire creativity and imagination. Additionally, these principles supplement our Community Guidelines (available at <https://support.google.com/youtube/answer/9288567>), which help provide a safe viewing experience for our users.

In 2015, we created YouTube Kids as a way for kids to more safely explore their interests and curiosity while providing parents more tools to control and customize the experience for their families. Videos on YouTube Kids include popular children's videos, as well as collections of channels selected by the YouTube Kids team and trusted kids content partners like Sesame Workshop, PBS Kids, and UNICEF. Our approach to determining eligibility in YouTube Kids is to first identify a very small subset of channels that have a low likelihood of uploading inappropriate content from the full YouTube corpus of content that is available.

We also provide parents the ability to fully customize what their kids watch. For example, parents can block a video or channel, handpick which videos to make available for their kids, and select an option to only allow content selected by trusted partners. We are constantly working to improve our safeguards and offer more features to help parents create the right experience for their families. For more information on the content policies for YouTube Kids, please see <https://support.google.com/youtube/answer/10938174?hl=en>.

After we launched YouTube Kids, we heard from parents of older children that tweens and teens have different needs which were not fully met by our products. That is why we worked with parents and experts across the globe in areas related to child safety, child development, and digital literacy to develop a solution for parents of tweens — which we call Supervised Experience — that launched in March 2021 on the main YouTube platform. Supervised Experience allows parents to choose between three different content choices: content generally suitable for viewers aged 9+; content generally suitable for viewers 13+; and the 'Most of YouTube' option, which excludes all age-restricted content (18+).

To be clear, kids under 13 who are not in a Supervised Experience are not allowed on YouTube. When a parent opens a supervised Google account for their child, the experience feels much

like regular YouTube but certain features are disabled for younger audiences. Underlying the ads and creator policies applicable to both YouTube Kids and Supervised Experiences are core principles aimed at maintaining an appropriate viewing environment for children and families and protecting young users' privacy. For example, on both YouTube Kids and Supervised Experiences we prohibit personalized advertising and age-sensitive ad categories (such as those available at <https://support.google.com/youtube/answer/9713557?hl=en>). We do not allow ads that contain adult, dangerous, sexualized or violent content; ads that mislead or make deceptive claims; and sensitive product categories such as ads for pharma/healthcare, dating, politics, weight loss, beauty and fitness.

To help parents understand more about the YouTube Supervised Experience, we have developed guides (such as <https://www.youtube.com/myfamily/>), videos (such as <https://www.youtube.com/watch?v=oVOa6nDU7HQ>), and support pages (available at <https://support.google.com/youtube/answer/10314940?hl=en>), building on Google's successful Be Internet Awesome digital literacy resources, and in partnership with the National PTA, Parent Zone UK and other leading experts. We will continue to partner with these and other groups to provide easy to use resources specifically for parents to help them keep their kids safe online.

Question 2. From your firm's perspective, what is the ideal amount of time for a user to spend on the platform over a specified period of time? How much is too much?

Our business relies on the trust of our users. We want to build these relationships for the long term. That is why we bake user choice and user control right into the product with things like timers and the opportunity to turn auto-play off.

We also employ systems that enable our users to make decisions on screen time that are best for them and their families. For example, Family Link is a downloadable app (available at <https://families.google.com/familylink/>) now available by default in the latest Android operating system and Chromebooks that helps parents guide their child's experience as they explore online. The app lets parents set digital ground rules for their family, such as managing the apps their child can use, keeping an eye on screen time, or setting a bedtime and daily limits for their child's device. SafeSearch, a default Family Link supervision feature, helps filter out explicit search results on Google Search. And, as discussed in the response above, YouTube Kids provides a separate YouTube experience designed especially for children that parents can customize. Parents can decide what content to make available for their child to watch, set a timer to control screen time, block videos or channels, and more. For more information on the tools we provide on YouTube Kids to help parents manage screen time, please see <https://support.google.com/youtubekids/answer/6130558>.

Question 3. If an individual pulled up her account on your platforms, could somebody at your company walk her through the exact calculation behind why the first post ended up where it is? And if not, what are the limitations, and how close could you get to a full explanation?

Recommendations on YouTube help users discover videos they may enjoy and creators reach new viewers to grow their audience across the platform. Recommendations occur in two main places: a user's homepage and the "Up Next" panel. Our recommendation system takes into account many signals, including: a user's viewing habits, such as clicks, watchtime, survey responses, sharing, likes, and dislikes; and engagement signals about the video itself — for example, whether others who clicked on the same video watched it to completion or clicked away shortly after starting to view the video.

We also empower our users by giving them control over personalized recommendations. In June 2019, we launched new features that give users more control over what recommendations appear on the homepage and in their "Up Next" suggestions. Three specific changes were made in these new features to make it easier for users to block channels from recommendations, give users the option to filter recommendations on Home and on "Up Next," and give users more information about why we are suggesting a video:

- **Explore topics and related videos on your Homepage and in Up Next videos:** Users can now more easily explore topics and related videos on their Homepage and in Up Next videos to see more of what they love on YouTube. Users now see options based on existing personalized suggestions that are meant to help users find what they are looking for faster.
- **Remove suggestions from channels you don't want to watch:** Users can easily let us know to stop suggesting videos from a particular channel by tapping the three-dot menu next to a video on the homepage or Up Next, then "Don't recommend channel." After that, users should no longer see videos from that channel suggested to them on YouTube.
- **Learn more about why a video may be suggested to you:** Sometimes we recommend videos from channels users have not seen before based on what other viewers with similar interests have liked and watched in the past. When we're suggesting videos based on this, users now see more information underneath the video in a small box. Our goal is to explain why these videos surface on your homepage in order to help you find videos from new channels you might like.

Question 4. If a family member asked you for advice on whether and under what conditions the family member should allow her child to use your firm's social media platform, what advice would you offer? How would it change based on age?

Please refer to the answer to Question No. 1, above.

Question 5. Do you think that there is anything fundamentally wrong with the business model of your firm along the lines of the criticisms outlined by Dr. Donovan and Mr. Harris?

At YouTube, we take a comprehensive approach to combating online extremism, abuse and other harmful content. Over the last few years, we have significantly increased our investments in the systems and processes that allow us to remove violative content, raise up authoritative content, reward trusted creators and artists, and reduce the spread of borderline content. Our teams work hard every day to ensure that we are allowing for a broad range of political speech while adhering to our responsibility of making sure that our platform is not abused to incite real-world harm or spread harmful misinformation.

