S. HRG. 116–572

# OPTIMIZING FOR ENGAGEMENT: UNDERSTANDING THE USE OF PERSUASIVE TECHNOLOGY ON INTERNET PLATFORMS

## HEARING

BEFORE THE

SUBCOMMITTEE ON COMMUNICATIONS, TECHNOLOGY, INNOVATION AND THE INTERNET

OF THE

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION
UNITED STATES SENATE

ONE HUNDRED SIXTEENTH CONGRESS

FIRST SESSION

JUNE 25, 2019

Printed for the use of the Committee on Commerce, Science, and Transportation

Available online: http://www.govinfo.gov

U.S. GOVERNMENT PUBLISHING OFFICE
52–609 PDF WASHINGTON : 2024

SENATE COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION

ONE HUNDRED SIXTEENTH CONGRESS

FIRST SESSION

ROGER WICKER, Mississippi, *Chairman*

JOHN THUNE, South Dakota
ROY BLUNT, Missouri
TED CRUZ, Texas
DEB FISCHER, Nebraska
JERRY MORAN, Kansas
DAN SULLIVAN, Alaska
CORY GARDNER, Colorado
MARSHA BLACKBURN, Tennessee
SHELLEY MOORE CAPITO, West Virginia
MIKE LEE, Utah
RON JOHNSON, Wisconsin
TODD YOUNG, Indiana
RICK SCOTT, Florida

MARIA CANTWELL, Washington, *Ranking*
AMY KLOBUCHAR, Minnesota
RICHARD BLUMENTHAL, Connecticut
BRIAN SCHATZ, Hawaii
EDWARD MARKEY, Massachusetts
TOM UDALL, New Mexico
GARY PETERS, Michigan
TAMMY BALDWIN, Wisconsin
TAMMY DUCKWORTH, Illinois
JON TESTER, Montana
KYRSTEN SINEMA, Arizona
JACKY ROSEN, Nevada

JOHN KEAST, *Staff Director*
CRYSTAL TULLY, *Deputy Staff Director*
STEVEN WALL, *General Counsel*
KIM LIPSKY, *Democratic Staff Director*
CHRIS DAY, *Democratic Deputy Staff Director*
RENAE BLACK, *Senior Counsel*

---

SUBCOMMITTEE ON COMMUNICATIONS, TECHNOLOGY, INNOVATION AND THE INTERNET

JOHN THUNE, South Dakota, *Chairman*
ROY BLUNT, Missouri
TED CRUZ, Texas
DEB FISCHER, Nebraska
JERRY MORAN, Kansas
DAN SULLIVAN, Alaska
CORY GARDNER, Colorado
MARSHA BLACKBURN, Tennessee
SHELLEY MOORE CAPITO, West Virginia
MIKE LEE, Utah
RON JOHNSON, Wisconsin
TODD YOUNG, Indiana
RICK SCOTT, Florida

BRIAN SCHATZ, Hawaii, *Ranking*
AMY KLOBUCHAR, Minnesota
RICHARD BLUMENTHAL, Connecticut
EDWARD MARKEY, Massachusetts
TOM UDALL, New Mexico
GARY PETERS, Michigan
TAMMY BALDWIN, Wisconsin
TAMMY DUCKWORTH, Illinois
JON TESTER, Montana
KYRSTEN SINEMA, Arizona
JACKY ROSEN, Nevada

# CONTENTS



WITNESSES



APPENDIX

# OPTIMIZING FOR ENGAGEMENT: UNDERSTANDING THE USE OF PERSUASIVE TECHNOLOGY ON INTERNET PLATFORMS

**TUESDAY, JUNE 25, 2019**

U.S. SENATE,
SUBCOMMITTEE ON COMMUNICATIONS, INNOVATION, AND THE INTERNET,
COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION,
*Washington, DC.*

The Subcommittee met, pursuant to notice, at 10:05 a.m. in room SH–216, Hart Senate Office Building, Hon. John Thune, Chairman of the Subcommittee, presiding.

Present: Senators Thune [presiding], Schatz, Fischer, Blumenthal, Blackburn, Peters, Johnson, Tester, Rosen, Udall, Sullivan, Markey, Young, and Cruz.

## OPENING STATEMENT OF HON. JOHN THUNE, U.S. SENATOR FROM SOUTH DAKOTA

Senator THUNE. I want to thank everyone for being here today to Examine the Use of Persuasive Technologies on Internet Platforms.

Each of our witnesses today has a great deal of expertise with respect to the use of artificial intelligence and algorithms more broadly as well as in the more narrow context of engagement and persuasion and brings unique perspectives to these matters.

Your participation in this important hearing is appreciated, particularly as this Committee continues to work on drafting data privacy legislation.

I convened this hearing in part to inform legislation that I'm developing that would require Internet platforms to give consumers the option to engage with the platform without having the experience shaped by algorithms driven by users' specific data.

Internet platforms have transformed the way we communicate and interact and they have made incredibly positive impacts on society in ways that are too numerous to count.

The vast majority of content on these platforms is innocuous and at its best, it is entertaining, educational, and beneficial to the public. However, the powerful mechanisms behind these platforms meant to enhance engagement also have the ability or at least potential to influence the thoughts and behaviors of literally billions of people.

As one reason why there's widespread unease about the power of these platforms and why it is important for the public to better un-

derstand how these platforms use artificial intelligence and opaque algorithms to make inferences from the reams of data about us that affect behavior and influence outcomes.

Without safeguards, such as real transparency, there is a risk that some Internet platforms will seek to optimize engagement to benefit their own interests and not necessarily to benefit the consumers' interests.

In 2013, former Google Executive Chairman Eric Schmidt wrote that modern technology platforms, and I quote, "are even more powerful than most people realize and our future will be profoundly altered by their adoption and successfulness in societies everywhere."

Since that time, algorithms and artificial intelligence have rapidly become an important part of our lives, largely without us even realizing it. As online content continues to grow, large technology companies rely increasingly on AI-powered automation to select and display content that will optimize engagement.

Unfortunately, the use of artificial intelligence and algorithms to optimize engagement can have an unintended and possibly even dangerous downside. In April, Bloomberg reported that YouTube has spent years chasing engagement while ignoring internal calls to address toxic videos, such as vaccination conspiracies and disturbing content aimed at children.

Earlier this month, *New York Times* reported that YouTube's automated recommendation system was found to be automatically playing a video of children playing in their backyard pool to other users who had watched sexually themed content. That is truly troubling and it indicates the real risks in a system that relies on algorithms and artificial intelligence to optimize for engagement.

And these are not isolated examples. For instance, some have suggested that the so-called "filter level" created by social media platforms like Facebook may contribute to our political polarization by encapsulating users within their own comfort zones or echo chambers.

Congress has a role to play in ensuring companies have the freedom to innovate but in a way that keeps consumers' interests and well-being at the forefront of their progress.

While there must be a healthy dose of personal responsibility when users participate in seemingly free online services, companies should also provide greater transparency about how exactly the content we see is being filtered. Consumers should have the option to engage with the platform without being manipulated by algorithms powered by their own personal data, especially if those algorithms are opaque to the average user.

We are convening this hearing in part to examine whether algorithmic explanation and transparency are policy options that Congress should be considering.

Ultimately, my hope is that at this hearing today, we are able to better understand how Internet platforms use algorithms, artificial intelligence, and machine learning to influence outcomes, and we have a very distinguished panel before us.

Today, we are joined by Tristan Harris, the Co-Founder of the Center for Humane Technology; Ms. Maggie Stanphill, the Director of Google User Experience; Dr. Stephen Wolfram, Founder of Wol-

fram Research; and Ms. Rashida Richardson, the Director of Policy Research at the AI Now Institute.

Thank you all again for participating on this important topic.

I want to recognize Senator Schatz for any opening remarks that he may have.

**STATEMENT OF HON. BRIAN SCHATZ,**
**U.S. SENATOR FROM HAWAII**

Senator SCHATZ. Thank you, Mr. Chairman.

Social media and other Internet platforms make their money by keeping users engaged and so they've hired the greatest engineering and tech minds to get users to stay longer inside of their apps and on their websites.

They've discovered that one way to keep us all hooked is to use algorithms that feed us a constant stream of increasingly more extreme and inflammatory content and this content is pushed out with very little transparency or oversight by humans. This set-up and also basic human psychology makes us vulnerable to lies, hoaxes, and mis-information.

The *Wall Street Journal* investigation last year found that YouTube's recommendation engine often leads users to conspiracy theories, partisan viewpoints, and misleading videos, even when users aren't seeking out that kind of content, and we saw that YouTube's algorithms were recommending videos of children after users watched sexualized content that did not involve children, and this isn't just a YouTube problem.

We saw all of the biggest platforms struggle to contain the spread of videos of the Christchurch massacre and its anti-Muslim propaganda. The shooting was live-streamed on Facebook and over a million copies were uploaded across platforms. Many people reported seeing it on auto-play on their social media feeds and not realizing what it was. And just last month, we saw a fake video of the Speaker of the House go viral.

I want to thank the Chairman for holding this hearing because as these examples illustrate, something is really wrong here and I think it's this: Silicon Valley has a premise. It's that society would be better, more efficient, smarter, more frictionless if we would just eliminate steps that include human judgment, but if YouTube, Facebook, or Twitter employees rather than computers were making the recommendations, would they have recommended these awful videos in the first place?

Now I'm not saying that employees need to make every little decision, but companies are letting algorithms run wild and only using humans to clean up the mess. Algorithms are amoral. Companies design them to optimize for engagement as their highest priority and by doing so, they eliminated human judgment as part of their business models.

As algorithms take on an increasingly important role, we need for them to be more transparent and companies need to be more accountable for the outcomes that they produce. Imagine a world where pharmaceutical companies were not responsible for the long-term impacts of their medicine and we couldn't test their efficacy or if engineers were not responsible for the safety of the structure they designed and we couldn't review the blueprints.

We are missing that kind of accountability for Internet platform companies. Right now, all we have are representative sample sets, data scraping, and anecdotal evidence. These are useful tools, but they are inadequate for the rigorous systemic studies that we need about the societal effects of algorithms.

These are conversations worth having because of the significant influence that algorithms have on people's daily lives, and this is a policy issue that will only grow more important as technology continues to advance.

And so thank you, Mr. Chairman, for holding this hearing and I look forward to hearing from our experts.

Senator THUNE. Thank you, Senator Schatz.

We do, as I said, have a great panel to hear from today, and we're going to start on my left and your right with Mr. Tristan Harris, who's Co-Founder and Executive Director of Center for the Humane Technology; Ms. Maggie Stanphill, as I mentioned, who's the Google User Experience Director at Google, Inc.; Dr. Stephen Wolfram, who's the Founder and Chief Executive Officer of Wolfram Research; and Ms. Rashida Richardson, Director of Policy Research at AI Now Institute.

So if you would, confine your oral remarks to as close as five minutes as possible. It will give us an opportunity to maximize the chance for Members to ask questions. But thank you all for being here. We look forward to hearing from you.

Mr. Harris.

## STATEMENT OF TRISTAN HARRIS, EXECUTIVE DIRECTOR, CENTER FOR HUMANE TECHNOLOGY

Mr. HARRIS. Thank you, Senator Thune and Senator Schatz.

Everything you said, it's sad to me because it's happening not by accident but by design because the business model is to keep people engaged. Which, in other words, this hearing is about persuasive technology and persuasion is about an invisible asymmetry of power.

When I was a kid, I was a magician and magic teaches you that, you know, you can have asymmetric power without the other person realizing it. You can masquerade to have asymmetric power while looking like you have an equal relationship. You say "pick a card, any card," meanwhile, you know exactly how to get that person to pick the card that you want, and essentially what we're experiencing with technology is an increasing asymmetry of power that has been masquerading itself as a equal or contractual relationship where the responsibility is on us.

So let's walk through why that's happening. In the race for attention because there's only so much attention, companies have to get more of it by being more and more aggressive. I call it the race to the bottom of the brain stem.

So it starts with techniques—like pull to refresh. So you pull to refresh your newsfeed. That operates like a slot machine. It has the same kind of addictive qualities that keep people in Las Vegas hooked to the slot machine.

Other examples are moving stopping cues. So if I take the bottom out of this glass and I keep refilling the water or the wine, you won't know when to stop drinking. So that's what happens with in-

finitely strolling feeds. We naturally remove the stopping cues and this is what keeps people scrolling.

But the race for attention has to get more and more aggressive and so it's not enough just to get your behavior and predict what will take your behavior, we have to predict how to keep you hooked in a different way, and so it crawls deeper down the brain stem into our social validations.

That was the introduction of likes and followers. How many followers do I have? That got every—it was much cheaper, instead of getting your attention, to get you addicted to getting attention from other people and this has created the kind of mass narcissism and mass cultural thing that's happening with young people especially today. And after two decades in decline, the mental health of 10-to-14-year-old girls has actually shot up 170 percent in the last 8 years and this has been very characteristically the cause of social media.

And in the race for attention, it's not enough just to get people addicted to attention, the race has to migrate to AI. Who can build a better predictive model of your behavior? And so if you give an example of YouTube, so there you are, you're about to hit play on a YouTube video and you hit play and then you think you're going to watch this one video and then you wake up 2 hours later and say, "oh, my God, what just happened," And the answer is because you had a super computer pointed at your brain and the moment you hit play, it wakes up an avatar voodoo doll-like version of you inside of a Google server and that avatar, based on all the clicks and likes and everything you ever made, those are like your hair clippings and toenail clippings and nail filings that make the avatar look and act more and more like you, so that inside of a Google server they can simulate more and more possibilities if I prick you with this video, if I prick you with this video, how long would you stay and the business model is simply what maximizes watch time.

This leads to the kind of algorithmic extremism that you've pointed out and this is what's caused 70 percent of YouTube's traffic now to be driven by recommendations, not by human choice but by the machines, and it's a race between Facebook's voodoo doll where you flick your finger can they predict what to show you next and Google's voodoo doll, and these are abstract metaphors that apply to the whole tech industry—where it's a race between who can better predict your behavior.

Facebook has something called Loyalty Prediction, where they can actually predict to an advertiser when you're about to become disloyal to a brand. So, if you're a mother and you take Pampers diapers, they can tell Pampers, "hey, this user's about to become disloyal to this brand."

So, in other words, they can predict things about us that we don't know about our own selves and that's a new level of asymmetric power. And we have a name for this asymmetric relationship which is a fiduciary relationship or a duty of care relationship. The same standard we apply to doctors to priests, to lawyers.

Imagine a world in which priests only make their money by selling access to the confession booth to someone else, except in this case Facebook listens to two billion people's confessions, has a super computer next to them and is calculating and predicting con-

fessions you're going to make before you know you're going to make them and that's what's causing all this havoc.

So, I'd love to talk about more of these things later. I just want to finish up by saying this affects everyone, even if you don't use these products. You still send your kids to a school where other people believing in the anti-vaccine conspiracy theories causes impact for your life or other people voting in your elections and when Mark Andreesen said in 2011 that "software is going to eat the world," and what he meant by that, Mark Andreesen was the founder of Netscape, what he meant by that was that software can do every part of society more efficiently than non-software, right, because its' just adding efficiencies.

So we're going to allow software to eat up our elections, we're going to allow it to eat up our media, our taxi, our transportation, and the problem was that software was eating the world without taking responsibility for it.

We used to have rules and standards around Saturday morning cartoons and when YouTube gobbles up that part of society, it just takes away all of those protections.

I want to finish up by saying that I know Mr. Rogers, Fred Rogers testified before this Committee 50 years ago concerned about the animated bombardment that we were showing children. I think he would be horrified today about what we're doing now and at that same time, he was able to talk to the Committee and that Committee made a choice differently. So, I'm hoping we can talk more about that today.

Thank you.

[The prepared statement of Mr. Harris follows:]

PREPARED STATEMENT OF TRISTAN HARRIS, EXECUTIVE DIRECTOR, CENTER FOR HUMAN TECHNOLOGY

Good morning.

I want to argue today that persuasive technology is a massively underestimated and powerful force shaping the world and that it has taken control of the pen of human history and will drive us to catastrophe if we don't take it back. Because technology shapes where 2 billion people place their attention on a daily basis shaping what we believe is true, our relationships, our social comparison and the development of children. I'm excited to be here with you because you are actually in a position to change this.

Let's talk about how we got here. While we often worried about the point at which technology's asymmetric power would overwhelm *human strengths* and take our jobs, we missed this earlier point when technology hacks *human weaknesses.* And that's all it takes to gain control. That's what persuasive technology does. I first learned this lesson as a magician as a kid, because in magic, you don't have to know more than your audience's intelligence—their PhD in astrophysics—you just have to know their weaknesses.

Later in college, I studied at the Stanford Persuasive Technology Lab with the founders of Instagram, and learned about the ways technology can influence people's attitudes, beliefs and behaviors.