Managing misinformation and harmful conspiracy theories is challenging because the content is always evolving, but we take this issue very seriously. Due to the shifting tactics of groups promoting conspiracy theories, we continuously invest in the policies, resources, and products needed to protect our users from harmful content. We have clear and public policies that we apply consistently. Across all of Google platforms, we have dedicated significant resources to content moderation efforts: we spent more than one billion dollars on content moderation efforts in 2020 and have nearly twenty thousand employees dedicated to ensuring the protection of our platforms.

In addition to removing content that violates our policies, we also work to ensure that we do not proactively expose users to content that is potentially harmful. We reduce recommendations of the "borderline content", or content which comes close to but does not violate our policies, that represents less than 1% of the content watched on YouTube. In January 2019, we announced that we would begin reducing recommendations of borderline content or videos that could misinform users in harmful ways, and we continue to extend these efforts to more countries outside of the United States and into non-English-language markets. We have launched more than 30 different changes to our recommendations systems on YouTube in order to reduce recommendations of borderline content and harmful misinformation. In 2019, we saw a more than 70% average drop in "watch time" of this content coming from non-subscribed recommendations in the United States.

Determining what is harmful misinformation or borderline content is challenging, especially given the wide variety of videos uploaded to YouTube. To do it, we rely on external evaluators from around the world to provide input on the quality of a set of videos. These evaluators use the same rater guidelines as Google Search to guide their work (these guidelines are available at <https://support.google.com/youtube/answer/9230586>). Each evaluated video receives up to nine different ratings, with some content requiring ratings from certified experts in the field. For example, medical doctors provide guidance on the validity of videos about specific medical treatments to limit the spread of medical misinformation. Based on consensus input from these raters, we use well-tested machine learning systems to build models that help

review hundreds of thousands of hours of videos every day to identify and limit the spread of borderline content. The accuracy of these systems continues to improve over time.

We also raise up authoritative voices by providing users with more information about the content they are seeing to allow them to make educated choices. For more information on information panels, please see <https://support.google.com/youtube/answer/9004474>, and on fact-checks, please see <https://support.google.com/youtube/answer/9229632>.

The openness of our platforms has helped creativity and access to information thrive. It is our responsibility to protect that, and to prevent our platforms from being used to incite hatred, harassment, discrimination, and violence. We are committed to taking the steps needed to live up to this responsibility today, tomorrow, and in the years to come.

Question 6. If somebody like Dr. Donovan who spends her time researching online extremists has her feed filled with extremist content, should we consider that to be a job well done in optimizing her user experience or something that should worry us?

The rise of misinformation in recent years led us to further expand the ways we use our recommendation system to address problematic misinformation and borderline content. This includes conspiracy theory videos (“the moon landing was faked”) or other content that spreads misinformation (“orange juice can cure cancer”).

We are able to do this by using classifiers to identify whether a video is “authoritative” or “borderline”. To determine authoritativeness, we look at a few key questions: Does the content deliver on its promise or achieve its goal? What kind of expertise is needed to achieve the video goal? What is the reputation of the speaker in the video and the channel it is on? What is the main topic of the video (e.g., News, Sports, History, Science, etc)? Is the content primarily meant to be satire? These answers and more determine how authoritative a video is. The higher the score, the more the video is promoted when it comes to news and information content.

To determine borderline content, we assess factors that include — but aren’t limited to — whether the content is: inaccurate, misleading or deceptive; insensitive or intolerant; or harmful or with the potential to cause harm. The results are combined to give a score for how likely the video contains harmful misinformation or is borderline. Any video classified borderline is demoted in recommendations. These human evaluations then train our system to model their decisions, and we now scale their assessments to all videos across YouTube.

To create a safe environment for our users at YouTube, we approach our content recommendations with our four “Rs” of responsibility: we remove content that violates our policies as quickly as possible, we raise up authoritative voices, we reduce the spread of content that brushes up against our policy (i.e., borderline content), and we reward trusted creators.

- **We remove content that violates our policies.** In Q2 2022 alone, we removed more than 4.4 million videos for violating our Community Guidelines, more than 93% of which were first flagged by our automated systems. More than 72% of those first flagged by our systems received 10 or fewer views. Most recently, we began disclosing our violative view rate (VVR), or the number of views on content that violates our Community Guidelines as a portion of total viewership. More information on our VVR is available at <https://transparencyreport.google.com/youtube-policy/views>. In Q2 2022 our VVR was 0.09-0.11%, meaning that out of every 10,000 views on YouTube, 9-11 came from violative content.
- **We raise up authoritative voices when people are looking for news and information.** For example, we raise content from authoritative channels for newsworthy events in YouTube search results and “Up Next” recommendations panels. In addition to current events, authoritative content is also important when it comes to topics prone to misinformation, such as vaccines. In these cases, we aim to surface videos from experts, like public health institutions, in search results. We also provide informational panels to provide contextual information on specific content in both the search results page and the video watch page. In October 2020, we updated YouTube’s hate and harassment policy to prohibit content targeting individuals or protected groups based on conspiracy theories used to justify real world violence (e.g., QAnon and Pizzagate).
- **We reduce recommendations of borderline content.** As referenced above, in January 2019, we announced improvements to our recommendation system to greatly reduce recommendations of borderline content and content that could misinform users in harmful ways. Globally, borderline content that comes from our recommendations represents less than 1% of the content watched on YouTube.
- **We reward trusted creators by allowing those channels to make money on our site.** We have strict policies on the kind of videos we allow ads to appear on that are strictly enforced, and creators must meet specific eligibility requirements to join our monetization program.

We are proud of these efforts, and believe they work to help curb the spread of extremist content. A growing number of independent researchers have been looking into how tech platforms impact the consumption of borderline content, and, while ongoing study continues, certain recently published papers conclude YouTube recommendations do not steer viewers towards extreme content. Instead, consumption of news and political content on YouTube more generally reflects personal preferences that can be seen across their online habits. To encourage further research, we recently launched the YouTube Research Program (more information is available at <https://research.youtube/>), which gives academic researchers access to our raw data in a user privacy sensitive way, allowing them to derive metrics or develop insights of their own based on that data.

Question 7. If a user's behavior indicates that he or she consistently chooses to focus on content that makes them angry or anxious or depressed, what is the responsibility of the platform in terms of curating the content that it places in front of that user?

Please refer to the answer to Question No. 8, below.

Question 8. Do you measure whether content that stimulates a reaction of anger elicits more engagement from users? If so, what are the results, and how are those results used to shape the platform?

Because the answers to these questions are related, we have grouped together our response to Question Numbers 7 and 8.

Across our different platforms, we strive to connect people with “high-quality information” — the most useful, trustworthy, and helpful content — at the moment a person needs it. At the same time, we work to prevent user and societal harm and limit the reach of “low-quality information” — content that strays furthest from those qualities. We build products that help users access high-quality information that supports their mental health and wellbeing.