At Google, I was a design ethicist where I thought about how do you ethically wield this influence over 2 billion people's thoughts. Because in an attention economy, there's only so much attention and the advertising business model always wants more. So, it becomes a race to the bottom of the brain stem. Each time technology companies go lower into the brain stem, it takes a little more control of society. It starts small. First to get your attention, I add slot machine "pull to refresh" rewards which create little addictions. I remove stopping cues for "infinite scroll" so your mind forgets when to do something else. But then that's not enough. As attention gets more competitive, we have to crawl deeper down the brainstem to your identity and get you addicted to getting attention from other people. By adding the

number of followers and likes, technology hacks our social validation and now people are obsessed with the constant feedback they get from others. This helped fuel a mental health crisis for teenagers. And the next step of the attention economy is to compete on algorithms. Instead of splitting the atom, it splits our nervous system by *calculating* the perfect thing that will keep us there longer– the perfect YouTube video to autoplay or news feed post to show next. Now technology analyzes everything we've done to create an avatar, voodoo doll simulations of us. With more than a billion hours watched daily, it takes control of what we believe, while discriminating against our civility, our shared truth, and our calm.

As this progression continues the asymmetry only grows until you get deep fakes which are checkmate on the limits of the human mind and the basis of our trust.

But, all these problems are connected because they represent a growing asymmetry between the power of technology and human weaknesses, that's taking control of more and more of society.

The harms that emerge are not separate. They are part of an interconnected system of compounding harms that we call "human downgrading". How can we solve the world's most urgent problems if we've downgraded our attention spans, downgraded our capacity for complexity and nuance, downgraded our shared truth, downgraded our beliefs into conspiracy theory thinking that we can't construct shared agendas to solve our problems? This is destroying our sensemaking at a time we need it the most. And the reason why I'm here is because every day it's incentivized to get worse.

We have to name the cause which is an *increasing asymmetry between the power of technology and the limits of human nature.* So far, technology companies have attempted to pretend they are in a relationship of equals with us when it's actually been asymmetric. Technology companies have said that they are neutral, and users have equal power in the relationship with users. But it's much closer to the power that the therapist, a lawyer or a priest has since they have massively superior compromising and sensitive information about what will influence user behavior, so we have to apply fiduciary law. Unlike a doctor or a lawyer, these platforms have the truth, the whole truth and nothing but the truth about us, and they can increasingly predict invisible facts about us that you couldn't get otherwise. And with thee extractive business model of advertising, they are forced to use this asymmetry to profit in ways that we know cause harm.

The key in this is to move the business model to be responsible. With asymmetric power, they have to have asymmetric responsibility. And that's the key to preventing future catastrophes from technology that out-predicts human nature.

Government's job is to protect citizens. I tried to change Google from the inside, but I found that it's only been through external pressure—from government policymakers, shareholders and media—that has changed companies' behavior.

Government is necessary because human downgrading changes our global competitiveness with other countries, especially with China. Downgrading public health, sensemaking and critical thinking while they do not would disable our long-term capacity on the world stage.

Software is eating the world, which Netscape founder Marc Andreesen said, but it hasn't been made responsible for protecting the society that it eats. Facebook "eats" election advertising, while taking away protections for equal price campaign ads. YouTube "eats" children's development while taking away the protections of Saturday morning cartoons.

50 years ago, Mr. Rogers testified before this committee about his concern for the race to the bottom in television that rewarded mindless violence. YouTube, TikTok, Instagram can be far worse, impacting exponentially greater number of children with more alarming material. And in today's world, Mr. Rogers wouldn't have a chance. But in his hearing 50 years ago, the committee made a decision that permanently changed the course of children's television for the better. I'm hoping that a similar choice can be made today.

Thank you.

## PERSUASIVE TECHNOLOGY & OPTIMIZING FOR ENGAGEMENT

Tristan Harris, *Center for Humane Technology*
Thanks to Yael Eisenstat, Roger McNamee for contributions.

### INTRODUCTION TO WHO & WHY

I tried to change Google from the inside as a design ethicist after they bought my company in 2011, but I failed because companies don't have the right incentive to change. I've found that it is only pressure from outside—from policymakers like you, shareholders, the media, and advertisers—that can create the conditions for real change to happen.

### WHO I AM: PERSUASION & MAGIC

Persuasion is about an *asymmetry of power*.

I first learned this as a magician as a kid. I learned that the human mind is highly vulnerable to influence. Magicians say "pick any card." You feel that you've made a "free" choice, but the magician has actually influenced the outcome upstream because they have asymmetric knowledge about how your mind works.

In college, I studied at the Stanford Persuasive Technology Lab understanding how technology could persuade people's attitudes, beliefs and behaviors. We studied clicker training for dogs, habit formation, and social influence. I was project partners with one of the founders of Instagram and we prototyped a persuasive app that would alleviate depression called "Send the Sunshine". Both magic and persuasive technology represent an *asymmetry in power*– an increasing ability to influence other people's behavior.

### SCALE OF PLATFORMS AND RACE FOR ATTENTION

Today, tech platforms have more influence over our daily thoughts and actions than most governments. 2.3 billion people use Facebook, which is a psychological footprint about the size of Christianity. 1.9 billion people use YouTube, a larger footprint than Islam and Judaism combined. And that influence isn't neutral.

The advertising business model links their profit to how much attention they capture, creating a "race to the bottom of the brain stem" to extract attention by hacking lower into our lizard brains– into dopamine, fear, outrage—to win.

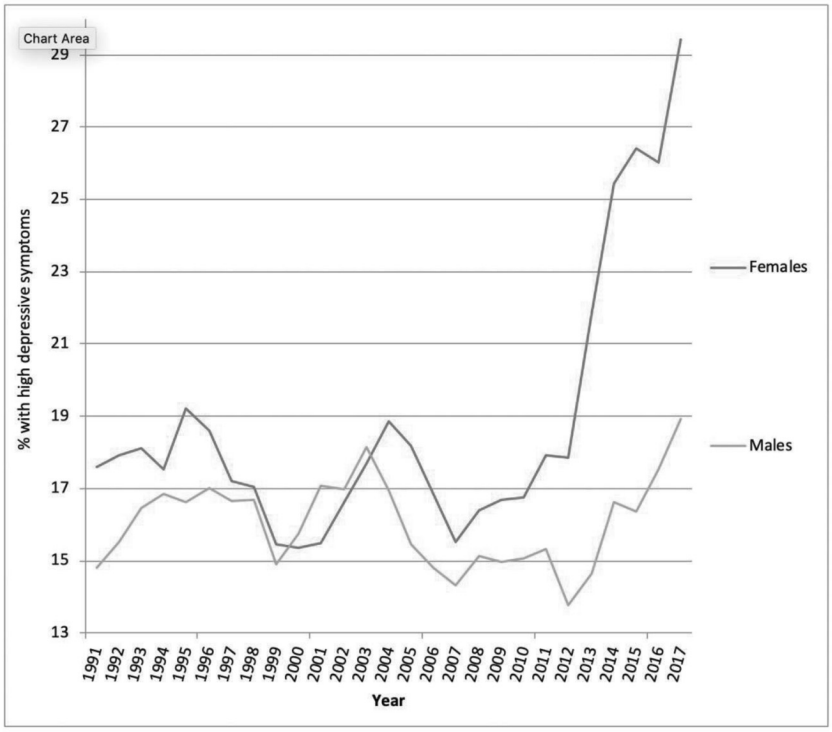
### HOW TECH HACKED OUR WEAKNESSES

It starts by *getting our attention.* Techniques like "pull to refresh" act like a slot machine to keep us "playing" even when nothing's there. "Infinite scroll" takes away stopping cues and breaks so users don't realize when to stop. You can try having self-control, but there are a thousand engineers are on the other side of the screen working against you.

Then design evolved to *get people addicted to getting attention from other people.* Features like "Follow" and "Like" drove people to independently grow their audience with drip-by-drip social validation, fueling social comparison and the rise of "influencer" culture: suddenly everyone cares about being famous.

The race went deeper into *persuading our identity:* Photo-sharing apps that include "beautification filters" that alter our self-image work better at capturing attention than apps that don't. This fueled "Body Dysmorphic Disorder," anchoring the self-image of millions of teenagers to unrealistic versions of themselves, reinforced with constant social feedback that *people only like you if you look different than you actually do.* 55 percent of plastic surgeons in a 2018 survey said they'd seen patients whose primary motivation was to look better in selfies, up from just 13 percent in 2016. Instead of companies competing for attention, now *each person competes for attention using a handful of tech platforms.*

Constant visibility to others fueled mass social anxiety and a mental health crisis. It's impossible to disconnect when you fear your social reputation could be ruined by the time you get home. After nearly two decades in decline, "high depressive symptoms" for 13–18 year old teen girls suddenly rose 170 percent between 2010—2017. Meanwhile, most people aren't aware of the growing asymmetry between persuasive technology and human weaknesses.




**USING AI TO EXTRACT ATTENTION, ROLE OF ALGORITHMS**

The arms race for attention then moved to algorithms and A.I.: companies compete on whose algorithms more accurately predict what will keep users there the longest.

For example, you hit 'play' on a YouTube video and think, "I know those other times I get sucked into YouTube, but *this time* it will be different." Two hours later you wake up from a trance and think "I can't believe I did that again." Saying we should have more self control hides an invisible asymmetry in power: *YouTube has a supercomputer pointed at your brain.*

When you hit play, YouTube wakes up an avatar, voodoo doll-like model of you. All of your video clicks, likes and views are like the hair clippings and toenail filings that make your voodoo doll look and act more like you so it can more accurately predict your behavior. YouTube then 'pricks' the avatar with millions of videos to simulate and make predictions about which ones will keep you watching. It's like playing chess against Garry Kasparov, you're going to lose. YouTube's machines are playing too many moves ahead.

That's exactly what happened: 70 percent of YouTube's traffic is now driven by recommendations, "because of what our recommendation engines are putting in front of you," said Neal Mohan, CPO of YouTube. With over a billion hours watched daily, algorithms have already taken control of two billion people's thoughts.

**TILTING THE ANT COLONY TOWARDS CRAZYTOWN**

Imagine a spectrum of videos on YouTube, from the "calm" side—rational, science-based, long, Walter Cronkite section, to the side of "crazytown"

Because YouTube wants to maximize watch time, it tilts the entire ant colony of humanity towards crazytown. It's "algorithmic extremism":

- Teen girls that played "diet" videos on YouTube were recommended anorexia videos.
- AlgoTransparency.org revealed that the most frequent keywords in recommended YouTube videos were: *get schooled, shreds, debunks, dismantles, debates, rips confronts, destroys, hates, demolishes, obliterates.*

- Watching a NASA Moon landing videos YouTube recommended "Flat Earth" conspiracies, recommended hundreds of millions of times before being downranked.
- YouTube recommended Alex Jones InfoWars videos 15 billion times—more than the combined traffic of NYTimes, Guardian, Washington Post and Fox News.
- More than 50 percent of fascist activists in a Bellingcat study credit the Internet with their red-pilling. YouTube was the single most frequently discussed website.
- When the Mueller report was released about Russian interference in the 2016 election, RussiaToday's coverage was the most recommended of 1,000+ monitored channels.
- Adults watching sexual content were recommended videos that increasingly feature young women, then girls to then children playing in bathing suits (*NYT article*)
- Fake news spreads six times faster than real news, because it's free to evolve to confirm existing beliefs unlike real news, which is constrained by the limits of what is true (*MIT Twitter study*)

*Freedom of speech is not the same as freedom of reach.* Everyone has a right to speak, but not a right to a megaphone that reaches billions of people. Social platforms amplify salacious speech without upholding any of the standards and practices required for traditional media and broadcasters. If you derived a motto from technology platforms from their observed behavior, it would be, *"with great power comes no responsibility."*

They are debasing the information environment that powers our democracy. Beyond discriminating against any party, tech platforms are discriminating against the values that make democracy work: discriminating against civility, thoughtfulness, nuance and open-mindedness.

**EQUAL, OR ASYMMETRIC?**

Once you see the extent to which technology has taken control, we have to ask, is the nature of the business relationship between platforms and users one that is contractual, a relationship between parties of equal power, or is it asymmetric?

There has been a misunderstanding about the nature of the business relationship between the platform and the user, that they have asserted that it is a contractual relationship of parties with equal power. In fact, it is much closer to the relationship of a therapist, lawyer, priest. They have superior information, such an asymmetry of power, that you have to apply fiduciary law.

Saying "we give people what they want" or "we're a neutral platform" hides a dangerous asymmetry: Google and Facebook hold levels of compromising information on two billion users that vastly exceed that of a psychotherapist, lawyer, or priest, while being able to extract benefit towards their own goals of maximizing certain behaviors.

**THE ASYMMETRY WILL ONLY GET EXPONENTIALLY WORSE**

The reason we need to apply fiduciary law now is because the situation is only going to get worse. A.I. will make technology exponentially more capable of predicting what will manipulate humans, not less.

There's a popular conspiracy theory that Facebook listens to your microphone, because the thing *you were just talking about with your friend* just showed up in your news feed. But forensics show they don't listen. More creepy: they don't have to, because they can wake up one of their 2.3 billion avatar, voodoo dolls of you to *accurately predict* the conversations you're most likely to have.

This will only get worse.

Already, platforms are easily able to:

- Predict whether you are lonely or suffer from low self-esteem
- Predict your big 5 personality traits with your temporal usage patterns alone
- Predict when you're about to get into a relationship
- Predict your sexuality before you know it yourself
- Predict which videos will keep you watching

Put together, Facebook or Google are like a priest in a confession booth who listens to two billion people's confessions, but whose only business model is to shape and control what those two billion people do while being paid by a 3rd party. Worse, the priest has a supercomputer calculating patterns between two billion people's confessions, so they can predict what confessions you're going to make, before you know you're going to make them—and sell access to the confession booth.

Technology, unchecked, will only be able to better predict what will influence our behavior, not less.

There are two ways to take control of human behavior—1) you can build more advanced A.I. to accurately predict what will manipulate someone's actions, 2) you can simplify humans by making them more predictable and reactive. Today, technology is doing both: profits within Google and Facebook get reinvested into better predictive models and machine learning to manipulate behavior, while simultaneously simplifying humans to respond to simpler and simpler stimuli. This is *checkmate humanity.*

**THE HARMS ARE A SELF–REINFORCING SYSTEM**

We often consider problems in technology as separate—addiction, distraction, fake news, polarization and teen suicides and mental health. They are not separate. They are part of an interconnected system of harms that are a direct consequence of a race to the bottom of the brain stem to extract attention.

Shortening attention spans, breakdown our shared truth, increase polarization, rewarding outrage, depressed critical thinking, increase loneliness and social isolation, increasing teen suicide and self-harm—especially among girls, rising extremism, and conspiracy thinking—and ultimately debase the information environment and social fabric we depend on.

These harms reinforce each other. When it shrinks our attention spans, we can only say simpler, 140 character messages about increasingly complex problems—driving polarization: half of people might agree with the simple call to action, but will automatically enrage the rest. NYU psychology researchers found that each word of moral outrage added to a tweet raises the retweet rate by 17 percent. Reinforcing outrage compounds mob mentality, where people become increasingly angry about things happening at increasing distances.

This leads to "callout culture" that angry mobs trolling and yelling at each other for the least charitable interpretation of simpler and simpler message. Misinterpreted statements lead to more defensiveness. This leads to more victimization, more baseline anger and polarization, and less social trust. "Callout culture" creates a chilling effect, and crowds out inclusive thinking that reflects the complex world we live in and our ability to construct shared agendas of action. More isolation also means more vulnerability to conspiracies.

As attention starts running out, companies have to "frack" for attention by splitting our attention into multiple streams—multi-tasking three or four simultaneous things at once. They might quadruple the size of the attention economy, but downgraded our attention spans. The average time we focus drops. Productivity drops.

**NAMING THE INTERCONNECTED SYSTEM OF HARMS**

These effects are interconnected and mutually reinforcing. Conservative pollster Frank Luntz calls it the "the climate change of culture." We at the Center for Humane Technology call it "human downgrading":

While tech has been upgrading the machines, they've been downgrading humans—downgrading attention spans, civility, mental health, children, productivity, critical thinking, relationships, and democracy.

**IT AFFECTS EVERYONE**

Even if you don't use these platforms, it still affects you. You still live in a country where other people vote based on what they are recommended. You still send your kids to schools with *other parents* who believe anti-vaxx conspiracies recommended to them on social media. Measles cases increased 30 percent between 2016 and 17 and leading WHO to call 'vaccine hesitancy' a top 10 global health threat.

We're all in the boat together. Human downgrading is like a dark cloud descending upon society that *affects everyone.*

**COMPETITION WITH CHINA**

But human downgrading matters for global competition. Competing with China, whichever nation least downgrades its populations' attention spans, critical thinking, mental health, and political polarization will win be more productive, healthy and fast-moving on the global stage.