At YouTube, we take the health and well-being of all our creators and viewers seriously. Awareness and understanding of mental health is important and we support creators sharing their stories, such as posting content discussing their experiences with depression, self-harm, or other mental health issues. We also have a clear set of publicly available Community Guidelines (available at <https://support.google.com/youtube/answer/9288567>) that outline content that is not allowed on the platform.

Per our suicide and self-injury policy, we prohibit content that promotes self-harm or is intended to shock or disgust viewers. We remove content promoting or glorifying suicide, content providing instructions on how to self-harm or die by suicide and content containing graphic images of self-harm posted to shock or disgust viewers. The Stanford Internet Observatory highlighted how we handle searches related to self-harm and suicidal ideation, writing that they were “impressed that YouTube’s Community Guidelines on suicide and self-injury provide resources, including hotlines and websites, for those having thoughts of suicide or self-harm, for 27 countries,” and that creator platforms like YouTube “offer meaningful support in their policies both for people who are recovering from mental health issues and those who may be considering self-harm.” The report detailing these findings is available at <https://cyber.fsi.stanford.edu/io/self-harm-policies-report>.

YouTube uses a combination of machine learning and human review to enforce our policies and we regularly report on the content removed for violating our policies in our quarterly Community Guidelines Enforcement Report (available at <https://transparencyreport.google.com/youtube-policy/removals>). In Q2 2022, we removed over 900,000 videos for violating our violent or graphic content policy and over 1.3 million videos for violating our child safety policy, the majority of which were removed before they

had received 10 views. Most recently, we began disclosing our violative view rate (VVR), or the number of views on content that violates our Community Guidelines as a portion of total viewership. More information on our VVR is available at <https://transparencyreport.google.com/youtube-policy/views>. In Q2 2022, our VVR was 0.09-0.11%, meaning that out of every 10,000 views on YouTube, 9-11 came from violative content.

Question 9. In the hearing, I noted that political debate about regulating social media platforms often fails to respect the distinctions between the First Amendment limitations of government actors and the greater freedom enjoyed by private actors. While your firm is not a government entity, we should not overlook the history of cooperation between social media platforms and federal, state, and local governments in some content decisions. Many of these efforts, such as limiting access to terrorist propaganda, have long earned bipartisan support. Nevertheless, learning more about the role of government actors in your content decisions would be a significant and helpful step in informing the ongoing debate. Please give an overview of your partnerships with federal, state, and local governments in making all aspects of content decisions (e.g., developing standards, removing content that violates these standards, deprioritizing or demonetizing content, flagging content as potentially inaccurate or misleading or providing links to authoritative resources, promoting certain kinds of content).

Please refer to the answer to Question Number 10, below.

Question 10. At the federal level, to what extent is this liaising with the federal government done by political appointees instead of career officials? Does contact occur at the agency level or the White House level? Please characterize the nature of this contact.

Because the answers to these questions are related, we have grouped together our responses to Question Numbers 9, 10, and 12.

At YouTube, we develop and maintain Community Guidelines that outline what type of content is not allowed on our platform. These policies (available at <https://support.google.com/youtube/answer/9288567>) act as our 'rules of the road' and we aim to make them clear and easily accessible to all users and content creators. Our Community Guidelines also define what content and behaviors are not allowed, and the process by which a piece of content, or its creator, may be removed from service. These Community Guidelines are one of the ways we balance allowing diverse voices on the platform while limiting harmful content. We enforce our content policies at scale and take tens of millions of actions every day against content that does not abide by the policies for one or more of our products.

Courts of law and government agencies around the world regularly request that we remove content from YouTube. Some requests may allege infringement of intellectual property rights, while others claim violation of local laws prohibiting certain types of user generated content,

such as defamation or Neo-Nazi content in certain members states of the European Union. We review these requests closely to determine if content should be removed or restricted because it violates a local law or is contrary to our terms of service and content policies. In order for us to evaluate a request from a government entity, it must be in writing and provide a clear explanation of how the content is violative of law or policies.

Legal standards vary by country, and content that violates a specific law in one country may be legal in others. Typically, we remove or restrict access to the content only in the country where it is deemed to be illegal. But when content is found to violate our Community Guidelines we remove it globally.

It is important to note that YouTube does not always remove content in response to a government request. Some requests may not be specific enough for us to know what the government wanted us to remove (for example, no URL is listed in the request) or lack sufficient explanation of why the government believes content to violate a particular law. For defamation claims, we generally rely on courts to decide if a statement is defamatory according to local law, and do not rely on informal letters from government agencies.

With regard to government requests for user information, a variety of laws allow government agencies around the world to request the disclosure of user information for civil, administrative, criminal, and national security purposes. Each request is carefully reviewed to make sure it satisfies applicable laws. For more information, see our policies for how Google handles government requests for user information (available at <https://policies.google.com/terms/information-requests>).

We also publicly share information on government requests to remove content to shed light on the scale and scope of government requests for content removals. Data on government removal requests can be located on the Google Transparency Report (available at <https://transparencyreport.google.com/government-removals/overview?hl=en>).

There are also a number of ways we proactively work with law enforcement agencies to assess threats and to counter attempts to deceive, harm, or take advantage of people using our platforms. Our interactions with law enforcement include the following:

- First, we maintain regular communication channels with law enforcement, as communication with law enforcement, industry partners, and the federal government is a key component of our efforts to keep users safe. We rely on information from industry peers and law enforcement to ensure the integrity of our platform and act swiftly in response to crises or when we detect abuse that may threaten public safety or the integrity of democratic processes.
- Second, we have dedicated teams of analysts and security experts who work around the clock to identify and investigate possible instances of coordinated influence

operations on our platform. Google's Threat Analysis Group works to identify bad actors and terminate their channels and accounts, works with other technology companies to share intelligence and best practices, and shares threat information with law enforcement. For more information regarding our Threat Analysis Group, please see <https://blog.google/threat-analysis-group/>.

- Third, we have processes in place to proactively refer to law enforcement imminent threats and certain illegal activity occurring on our platform. These imminent threats are raised to us in a variety of ways. In addition to our ongoing threat analysis just noted, we also have processes in place for our various internal product teams, as well as external sources, to escalate potential threats and criminal activity if they see it on our platforms. Google has a team - the CyberCrime Investigation Group (CCIG) - in place to assess escalated threats, and that team refers matters to law enforcement whenever appropriate.