**CONCLUSION**

Government's job is to protect citizens. All of this, I genuinely believe, can be fixed with changes in incentives that match the scope of the problem.

I am not against technology. The genie is out of the bottle. But we need a renaissance of "humane technology" that is designed to protect and care for human wellbeing and the social fabric upon which these technologies are built. We cannot rely on the companies alone to make that change. We need our government to create

the rules and guardrails that market forces to create competition for technology that strengthens society and human empowerment, and protects us from these harms.

Netscape founder Marc Andreesen said in 2011, "software is eating the world" because it will inevitably operate aspects of society more efficiently than without technology: taxis, election advertising, content generation, etc.

But technology shouldn't take over our social institutions and spaces, without taking responsibility for protecting them:

- Technology "ate" election campaigns with Facebook, while taking away FEC protections like equal price campaign ads.
- Tech "ate" the playing field for global information war, while replacing the protections of NATO and the Pentagon with a small teams at Facebook, Google or Twitter.
- Technology "ate" our dopamine centers of our brains—without the protection of an FDA.
- Technology "ate" children's development with YouTube, while taking away the protections of Saturday morning cartoons.

Exactly 50 years ago, children's TV show host Fred "Mister" Rogers testified to this committee about his concern for how the race to the bottom in TV rewarded mindless violence and harmed children's development. Today's world of YouTube and TikTok are far worse, impacting exponentially greater number of children with far more alarming material. Today Mister Rogers wouldn't have a chance.

But on the day Rogers testified, Senators chose to act and funded a caring vision for children in public television. It was a decision that permanently changed the course of children's television for the better. Today I hope you choose protecting citizens and the world order—by incentivizing a caring and "humane" tech economy that strengthens and protects society instead of being destructive.

The consequences of our actions as a civilization are more important than they have ever been, while technology that informs these decisions are being downgraded. If we're disabling ourselves from making good choices, that's an existential outcome.

Thank you.

---

**VIDEO: HUMANE: A NEW AGENDA FOR TECH**

You can view a video presentation of most of this material at: *https://humanetech.com/newagenda/*

---

TECHNOLOGY IS DOWNGRADING HUMANITY; LET'S REVERSE THAT TREND NOW

*Summary:* Today's tech platforms are caught in a race to the bottom of the brain stem to extract human attention. It's a race we're all losing. The result: addiction, social isolation, outrage, misinformation, and political polarization are all part of one interconnected system, called *human downgrading,* that poses an existential threat to humanity. The Center for Humane Technology believes that we can reverse that threat by redesigning tech to better protect the vulnerabilities of human nature and support the social fabric.

**THE PROBLEM: Human Downgrading**

*What's the underlying problem with technology's impact on society?*

*We're surrounded by a growing cacophony of grievances and scandals.* Tech addiction, outrage-ification of politics, election manipulation, teen depression, polarization, the breakdown of truth, and the rise of vanity/micro-celebrity culture. If we continue to complain about separate issues, nothing will change. The truth is, these are not separate issues. They are an interconnected systems of harms we call *human downgrading.*

*The race for our attention is the underlying cause of human downgrading.* More than two billion people—a psychological footprint bigger than Christianity—are jacked into social platforms designed with the goal of not just getting our attention, but getting us addicted to getting attention from others. This an extractive attention economy. Algorithms recommend increasingly extreme, outrageous topics to keep us glued to tech sites fed by advertising. Technology continues to tilt us toward outrage. It's a race to the bottom of the brainstem that's downgrading humanity.

*By exploiting human weaknesses, tech is taking control of society and human history.* As magicians know, to manipulate someone, you don't have to overwhelm their strengths, you just have to overwhelm their weaknesses. While futurists were look-

ing out for the moment when technology would surpass human strengths and steal our jobs, we missed the much earlier point where technology surpasses human weaknesses. It's already happened. By preying on human weaknesses—fear, outrage, vanity—technology has been downgrading our well-being, while upgrading machines.

Consider these examples:

- Extremism exploits our brains: With over a billion hours on YouTube watched daily, 70 percent of those billion hours are from the recommendation system. The most recommended keywords in recommended videos were *get schooled, shreds, debunks, dismantles, debates, rips confronts, destroys, hates, demolishes, obliterates.*
- Outrage exploits our brains: For each moral-emotional word added to a tweet it raised its retweet rate by 17 percent (*PNAS*).
- Insecurity exploits our brains: In 2018, if you were a teen girl starting on a dieting video, YouTube's algorithm recommended anorexia videos next because those were better at keeping attention.
- Conspiracies exploit our brains: And if you are watching a NASA moon landing, YouTube would recommend Flat Earth conspiracies millions of time. YouTube recommended Alex Jones (InfoWars) conspiracies 15 billion times (*source*).
- Sexuality exploits our brains: Adults watching sexual content were recommended videos that increasingly feature young women, then girls to then children playing in bathing suits (*NYT article*)
- Confirmation bias exploits our brains: Fake news spreads six times faster than real news, because it's unconstrained while real news is constrained by the limits of what is true (*MIT Twitter study*)

*Why did this happen in the first place? Because of the advertising business model.*

Free is the most expensive business model we've ever created. We're getting "free" destruction of our shared truth, "free" outrage-ification of politics, "free" social isolation, "free" downgrading of critical thinking. Instead of paying professional journalists, the "free" advertising model incentivizes platforms to extract "free labor" from users by addicting them to getting attention from others and to generate content for free. Instead of paying human editors to choose what gets published to whom, it's *cheaper* to use automated algorithms that match salacious content to responsive audiences—replacing news rooms with amoral server farms.

*This has debased trust and the entire information ecology.*

*Social media has created an uncontrollable digital Frankenstein.* Tech platforms can't scale safeguards to these rising challenges across the globe, more than 100 languages, in millions of FB groups or YouTube channels producing hours of content. With two billion automated channels or "Truman shows" personalized to each user, hiring 10,000 people is inadequate to the exponential complexity—there's no way to control it.

- The 2017 genocide in Myanmar was exacerbated by unmoderated fake news with only four Burmese speakers at Facebook to monitor its 7.3M users (*Reuters report*)
- Nigeria had 4 fact checkers in a country where 24M people were on Facebook. (*BBC report*)
- India's population has 22 languages in their recent election. How many engineers or moderators at Facebook or Google know those languages?

*Human downgrading is existential for global competition.* Global powers that downgrade their populations will harm their economic productivity, shared truth, creativity, mental health and wellbeing the next generations—solving this issue is urgent to win the global competition for capacity.

*Society faces an urgent, existential threat from parasitic tech platforms.* Technology's outpacing of human weaknesses is only getting worse– from more powerful addiction to more power deep fakes. Just as our world problems go up in complexity and urgency—climate change, inequality, public health—our capacities to make sense of the world and act together is going down. Unless we change course right now, this is checkmate on humanity.

**WE CAN SOLVE THIS PROBLEM: Catalyzing a Transition to Humane Technology**

*Human downgrading is like the global climate change of culture.* Like climate change it can be catastrophic. But unlike climate change, *only about 1,000 people need to change what they're doing.*

Because each problem—from "slot machines" hacking our lizard brains to "Deep Fakes" hacking our trust have to do with *not protecting human instincts,* if we design all systems to protect humans, we can not only avoid downgrading humans, but we can upgrade human capacity.

*Giving a name to the connected systems—the entire surface area—of human downgrading is crucial* because without it, solution creators end up working in silos and attempt to solve the problem by playing an infinite "whack-a-mole" game.

*There are three aspects to catalyzing Humane Technology:*

1. *Humane Social Systems.* We need to get deeply sophisticated about not just technology, but *human nature* and the ways one impacts the other. Technologists must approach innovation and design with an awareness of protecting of the ways we're manipulated as human beings. Instead of more artificial intelligence or more advanced tech, we actually just need more sophistication about what protects and heals human nature and social systems.

   CHT has developed a starting point that technologists can use to explore and assess how tech affects us at the individual, relational and societal levels. *(design guide.)*

   - *Phones protecting against slot machine "drip" rewards*
   - *Social networks protecting our relationships off the screen*
   - *Digital media designed to protect against DeepFakes by recognizing the vulnerabilities in our trust*

2. *Humane AI, not overpowering AI.* AI already has asymmetric power over human vulnerabilities, by being able to perfectly predict what will keep us watching or what can politically manipulate us. Imagine a lawyer or a priest with asymmetric power to exploit you whose business model was to sell access to perfectly exploit you to another party. We need to convert that into AI to acts in our interest by making them fiduciaries to our values—that means prohibiting advertising business models that extract from that intimate relationship.
3. *Humane Regenerative Incentives, instead of Extraction.* We need to stop fracking people's attention. We need to develop a new set of incentives that accelerate a market competition to fix these problems. We need to create a race to the top to align our lives with our values instead to the bottom of the brain stem.

   Policy and organizational incentives that guide operations of technology makers to emphasize the qualities that enliven the social fabric
   We need an AI sidekick that's designed to protect the limits of human nature and be acting in our interests like a GPS for life that helps us get where we need to go.

The *Center for Humane Technology* supports the community in catalyzing this change:

- *Product teams at tech companies* can integrate humane social systems design into products that protect human vulnerabilities and support the social fabric.
- *Tech gatekeepers* such as Apple and Google can encourage apps to competing for our trust, not our attention, to fulfill values—by re-shaping App Stores, business models, and the interaction between apps competing on Home Screens and Notifications.
- *Policymakers* can protect citizens and shift incentives for tech companies.
- *Shareholders* can demand commitments from companies to shift away from engagement-maximizing business models that are a huge source of investor risk.
- *VCs* can fund that transition
- *Entrepreneurs* can build products that are sophisticated about humanity.
- *Journalists* can shine light on the systemic problems and solutions instead of the scandals and the grievances.
- *Tech workers* can raise their voices around the harms of human downgrading.
- *Voters* can demand policy from policymakers to reverse kids being downgraded.

There's change afoot. When people start speaking up with shared language and a humane tech agenda, things will change. For more information, please visit the *Center for Humane Technology.*

Senator THUNE. Thank you, Mr. Harris.

Ms. Stanphill.

**STATEMENT OF MAGGIE STANPHILL,**
**DIRECTOR OF USER EXPERIENCE, GOOGLE**

Ms. STANPHILL. Chairman Thune, Ranking Member Schatz, Members of the Committee, thank you for inviting me to testify today on Google's efforts to improve the digital well-being of our users.

I appreciate the opportunity to outline our programs and to discuss our research in this space.

My name is Maggie Stanphill. I'm the User Experience Director and I lead the Cross-Google Digital Well-Being Initiative.

Google's Digital Well-Being Initiative is an initiative that's a top company goal and we focus on providing users with insights about their individual tech habits and the tools to support an intentional relationship with technology.

At Google, we've heard from many of our users all over the world that technology is a key contributor to their sense of well-being. It connects them to those they care about, it provides information and resources, it builds their sense of safety and security, and this access has democratized information and provided services for billions of users around the world.

For most people, their interaction with technology is positive and they are able to make healthy choices about screen time and overall use. But as technology becomes increasingly prevalent in our day-to-day lives, for some people it can distract from the things that matter most. We believe technology should play a useful role in people's lives and we've committed to helping people strike a balance that feels right for them.

This is why our CEO, Sundar Pichai, first announced the Digital Well-Being Initiative with several new features across Android, Family Link, YouTube, Gmail, all of these to help people better understand their tech usage and focus on what matters most.

In 2019, we applied what we learned from users and experts and introduced a number of new features to support our Digital Well-Being Initiative. I'd like to go into more depth about our products and tools we've developed for our users.

On Android, the latest version of our Mobile Operating System, we added key capabilities to help users take a better balance with technology and make sure that they can focus on raising awareness of tech usage and providing controls to help them oversee their tech use.

This includes a dashboard. It shows information about their time on devices. It includes app timers so people can set time on specific apps. It requires a do not disturb function to silence phone calls and texts as well as those visual interruptions that pop up, and we've introduced a new wind-down feature that automatically puts the users' display into night light mode and that reduces blue light and gray scale to remove color and ultimately the temptation to scroll.

Finally, we've got a new setting called Focus Mode. This allows pausing specific apps and notifications that users might find distracting.

On YouTube, we have similarly launched a series of updates to help our users define their own sense of well-being. This includes time-watched profiles, take-a-break reminders, the ability to disable audible notifications, and the option to combine all YouTube app notifications into one notification.

We've also listened to the feedback about the YouTube recommendation system. Over the past year, we've made a number of improvements to these recommendations, raising up content from authoritative sources when people are coming to YouTube for news as well as reducing recommendations of content that comes close to violating our policies or spreads harmful misinformation.

When it comes to children, we believe the bar is even higher. That's why we've created Family Link to help parents stay in the loop when their child explores on Android and on Android Q, parents will be able to set screen time limits and bedtimes and remotely lock their child's device.

Similarly, YouTube Kids was designed with the goal of ensuring that parents have control over the content their children watch. In order to keep the videos in the YouTube Kids' app family friendly, we use a mix of filters, user feedback and moderators. We also offer parents the option to take full control over what their children watch by hand-selecting the content that appears in their app.

We're also actively conducting our own research and engaging in important expert partnerships with independent researchers to build a better understanding of the many personal impacts of digital technology.

We believe this knowledge can help shape new solutions and ultimately drive the entire industry toward creating products that support a better sense of well-being.

To make sure we are evolving the strategy, we have launched a longitudinal study to better understand the effectiveness of our digital well-being tools. We believe that this is just the beginning.

As technology becomes more integrated into people's daily lives, we have a responsibility to ensure that our products support their digital well-being. We are committed to investing more, optimizing our products, and focusing on quality experiences.

Thank you for the opportunity to outline our efforts in this space. I'm happy to answer any questions you might have.

[The prepared statement of Ms. Stanphill follows:]

PREPARED STATEMENT OF MAGGIE STANPHILL, DIRECTOR OF USER EXPERIENCE, GOOGLE

**I. Introduction**

Chairman Thune, Ranking Member Schatz, Members of the Committee: Thank you for inviting me to testify today on Google's efforts to improve the digital wellbeing of our users. I appreciate the opportunity to outline our programs and discuss our research in this space.

My name is Maggie Stanphill. I am a User Experience Director at Google, and I lead our global Digital Wellbeing Initiative.[1] Google's Digital Wellbeing Initiative

[1] *https://wellbeing.google*

is a top company goal, focused on providing our users with insights about their digital habits and tools to support an intentional relationship with technology.

At Google, our goal has always been to create products that improve the lives of the people who use them. We're constantly inspired by the ways people use technology to pursue knowledge, explore their passions and the world around them, or simply make their everyday lives a little easier. We've heard from many of our users—all over the world—that technology is a key contributor to their sense of wellbeing. It connects them to those they care about and it provides information and resources that build their sense of safety and security. This access has democratized information and provided services for billions of people around the world. In many markets, smartphones are the main connection to the digital world and new opportunities, such as education and work. For most people, their interaction with technology is positive and they are able to make healthy choices about screen time and overall use.

But for some people, as technology becomes increasingly prevalent in our day-to-day lives, it can distract from the things that matter most. We believe technology should play a helpful, useful role in all people's lives, and we're committed to helping everyone strike a balance that feels right for them. This is why last year, as a result of extensive research and investigation, we introduced our Digital Wellbeing Initiative: a set of principles that resulted in tools and features to help people find their own sense of balance. Many experts recommend self-awareness and reflection as an essential step in creating a balance with technology. With that in mind, at Google's 2018 I/O Developers Conference, our CEO Sundar Pichai first announced several new features across Android, Family Link, YouTube, and Gmail to help people better understand their tech usage, focus on what matters most, disconnect when needed, and create healthy habits for their families. These tools help people gain awareness of time online, disconnect for sleep, and manage their tech habits.

In 2019, we applied what we learned from users and experts. We know that one size doesn't fit all and behavior change is individual. Some people respond more readily to extrinsic motivation (like setting their app timer) and others to intrinsic motivations (based on personal goals like spending more time with family). With this ongoing and evolving approach to supporting our users' digital wellbeing, I'd like to go into more depth about key products and tools we have developed for our users.

**II. Android**

The latest version of our mobile operating system, *Android,* added key capabilities to help users achieve the balance with technology they are looking for, with a focus on raising awareness of tech usage and providing controls to help them interact with their devices the way they want.