Additionally, we developed the YouTube Trusted Flagger program to help provide more sophisticated reporting processes for government agencies and non-governmental organizations (NGOs) that are particularly effective at notifying YouTube of content that violates our Community Guidelines. The program provides these partners with dedicated reporting processes and a channel for ongoing discussion and feedback about YouTube's approach to various content areas. The program is part of a network of around 300 government partners and NGOs that bring valuable expertise to our enforcement systems. Participants in the Trusted Flagger program receive training in enforcing YouTube's Community Guidelines, and because their flags have a higher action rate than the average user, we prioritize them for review. Once flagged, our YouTube content reviewers make the call on what content is or isn't removed by applying our YouTube policies. We apply our policies transparently and consistently, without taking into account either the political leanings of the author, or their place in society.

Question 11. How often do you receive communications from legislators urging more aggressive content moderation?

Our mission at YouTube is to "give everyone a voice and show them the world." As such, our search and recommendation systems are not designed to filter or demote videos or channels based on specific political perspectives. We take extraordinary lengths to build our products and enforce our policies in a consistent manner. Our business model is dependent on being a useful and trustworthy source of information for everyone, so we have a natural, long-term business incentive to prevent anyone from interfering with the integrity of our products.

When developing and updating our policies, we solicit perspectives from a range of voices, including creators, subject-area experts, free speech proponents, and policy organizations from across the political spectrum. Once a policy has been developed, we invest significant time making sure new policies are consistently enforced by our global team of reviewers who

follow objective guidelines Our work is also guided by our four pillars of responsibility: removing violative content; raising up authoritative content; reducing the spread of borderline content; and rewarding trusted creators — what we refer to as the 4Rs of responsibility. YouTube’s Community Guidelines outline what content is permissible and what content is violative on our platform. We enforce these Community Guidelines using a combination of human reviewers and machine learning, and apply them to everyone equally — regardless of the subject or the creator’s background, political viewpoint, position, or affiliation.

We are committed to ensuring that our participation and associations with politics, trade associations, or third-party organizations is open, transparent and clear to our users, shareholders, and the public. Information on our public policy engagement, political contributions, lobbying efforts, and memberships — including our biannual list of politically engaged trade associations, independent third-party organizations, and other tax-exempt groups that receive the most substantial contributions from our U.S. Government Affairs and Public Policy Team — is published in our US Public Policy Transparency Reports (available at <https://www.google.com/publicpolicy/transparency/>).

Question 12. Has your firm ever been the subject of threats of negative consequences by federal officials for refusing to take certain content moderation steps?

Please refer to the answer to Question Number 10, above.

Question 13. Why is YouTube’s autoplay feature enabled by default?

The Autoplay feature on YouTube makes it easier to decide what to watch next. When Autoplay is on, another related video will automatically play after a video ends. Users can have different Autoplay settings for different devices. For example, users can set Autoplay to be “On” in the YouTube app on their mobile device, but set it to “Off” when watching YouTube on their computer. If users are connected to a mobile network, Autoplay will stop automatically if they have been inactive for 30 minutes. If users are on Wi-Fi, Autoplay will stop automatically after four hours. For more information on YouTube Autoplay, please see

<https://support.google.com/youtube/answer/6327615>

For teens and for those in our YouTube Supervised Experience and in YouTube Kids, autoplay is now off by default. We have also launched additional parental controls in the YouTube Kids app, including the ability for a parent to choose a “locked” default autoplay setting. Our “Take a Break” reminders and bedtime settings are also on by default in these experiences. In addition, we have worked with creators to develop a series of PSAs to help children reflect on the time they spend online, and build empathy for others.

Question 14. What percentage of your users subscribe to YouTube Premium?

As of September 2021, we have 50 million Music and Premium subscribers, including trial members. It is an honor to build a membership that allows people to more deeply immerse

themselves in music, learning, fashion, gaming, and more, all the while supporting the creators and artists that make it possible. Music and Premium subscriptions are key pillars of YouTube's monetization, enabling unique content and communities to flourish (more information on monetization is available at <https://blog.youtube/news-and-events/10-ways-monetize-youtube/>).

Question 15. Aside from disabling ads and the other advertised features of the service, does subscribing to YouTube Premium otherwise alter how content is presented to a user?

We offer YouTube Premium to amplify the user experience on YouTube. With YouTube Premium, subscribers can watch millions of videos without interruptions by ads before and during a video, including video overlay ads. Subscribers also do not see third-party banner ads and search ads, but may still see branding or promotions embedded in the content by the creator and promotional links, shelves and features in and around the content that are added or enabled by the creator. These links, shelves, and features could be for their website, merchandise, a membership to their channel, event tickets, or other related destinations they are promoting.

Written Questions Submitted by Sen. Hirono to Alexandra Veitch:

Question 1. Last fall, Jack Dorsey testified before the full Senate Judiciary Committee that “people should have choices about the key algorithms that affect their experience online,” including the ability to use algorithms created by third parties to rank and filter the content they see.

- a. Do you agree that “people should have choices about the key algorithms that affect their experience online?”**
- b. Does YouTube commit to allowing people to choose the algorithm that recommends videos on the platform, including a choice of algorithms developed by third parties?**

We give our users a number of choices that help form their experience on our platforms. First, recommendations on YouTube help users discover videos they may enjoy, and they help creators reach new viewers and grow their audience across the platform. Recommendations work in two main places: a user’s homepage and the “Up Next” panel. And users have control over their data that is stored and how it is saved and used on our platform. For example, users can access and delete their search and watch history, and any videos removed from their history will no longer influence their recommendations, and queries they have removed will not appear in their search bar as suggestions (more information on user choice concerning their data in YouTube is available at

<https://www.youtube.com/howyoutubeworks/user-settings/privacy/#your-data-in-youtube>).

Our recommendation system constantly evolves, learning every day from over 80 billion pieces of information we call signals. A number of signals build on each other to help inform our dynamic system about what a particular user finds satisfying: clicks, watchtime, survey responses, sharing, likes, and dislikes (more information about these signals is available at <https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/>).

- **Clicks:** Clicking on a video provides a strong indication a user will also find it satisfying. After all, a user would not click on content they do not want to watch.
- **Watchtime:** Watchtime — which videos a user watched and for how long — provides personalized signals to our system about what users most likely want to watch.
- **Survey Responses:** To ensure viewers are satisfied with the content they watch, we measure what we call “valued watchtime” — the time spent watching a video that viewers consider valuable. We measure valued watchtime through user surveys that ask viewers to rate the video watched from one to five stars, giving us a metric to determine how satisfying viewers found the content. If a video is rated one to two stars, we ask the viewer why they gave such a low rating. Similarly, if the viewer gives the video four to five stars, we ask why — was it inspirational or meaningful? Only videos that are rated highly with four or five stars are counted as valued watchtime.

- **Sharing, Likes, Dislikes:** On average, people are more likely to be satisfied by videos that they share or like. Our system uses this information to try to predict the likelihood that viewers will share or like further videos.