- First, since we know that people are motivated when they can reflect on tangible behaviors they want to change, we have a *dashboard* that provides information all in one place. This shows how much time they spend time on their devices, including time spent in apps, how many times they've unlocked their phone, and how many notifications they've received.
- With *app timers,* people can set time limits on specific apps. It nudges them when they are close to their limit, and then will gray out the app icon to help remind them of their goal. We have seen that app timers help people stick to their goals 90 percent of the time.
- Android's *Do Not Disturb function* is one way we address the impact that notifications have on the cycle of obligation we found in user research. Do Not Disturb silences the phone calls and texts as well as the visual interruptions that pop up on users' screens. And to make it even easier to use, we created a new gesture. If this feature is turned on, when people turn over their phone on the table, it automatically enters Do Not Disturb mode so they can focus on being present.
- Because there is extensive research that indicates the importance of sleep on people's overall wellbeing, we developed *Wind Down.* This function gets people and their phones ready for bed by establishing a routine that includes their phone going to into Night Light mode to reduce blue light and Grayscale to remove color and the attendant temptation to scroll. Since introducing this function, we have seen Wind Down lead to a 27 percent drop in nightly usage for those who use it.
- Finally, at Google's 2019 I/O Developer Conference, we introduced a new setting called "*focus mode.*" This works like Wind Down but can be used in other contexts. For example, if you're at university and you need to focus on a research

assignment, you can set "focus mode" to pause the apps and notifications you find distracting.

**III. YouTube**

Individuals use YouTube differently. Some of us use it to learn new things, while others use it when they need a laugh or to stay in touch with their favorite creators. Whatever their use case, we want to help everyone better understand their tech usage, disconnect when needed, and create healthy habits. That's why YouTube launched a series of updates to help users develop their own sense of digital wellbeing.

- *Time watched profile:* This profile in the main account menu gives users a better understanding of how much they watch. It lets users see how long they have watched YouTube videos today, yesterday, and over the past seven days.
- *Take a break reminder:* Users can opt-in to set a reminder that appears during long watch sessions. They receive a reminder to take a break after the amount of time they specified. We have served 1 billion reminders since the inception of the feature.
- *Scheduled Digest for Notifications:* This feature allows users to combine all of the daily push notifications they receive from the YouTube app into a single combined notification. Users set a specific time to receive their scheduled digest and from then on, they receive only one notification per day.
- *Disable notification sounds and vibrations:* This feature ensures that notifications from the YouTube app are sent silently to your phone during a specified time period each day. By default, all sounds and vibrations will be disabled between 10pm and 8am, but you can enable/disable the feature and customize the start and end times from your Settings.

In addition to our efforts to help improve users' awareness of their usage of the YouTube platform, we have also listened to feedback about the YouTube recommendations system. We understand that the system has been of particular interest to the Committee. We recognize we have a responsibility, not just in the content we decide to leave up or remove from our platform, but for what we choose to recommend to people. Recommendations are a popular and useful tool in the vast majority of situations, and help users discover new artists and creators and surface content to users that they might find interesting or relevant to watch next. YouTube is a vast library of content, and search alone is an insufficient mechanism to find content that might be relevant to you. YouTube works by surfacing recommendations for content that is similar to the content you have selected or is popular on the site, in the same way other online services recommend related TV shows, and this works well for the majority of users on YouTube when watching music or entertainment.

Over the past year, we've made a number of improvements to these recommendations, including raising up content from authoritative sources when people are coming to YouTube for news, as well as reducing recommendations of content that comes close to violating our policies or spreads harmful misinformation. Thanks to this change, the number of views this type of content gets from recommendations has dropped by over 50 percent in the U.S.

**IV. Family Link and YouTube Kids**

We believe the bar on digital wellbeing should be even higher when it comes to children. This is why we launched the Family Link app in 2017 to help parents stay in the loop as their child explores on their Android device. For Android Q, we have gone a step further and are making Family Link part of every device. When parents set up their child's device with Family Link, we'll automatically connect the child's device to the parent's device to supervise. We'll let parents set daily screen-time limits, set a device bedtime, and remotely lock their child's device when it's time to take a break. Family Link also allows parents to approve actions before their kids can download any app or make any purchases in apps, and after download they can see their child's app activity and block an app any time. These features will be available later this summer with the consumer launch of Android Q.

Similarly, YouTube Kids was designed with the goal of ensuring parents have control over the content their children watch. YouTube Kids uses a mix of filters, user feedback, and moderators to keep the videos in YouTube Kids family friendly. There are also built-in timers for length of use, no public comments, and easy ways to block or flag content. In Parent Approved Mode, parents can take full control over what their children watch by hand selecting the content that appears in the app.

**V. Other Efforts**

Wellbeing tools are also available on a range of other Google products and services. Using the Google Assistant, you can now voice-activate Do Not Disturb mode—silencing all notifications and communications—and the Bedtime Routine. On Google Wifi, parents can pause connectivity on one or all of their kids' devices simultaneously, or help them wind down by scheduling a time-out. While on Google Home you can also easily schedule breaks from one or all of the devices your family uses. Gmail now has the option to allow only high priority notifications and a snooze function to let you put off notifications until later.

*User education:* Beyond making tools available to help our users improve their digital wellbeing, we're also committed to helping through user education. We believe it's important to equip kids to make smart decisions online. That's why we're continuing to work with educators around the world on our "Be Internet Awesome" program. This is a Google-designed approach that teaches kids to be safer explorers of the digital world. The 5-part curriculum also teaches kids to be secure, kind, and mindful while online. We have committed to reach five million kids with this program in the coming year.

*Industry outreach:* Similarly, we are also thinking through our role in the broader Internet ecosystem and trying to find ways to help users find the content they're looking for quickly without extensive device usage. One major change we have made in this space can be found in our efforts to reduce the use of interstitials. Pages that show intrusive interstitials provide a poorer experience for users than other pages where content is immediately accessible. This can be problematic on mobile devices where screens are often smaller. To improve the mobile search experience, pages where content is not easily accessible to a user on the transition from the mobile search results may not rank as highly. Some examples of techniques that make content less accessible to a user can include showing a popup that covers the main content, either immediately after the user navigates to a page from the search results or while they are looking through the page.

*Research & strategy:* We're also actively conducting our own research and exploring partnerships with independent researchers and experts to build a better understanding of the many personal impacts of digital technology. We believe this knowledge can help shape new solutions and ultimately drive the entire technology industry toward creating products that support digital wellbeing.

From research in the U.S. in March 2019,[2] we know:

1. *One in three people* (33 percent) in the U.S. have *made or attempted to make changes in how they use technology* in order to address negative effects they've experienced.
2. *Taking action DOES help: 80+ percent* of users who took an action found the action to be helpful.

To make sure we are evolving this strategy, in 2018 we conducted research with more than 90,000 people globally, and we have launched a longitudinal study to better understand the effectiveness of our digital wellbeing tools in helping people achieve greater balance in their tech use. These findings will help us optimize our current offerings while inspiring brand new tools. We are also exploring new ways to understand people's overall satisfaction with our product experiences. Through emphasizing user goals (rather than solely measuring engagement and time spent on our platforms), we can deliver more helpful experiences that also support people's digital wellbeing.

*Partnerships:* To bolster our digital wellbeing efforts for kids, one key partner we work with is the Family Online Safety Institute (FOSI), an international, nonprofit organization that works to make the online world safer for kids. FOSI convenes leaders in industry, government, and nonprofit sectors to collaborate and innovate new solutions and policies in the field of online safety. Through research, resources, events, and special projects, FOSI promotes a culture of responsibility online and encourages a sense of digital citizenship for all.

We also support the bipartisan and bicameral Children and Media Research Advancement (CAMRA) Act, which proposes to authorize the National Institutes of Health to research technology's and media's effects on infants, children, and adolescents in core areas of cognitive, physical, and socio-emotional development.

*Wellbeing.google:* Finally, you can find can find more of our tools, as well as expert recommendations, at wellbeing.google.com.

---

[2] *https://www.blog.google/outreach-initiatives/digital-wellbeing/find-your-balance-new-digital-wellbeing-tools/*

**VI. Conclusion**

We believe this is just the beginning of our work in this space. As technology becomes more integrated into people's daily lives, we have a responsibility to ensure that our products support their digital wellbeing. We are committed to investing more, optimizing our products, and focusing on quality experiences.

Thank you for the opportunity to outline our efforts in this space. I'm happy to answer any questions you might have.

Senator THUNE. Thank you, Ms. Stanphill.

Mr. Wolfram.

**STATEMENT OF DR. STEPHEN WOLFRAM, FOUNDER AND CHIEF EXECUTIVE OFFICER, WOLFRAM RESEARCH, INC.**

Dr. WOLFRAM. Thanks for inviting me here today. I have to say that this is pretty far from my usual kind of venue, but I have spent my life working on the science and technology of computation and AI and perhaps some of what I know can be helpful here today.

So, first of all, here's a way I think one can kind of frame the issue. So many of the most successful Internet companies, like Google and Facebook and Twitter, are what one can call automated content selection businesses. They ingest lots of content and then they essentially use AI to select what to actually show to their users.