In June 2019, we launched new features that give users more control over what recommendations appear on the homepage and in their “Up Next” suggestions (more information is available at <https://blog.youtube/news-and-events/giving-you-more-control-over-homepage/>). Three specific changes were made:

- **Explore topics and related videos on your Homepage and in Up Next videos:** Users can now more easily explore topics and related videos on their Homepage and in Up Next videos to see more of what they love on YouTube. Users now see options based on existing personalized suggestions that are meant to help users find what they are looking for faster.
- **Remove suggestions from channels you don’t want to watch:** Users can easily let us know to stop suggesting videos from a particular channel by tapping the three-dot menu next to a video on the homepage or Up Next, then “Don’t recommend channel.” After that, users should no longer see videos from that channel suggested to them on YouTube.
- **Learn more about why a video may be suggested to you:** Sometimes we recommend videos from channels users have not seen before based on what other viewers with similar interests have liked and watched in the past. When we’re suggesting videos based on this, users now see more information underneath the video in a small box. Our goal is to explain why these videos surface on your homepage in order to help you find videos from new channels you might like.

These features make it easier for users to block channels from recommendations, give users the option to filter recommendations on Home and on “Up Next,” and give users more information about why we are suggesting a video.

In addition to providing choice, we believe that transparency around our processes is important. As an indication of this commitment, we recently launched the YouTube Research Program. The program gives academic researchers access to our raw data in a user privacy sensitive way, allowing them to derive metrics or develop insights of their own based on that data. This increased access includes expanded data quotas, ability to derive analytics, and in-house YouTube support. As we expand this access to our API, we’ll work closely with researchers to gain feedback and improve collaborations. Finally, we provide users with the ability to influence the ads they see online. As explained in the Google Privacy Policy, if you use a Google Account, Google may use information you provide to show you relevant ads. Users can always learn more information concerning personalized advertising and controlling the ads they see on Google’s support pages (see

<https://support.google.com/google-ads/answer/2549116> and <https://support.google.com/accounts/answer/2662856>). In addition, users can turn off ad personalization in their Google Account settings, available at <https://adssettings.google.com/>.

Question 2. Former-WDBJ reporter Alison Parker was murdered live on-the-air in 2015. Not long after, videos of her murder were uploaded to YouTube. Despite the videos violating YouTube’s Community Guidelines regarding “violent or dangerous content” and efforts by Alison’s father Andy Parker, the HONR Network, and others (including my office) to identify these videos to YouTube, many copies of the video remain easily accessible on the platform.

- a. **What specific steps has YouTube taken to ensure that videos of Ms. Parker’s murder are either removed from YouTube or not uploaded in the first place?**
- b. **Please explain YouTube’s process for proactively identifying content that may violate YouTube’s Community Guidelines, including any automated processes employed by YouTube such as hashing or machine learning.**
- c. **Once content has been flagged as possibly violating YouTube’s Community Guidelines—whether through YouTube’s internal processes or through the efforts of users—please explain the process for determining whether the content should be removed or have other action taken.**
- d. **When YouTube becomes aware of its platform being used to spread violent and graphic footage of a particular event, does YouTube take additional steps beyond its normal content moderation practices to remove such footage? If so, please describe those steps in detail.**

Violent footage filmed by perpetrators such as Alison Parker's murderer are not allowed on YouTube, and we are committed to removing any content that violates our policies. Our policies specifically prohibit videos that aim to shock with violence, or accuse victims of public violent events of being part of a hoax. For certain egregious content, like videos containing footage of Alison Parker’s murder, we use digital fingerprinting to detect and remove it from YouTube. We have automatically detected and removed over 400 reuploaded versions of the murderer’s video before they were ever flagged by a user.

[YouTube maintains detailed policies against violent or dangerous content](#)

At YouTube, we develop and maintain Community Guidelines that outline what type of content is not allowed on our platform (available at <https://support.google.com/youtube/answer/9288567?hl=en>). These policies act as our ‘rules of the road’ and we aim to make them clear and easily accessible to all users and content creators. Our Community Guidelines define what content and behaviors are not allowed, and the process by which a piece of content, or its creator, may be removed from service. These

Community Guidelines are one of the ways we balance allowing diverse voices on the platform while limiting harmful content.

Our Community Guidelines are designed to protect the YouTube Community and include a number of policies that protect users from violent or dangerous content. These include policies that address violent extremism, violent or graphic content, harmful or dangerous content, election misinformation and hate speech. They also include policies that prohibit content that praises, incites, or promotes violence, or that which is misleading and deceptive with serious risk of egregious harm (including content that interferes with democratic processes). For example:

- **Harmful or Dangerous Content:** YouTube prohibits content that encourages dangerous or illegal activities that risk serious physical harm or death. This policy applies to various types of conduct, including content that provides instructions to build a bomb meant to injure or kill others, and content that promotes or glorifies violent tragedies, such as school shootings.
- **Firearms:** YouTube prohibits content intended to sell firearms, instruct viewers on how to make firearms, ammunition, and certain accessories, or instruct viewers on how to install those accessories. We also do not allow live streams that show someone holding, handling, or transporting a firearm.
- **Harassment & Cyberbullying:** YouTube prohibits content that threatens individuals or targets an individual with prolonged or malicious insults based on intrinsic attributes. This policy includes, for example, targeting an individual and making claims that they are involved in human trafficking in the context of a harmful conspiracy theory where the conspiracy is linked to direct threats or violent acts.
- **Hate Speech:** YouTube removes content promoting violence or hatred against individuals or groups based on age, caste, disability, ethnicity, gender identity and expression, nationality, race, immigration status, religion, gender, sexual orientation, veteran status, or victims of a major violent event and their kin. This policy also prohibits videos alleging that a group is superior in order to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status. This would include, for example, videos that promote or glorify Nazi ideology, which is inherently discriminatory. We also prohibit content denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place.
- **Violent Criminal Organizations:** YouTube prohibits content intended to praise, promote, or aid violent criminal organizations is not allowed on YouTube. An example of content that violates this policy would be videos or comments directing users to sites hosting manifestos from the perpetrators of violent incidents, such as the Christchurch attacker. This policy also applies to channels and content produced by violent criminal or terrorist organizations, including that:

- praising or memorializing prominent terrorist or criminal figures to encourage others to carry out acts of violence;
 - praising or justifying violent acts carried out by violent criminal or terrorist organizations;
 - aimed at recruiting new members to violent criminal or terrorist organizations designated by the U.S. government;
 - depicting hostages or posted with the intent to solicit, threaten, or intimidate on behalf of a violent criminal or terrorist organization; and
 - depicting the insignia, logos, or symbols of violent criminal or terrorist organizations to praise or promote them.
- **Violent or Graphic Content:** YouTube prohibits violent or gory content intended to shock or disgust viewers, or content encouraging others to commit violent acts. This includes, but is not limited to, content that incites others to commit violent acts against individuals or a defined group of people; fights involving minors; and content where animals are encouraged or coerced to fight by humans. It also applies to footage, audio, or imagery involving road accidents, natural disasters, war aftermath, terrorist attack aftermath, street fights, physical attacks, sexual assaults, immolation, torture, corpses, protests or riots, robberies, medical procedures, or other such scenarios with the intent to shock or disgust viewers. Further, footage or imagery showing bodily fluids, such as blood or vomit, with the intent to shock or disgust viewers is prohibited by this policy.

Content that violates our policies against violent extremism, detailed above, includes material produced by government-listed foreign terrorist organizations. We do not permit these terrorist organizations to use YouTube for any purpose, including recruitment. YouTube also strictly prohibits content that promotes terrorism, such as content that glorifies terrorist acts or incites violence. Content produced by violent extremist groups that are not government-listed foreign terrorist organizations is subject to our policies, including those described above, regardless of the uploader or the group to which they belong.

Finally, we regularly review and update our policies to address new and emerging threats. In 2017, we introduced a tougher stance toward videos with supremacist content, including limiting recommendations and features like comments and the ability to share the video. In June 2019, we clarified our hate speech policy to specifically prohibit videos alleging that a group is superior to justify discrimination, segregation or exclusion based on qualities like age, gender, race, caste, religion, sexual orientation or veteran status. In October 2020, we expanded our hate and harassment policies to prohibit content that targets an individual or group with conspiracy theories that have been used to justify real-world violence, such as QAnon.

Due to the evolving nature and shifting tactics of groups promoting these conspiracy theories, we will continue to adapt our policies to stay current and remain committed to taking the steps needed to live up to this responsibility.

YouTube has robust methods to enforce its policies regarding violent or dangerous content

We rely on a combination of people and technology to enforce our policies. YouTube has invested significantly in our automated detection systems and our engineering teams continue to update and improve them month by month. In 2017, we significantly expanded our use of machine learning technology to help detect potentially violative content and send it for human review at scale. Machine learning is well-suited to detect patterns and find content similar (but not identical) to other content we've already removed, even before it's ever viewed. Automated flagging also allows us to identify and act more quickly and accurately to remove violative content, lessening both the burden on human reviewers and the time it takes to remove such content. For example, in Q2 2022, more than 93% of videos that we removed as violative were first flagged through our automated systems.

Our machine learning systems help our human review teams remove content at scale, with the speed and volume that could not be achieved with people alone. Recently, we removed nearly 73,000 videos or promotion of violence and violent extremism between April and June 2022. We also use hashes (or "digital fingerprints") to automatically identify copies of known violative content before they are ever made available for viewing. These systems automatically remove content only where there is high confidence of a policy violation and the other content is flagged for human review. YouTube has invested significantly in these human reviewers, who remain essential to both removing content and training machine learning systems to help make contextualized decisions on content.

Additionally, we release a quarterly Community Guidelines Enforcement Report (available at <https://transparencyreport.google.com/youtube-policy/removals>) that shows the progress we are making in removing violative content from our platform. In Q2 2022 alone, we removed more than 4.4 million videos for violating our Community Guidelines. More than 72% of those first flagged by our systems received 10 or fewer views and more than 93% were detected by our automated flagging system.

Written Questions Submitted by Sen. Coons to Alexandra Veitch:

Question 1. I appreciate your testimony that YouTube is focused on long-term usage of its products rather than prioritizing short-term engagement. Some have raised concern, however, about whether the goals and metrics that social media platforms use to measure performance and pay for product teams may create structural incentives within the company potentially at odds with this long-term interest.

- a. Please enumerate the metrics that YouTube uses to set goals for its product teams—particularly the team responsible for recommending videos—for purposes of performance and pay.**
- b. Please explain why YouTube chose those metrics.**
- c. Does YouTube take any integrity-related metrics into account for its performance and pay assessments for any of its teams, including those focused on trust, safety, and integrity?**

Please refer to the answer to Question Number 2, below.

Question 2. Some have raised concern about whether social media companies are internally organized to ensure that trust and safety teams have adequate voice, staffing, and resources in product development.

- a. What teams are responsible for ensuring that the recommendation algorithm does not promote violating content? Please elaborate on how this process is structured and coordinated internally.**
- b. To the extent there are tensions between topline goal metrics and integrity metrics in assessing business units or product features, how do you resolve them, and who makes those decisions?**

Because the answers to these questions are related, we have grouped together our responses to Question Numbers 1 and 2.

Our first priority is to protect our users. Content moderation at Google and YouTube is primarily managed by Trust and Safety teams across the company. These teams are made up of engineers, content reviewers, and others who work across Google to address content that violates any of our policies. They also work with our legal and public policy teams, oversee the people we hire to help us scale our content moderation efforts, and provide the native language expertise and the 24-hour coverage required of a global platform. Google employs review teams across many offices globally and across the U.S. to ensure we have a diverse set of reviewers who are reviewing publisher sites, apps, and content.

We do not have to choose between safety and growing our business. In fact, our business depends on our providing a safe environment for users and advertisers. How YouTube lives up

to its publicly declared goals is a core part of the product, and much of what we do on the product team is to deliver on Trust and Safety objectives and live up to YouTube's responsibilities.

As we consider the YouTube product experience, we must take into account how our content policies impact our viewers and our creators. The Community Guidelines act as the "rules of the road" for what is and is not allowed, and we apply them consistently across YouTube. This means the policies and how they work are a core part of the overall product experience for our entire ecosystem — including our viewers, creators, and advertisers. Our responsibility relies on using our systems to connect viewers to high-quality information and to minimize the chances they'll see problematic content. It also involves assessing how creators—the vast majority of whom are looking to do the right thing — are rewarded. These levers work together to form a comprehensive approach to responsibility.

We prioritize responsibility even when it comes at a cost. Our advertiser-friendly content guidelines and the YouTube Partner Program set out YouTube's revenue sharing program, and if a creator fails to comply with those guidelines, we cut them off from making money on the platform. When we do that, YouTube foregoes revenue, as well.

Question 3. In a response to questions for the record I posed last November, Facebook said that it has been building a "parallel viral content review system to flag posts that may be going viral—no matter what type of content it is" and prioritize these posts for review. The idea is that potential virality should be a signal that content would be particularly dangerous if it violates platform policies, and thus warrants a trigger for human review.

- a. Does YouTube incorporate potential virality in reviewing content?
- b. Please explain the factors YouTube considers when determining which content to review or fact-check.
- c. How long does it take on average for YouTube videos to be fact-checked?
- d. What is YouTube's process for providing corrections when it learns someone watched a video that later proved to be misinformation or violative?

YouTube's Community Guidelines outline what content is permissible and what content is violative on our platform. As noted above in our responses to Questions 1 and 2, these policies act as our "rules of the road" of what is and is not allowed, and we enforce our policies globally, independent of viewpoints or audience size.

For a number of years, the misinformation landscape online was dominated by a few main narratives — such as 9/11 "truthers," moon landing conspiracy theorists, and so-called "flat earthers." We were able to train our machine learning systems to reduce recommendations of those videos and other similar ones based on patterns in that type of content. But increasingly, a completely new narrative can quickly crop up and gain views. Or, narratives can slide from

one topic to another — for example, some general wellness content can lead to vaccine hesitancy. Each narrative can also look and propagate differently, and at times, even be hyperlocal.

We faced these challenges early on in the COVID-19 pandemic, such as when a conspiracy theory that 5G towers caused the spread of coronavirus led to people burning down cell towers in the UK. Due to the clear risk of real-world harm, we responded by updating our guidelines and making this type of content violative. In this case we could move quickly because we already had policies in place for COVID-19 misinformation based on local and global health authority guidance as well as our longstanding policies against inciting violence.

But not every fast-moving narrative in the future will have expert guidance that can inform our policies. And the fresher the misinformation, the fewer examples we have to train our systems. To address this, we continuously train our systems on new data.

In addition to reducing the spread of some content, our systems connect viewers to authoritative videos in search results and recommendations. For major news events that are developing, like a natural disaster, we surface developing news panels to point viewers to text articles. We also surface, in certain countries, fact checks when you search for a specific claim, or contextual information panels when you search for or watch videos related to topics prone to misinformation. We give viewers this context to help them make their own informed decisions about the videos they are watching. When viewers search YouTube for something related to a specific claim, sometimes we provide an information panel that includes a fact check from an independent third-party publisher. These information panels indicate whether claims related to your search are true, false, or something else like "partly true," according to the publisher's fact check. More information about our fact checking is available at <https://support.google.com/youtube/answer/9229632>, and more information about our information panels is available at <https://support.google.com/youtube/answer/9229632>.

We strive to remove violative content as quickly as possible. Further, as detailed in the YouTube Community Guidelines Enforcement Transparency Report (available at <https://transparencyreport.google.com/youtube-policy/removals>), in the second quarter of 2022, more than 4.4 million videos were removed from YouTube for violating our Community Guidelines — 93% of which were first flagged by machines rather than humans. Of those detected by machines, over 34% never received a single view, and just over 72% received fewer than 10 views.

Question 4. I appreciate YouTube’s new reporting of a “violative view rate,” which is an important estimation of what is not being caught and thus being viewed by users. To what extent can YouTube calculate this metric at a level of granularity that allows it to drive decision making about product features? For instance, can YouTube estimate this rate as to certain regions, or use it in potential A/B testing of products?

As discussed in our answers to Questions 1 and 2, in April 2021 we released a new metric called Violative View Rate (VVR) as part of our Community Guidelines Enforcement Report. It is now the primary metric used to measure our responsibility work across YouTube. We will continue to publish this metric so we can be held accountable to reducing it over time.

As noted in the responses above, we calculate VVR by taking a sample of videos on YouTube and sending it to our content reviewers who tell us which videos violate our policies and which do not. By sampling, we gain a more comprehensive view of the violative content we might not be catching with our systems.

Question 5. I appreciate your testimony that YouTube is committed to increasing transparency about your algorithmic recommendation engines. Many independent researchers have said it is essential to know not only what “signals” are considered but also how those signals are weighted; otherwise, it is not apparent which signals are actually driving the ranking.

- a. Please provide more information about the inputs, signals and weightings used in YouTube’s recommendation algorithm.
- b. Will YouTube provide greater transparency going forward into its weightings of different inputs for its recommendation algorithm, whether through reporting, granting access to independent researchers, or conducting an audit?

Our approach to recommendations is anchored in transparency and user choice and control. We have a number of resources available and are working on ways to make our systems easier to understand. We want to offer meaningful transparency to policymakers and users. We have an informative “How YouTube Works” site (available at <https://www.youtube.com/howyoutubeworks/>) with information about how we raise results and make improvements. We also provide significant resources to creators and work hard to inform creators about the ways our systems work and how to be successful on YouTube. This includes interactive websites, detailed Help Centers, videos, frequent blog posts, and user forums (examples of which are available at <https://www.youtube.com/channel/UckRfArvrzheW2E7b6SVT7vQ> and <https://support.google.com/youtube/answer/9002587>).

It is important to note that our recommendation system does not operate off of a “recipe book” of what to do. It constantly evolves, learning every day from over 80 billion pieces of

information we call signals. Providing more transparency is not as simple as listing a formula for recommendations, but involves understanding all the data that feeds into our system. A number of signals build on each other to help inform our dynamic system about what you find satisfying: clicks, watchtime, survey responses, sharing, likes, and dislikes (more information about these signals is available at <https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/>).

- **Clicks:** Clicking on a video provides a strong indication that users will also find it satisfying. After all, a user would not click on content they do not want to watch.
- **Watchtime:** Watchtime — which videos a user watched and for how long — provides personalized signals to our system about what users most likely want to watch.
- **Survey Responses:** To ensure viewers are satisfied with the content they watch, we measure what we call “valued watchtime” — the time spent watching a video that viewers consider valuable. We measure valued watchtime through user surveys that ask viewers to rate the video watched from one to five stars, giving us a metric to determine how satisfying viewers found the content. If a video is rated one to two stars, we ask the viewer why they gave such a low rating. Similarly, if the viewer gives the video four to five stars, we ask why — was it inspirational or meaningful? Only videos that are rated highly with four or five stars are counted as valued watchtime.
- **Sharing, Likes, Dislikes:** On average, people are more likely to be satisfied by videos that they share or like. Our system uses this information to try to predict the likelihood that viewers will share or like further videos.

We also empower our users by giving them control over personalized recommendations. As described in our response to Question 1, in June 2019, we launched new features that give users more control over what recommendations appear on the homepage and in their “Up Next” suggestions. These features make it easier for users to block channels from recommendations and provide users more information about why we are suggesting a video.

Question 6. In general, YouTube does not disclose data to the public regarding how many times any video has been recommended by its recommendation algorithm. This means that when YouTube removes videos from its platform for violating its content policies, the public has no way of knowing how many times those videos were recommended before their removal.

- a. During last Tuesday’s hearing, I asked whether YouTube could commit to disclosing this information publicly. You said this was an “interesting idea” and agreed to follow up. Can YouTube now commit to disclosing this information? If not, why not?
- b. In general, does YouTube collect data regarding the number of times that videos on the platform have been recommended by its algorithm?
- c. Does YouTube currently share that data with anyone outside the company?

YouTube has been a leader in transparency concerning how we operate our platform and when we enforce our Community Guidelines. We publish a quarterly YouTube Community Guidelines Enforcement Transparency Report (available at <https://transparencyreport.google.com/youtube-policy/removals>). As noted in some of our other answers, we strive to remove violative content as quickly as possible.

Our commitment to openness means we may have “borderline” content available on the platform that comes close to violating our policies but does not quite cross the line. It is important to protect freedom of expression, but we set a high bar for what videos we display prominently in our recommendations on the YouTube homepage or through the “Up Next” panel. We constantly work to improve our systems to reduce recommendations on borderline content and videos that could misinform users in a potentially harmful way, particularly in areas that rely on veracity — such as news, medicine, historical events, or science.

To determine borderline content, evaluators assess factors that include, but are not limited to, whether the content is: inaccurate, misleading or deceptive; insensitive or intolerant; and harmful or with the potential to cause harm. The results are combined to give a score for how likely it is that the video contains harmful misinformation or is borderline. Based on the consensus input from the evaluators, we use well-tested machine learning systems to build models. These models help review hundreds of thousands of hours of videos every day in order to find and limit the spread of borderline content. And over time, the accuracy of these systems will continue to improve. Today, consumption of borderline content that comes from our recommendations represents less than 1% of the content watched on YouTube. We have now launched changes to our recommendation system to reduce recommendations of borderline content in every market where we operate.

To encourage further research, we recently launched the YouTube Research Program, which gives academic researchers access to our data in a user privacy sensitive way, allowing them to derive metrics or develop insights of their own based on that data (more information is available at <https://research.youtube/>).

Question 7. During last Tuesday’s hearing, I asked whether YouTube generally requires employees to sign non-disclosure agreements (NDAs). You said you believe that YouTube has “standard agreements to protect proprietary information.”

- a. Please briefly describe the scope of the NDAs that YouTube employees are required to sign, including the types of information that employees and former employees are barred from disclosing.
- b. Do the NDAs cover more than trade secrets, private user data, and/or technical information that employees learn while employed by YouTube?

- c. Do the NDAs or any other contractual agreements bar employees from criticizing or disparaging YouTube after they leave the company?
- d. If an employee disagrees with a policy decision made by YouTube during the employee's period of employment, would the employee's NDA prohibit him or her from discussing that disagreement publicly after leaving the company?
- e. Do YouTube's NDAs contain binding arbitration clauses—meaning employees must submit to arbitration in the event of any legal dispute under the terms of the agreement?
- f. Has YouTube ever threatened legal action against an employee (current or former) based on potential violations of that employee's NDA?

In 2020, Google and Alphabet updated their workplace codes of conduct consistent with the following five guiding principles:

- **Commitment:** We will double down on our commitment to a respectful, safe, and inclusive working environment for all employees and members of our extended workforce.
- **Care:** We will continue to create an environment that emphasizes respect for each individual at all levels of the organization.
- **Transparency:** We will be open and transparent as an organization regarding the frequency of complaints of misconduct and our approach to investigating and responding to those allegations.
- **Fairness & Consistency:** We will make sure that everyone is treated respectfully, fairly, and compassionately, and we'll apply our employment policies consistently.
- **Accountability:** We will hold everyone responsible for their actions.

Google's employment agreement, which includes a NDA, does not prohibit the disclosure of facts underlying claims of harassment or discrimination. Specifically, its employment agreement provides that "nothing in this Agreement limits any right I may have to discuss terms, wages, and working conditions of employment, as protected by applicable law."

In addition, Google has made arbitration optional for all employees, temporary staff, vendors, and independent contractors with respect to harassment, discrimination, and retaliation disputes with Alphabet. It has also committed to annually review its sexual harassment and retaliation policies and practices to ensure consistency with the guiding principles.

Question 8. You stated in your written testimony that YouTube “disincentivize[s] creators seeking to profit off of misinformation by blocking ads on their content and suspending repeat offenders from our monetization program.” How does YouTube determine whether a content creator is a “repeat offender,” particularly with regard to misinformation?

- a. Does YouTube ever take public perception into consideration when determining whether or how to enforce its repeat offender policies? If so, how?**
- b. Are potential political repercussions ever taken into consideration when determining whether or how to enforce YouTube’s repeat offender policies? If so, how?**
- c. Are potential business consequences—such as potential dissatisfaction from advertisers—ever taken into consideration when determining whether or how to enforce YouTube’s repeat offender policies? If so, how?**

We believe strongly in the freedom of expression and access to information, and we know that the overwhelming majority of creators follow our guidelines and understand they are part of a large, influential, and interconnected community. However, we also know that we have a responsibility to protect our users, which is why we have policies prohibiting hate speech, terrorist content, and other content that violates our policies, as well as stricter standards for who can monetize their content. Each of the products and services we offer has a different purpose, and we tailor our approach carefully to the content that should be available on each product and service. However, our search and recommendation systems are not designed to filter or demote videos or channels based on specific political perspectives. We take extraordinary lengths to build our products and enforce our policies impartially.

While YouTube creates a space for ideas and expression, it is not a free-for-all. For example, it is a violation of YouTube’s hate speech policy for users to post videos that promote violence against particular ethnic or religious groups (for more information, please see <https://support.google.com/youtube/answer/2801939>). As described in our response to Question No. 2, creators who violate those rules may have their content removed or their accounts terminated. When we detect a video that violates our Community Guidelines, we remove the video and apply a strike to the channel. The strike restricts a creator’s ability to post or create content on the platform for one week. If the creator’s behavior warrants another strike within 90 days from the first, a new two-week prohibition from posting or creating content is implemented. A third strike within 90 days results in permanent removal of a channel from YouTube. Creators can appeal those strikes if they believe we are mistaken. We also terminate entire channels if they are dedicated to posting content prohibited by our Community Guidelines or contain a single egregious violation, like child sexual abuse material.

As referenced in our response to Questions 1 and 2, we reward our trusted creators through monetization programs. Creators who earn money on YouTube must follow YouTube’s channel

monetization policies (available at <https://support.google.com/youtube/answer/1311392>), which include YouTube's Community Guidelines, Terms of Service, Copyright, and Google AdSense program policies. Violation of our YouTube channel monetization policies may result in monetization being suspended or permanently disabled. Additionally, creators who follow our policies can apply to join the YouTube Partner Program, a program that sets a higher bar for which channels can make money on our site (for more information on our Partner Program, please see <https://support.google.com/youtube/answer/72851>).