How does that AI work? How one can tell if it's doing the right thing? People often assume that computers just run algorithms that someone sat down and wrote but modern AI systems don't work that way. Instead, lots of the programs they use are actually constructed automatically, usually by learning from some massive number of examples.

If you go look inside those programs, there's usually embarrassingly little that we humans can understand in there, and here's the real problem. It's sort of a fact of basic science that if you insist on explainability, then you can't get the full power of the computational system or AI.

So if you can't open up the AI and understand what it's doing, how about sort of putting external constraints on it? Can you write a contract that says what the AI is allowed to do? Well, partly actually through my own work, we're starting to be able to formulate computational contracts, contracts that are written not in legalese but in a precise executable computational language suitable for an AI to follow.

But what does the contract say? I mean, what's the right answer for what should be at the top of someone's newsfeed or what exactly should be the algorithmic rule for balance or diversity of content?

Well, as AIs start to run more and more of our world, we're going to have to develop a whole network of kind of AI laws and it's going to be super-important to get this right, probably starting off by agreeing on sort of the right AI constitution. It's going to be a hard thing kind of making computational how people want the world to work.

Right now that's still in the future, but, OK, so what can we do about people's concerns now about automatic content selection? I have to say that I don't see a purely technical solution, but I didn't

want to come here and say that everything is impossible, especially since I personally like to spend my life solving "impossible problems," but I think that if we want to do it, we actually can use technology to set up kind of a market-based solution.

I've got a couple of concrete suggestions about how to do that. Both are based on giving users a choice about who to trust for the final content they see.

One of the suggestions introduces what I call final ranking providers, the other introduces constraint providers. In both cases, these are third party providers who basically insert their own little AIs into the pipeline of delivering content to users and the point is that users can choose which of these providers they want to trust.

The idea is to leverage everything that the big automated content selection businesses have but to essentially add a new market layer. So users get to know that they're picking a particular way that content is selected for them.

It also means that you get to avoid kind of all or nothing banning of content and you don't have kind of a single point of failure for spreading bad content and you open up a new market potentially delivering even higher value for users.

Of course, for better or worse, unless you decide to force certain content or diversity of content, which you could, people can live kind of in their own content bubbles, though importantly, they get to choose those themselves.

Well, there are lots of technical details about everything I'm saying as well as some deep science about what's possible and what's not, and I tried to explain a little bit more about that in my written testimony.

I'm happy to try and answer whatever questions I can here.

Thank you.

[The prepared statement of Dr. Wolfram follows:]

PREPARED STATEMENT OF STEPHEN WOLFRAM, FOUNDER AND CHIEF EXECUTIVE OFFICER, WOLFRAM RESEARCH, INC.

**About Me**

I have been a pioneer in the science and technology of computation for more than 40 years. I am the creator of Wolfram|Alpha which provides computational knowledge for Apple's Siri and Amazon's Alexa, and is widely used on the web, especially by students. I am also the creator of the Mathematica software system, which over the course of more than 30 years has been used in making countless inventions and discoveries across many fields. All major U.S. universities now have site licenses for Mathematica, and it is also extensively used in U.S. government R&D.

My early academic work was in theoretical physics. I received my PhD in physics at Caltech in 1979 when I was 20 years old. I received a MacArthur Fellowship in 1981. I was on the faculty at Caltech, then was at the Institute for Advanced Study in Princeton, then moved to the University of Illinois as Professor of Physics, Mathematics and Computer Science. I founded my first software company in 1981, and have been involved in the computer industry ever since.

In the late 1980s, I left academia to found Wolfram Research, and have now been its CEO for 32 years. During that time, I believe Wolfram Research has established itself as one of the world's most respected software companies. We have continually pursued an aggressive program of innovation and development, and have been responsible for many technical breakthroughs. The core of our efforts has been the long-term development of the Wolfram Language. In addition to making possible both Mathematica and Wolfram|Alpha, the Wolfram Language is the world's only full-scale computational language. Among its many implications are the ubiquitous

delivery of computational intelligence, the broad enabling of "computational X" fields, and applications such as computational contracts.

In addition to my work in technology, I have made many contributions to basic science. I have been a pioneer in the study of the computational universe of possible programs. Following discoveries about cellular automata in the early 1980s, I became a founder of the field of complexity theory. My additional discoveries—with implications for the foundations of mathematics, physics, biology and other areas—led to my 2002 bestselling book *A New Kind of Science.* I am the discoverer of the simplest axiom system for logic, as well as the simplest universal Turing machine. My Principle of Computational Equivalence has been found to have wide implications not only in science but also for longstanding questions in philosophy.

My technological work has made many practical contributions to artificial intelligence, and the 2009 release of Wolfram|Alpha—with its ability to answer a broad range of questions posed in natural English—was heralded as a significant breakthrough in AI. My scientific work has been seen as important in understanding the theory and implications of AI, and issues such as AI ethics.

I have never been directly involved in automated content selection businesses of the kind discussed here. Wolfram|Alpha is based on built-in computational knowledge, not searching existing content on the web. Wolfram Research is a privately held company without outside investors. It employs approximately 800 people, mostly in R&D.

I have had a long commitment to education and to using the Wolfram Language to further computational thinking. In addition to writing a book about computational thinking for students, my other recent books include *Idea Makers* (historical biographies), and the forthcoming *Adventures of a Computational Explorer.*

For more information about me, see *http://stephenwolfram.com*

*Summary*

Automated content selection by internet businesses has become progressively more contentious—leading to calls to make it more transparent or constrained. I explain some of the complex intellectual and scientific problems involved, then offer two possible technical and market suggestions for paths forward. Both are based on giving users a choice about who to trust for the final content they see—in one case introducing what I call "final ranking providers", and in the other case what I call "constraint providers".

**The Nature of the Problem**

There are many kinds of businesses that operate on the internet, but some of the largest and most successful are what one can call *automated content selection businesses.* Facebook, Twitter, YouTube and Google are all examples. All of them deliver content that others have created, but a key part of their value is associated with their ability to (largely) automatically select what content they should serve to a given user at a given time—whether in news feeds, recommendations, web search results, or advertisements.

What criteria are used to determine content selection? Part of the story is certainly to provide good service to users. But the paying customers for these businesses are not the users, but advertisers, and necessarily a key objective of these businesses must be to maximize advertising income. Increasingly, there are concerns that this objective may have unacceptable consequences in terms of content selection for users. And in addition there are concerns that—through their content selection—the companies involved may be exerting unreasonable influence in other kinds of business (such as news delivery), or in areas such as politics.

Methods for content selection—using machine learning, artificial intelligence, etc.—have become increasingly sophisticated in recent years. A significant part of their effectiveness—and economic success—comes from their ability to use extensive data about users and their previous activities. But there has been increasing dissatisfaction and, in some cases, suspicion about just what is going on inside the content selection process.

This has led to a desire to make content selection more transparent, and perhaps to constrain aspects of how it works. As I will explain, these are not easy things to achieve in a useful way. And in fact, they run into deep intellectual and scientific issues, that are in some ways a foretaste of problems we will encounter ever more broadly as artificial intelligence becomes more central to the things we do. Satisfactory ultimate solutions will be difficult to develop, but I will suggest here two near-term practical approaches that I believe significantly address current concerns.

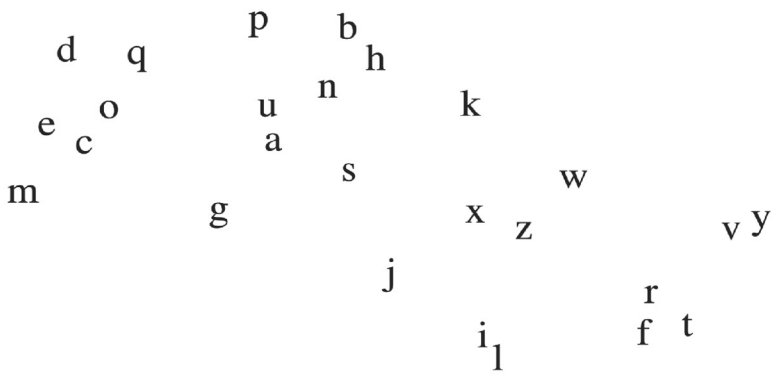
**How Automated Content Selection Works**

Whether one's dealing with videos, posts, webpages, news items or, for that matter, ads, the underlying problem of automated content selection (ACS) is basically always the same. There are many content items available (perhaps even billions of them), and somehow one has to quickly decide which ones are "best" to show to a given user at a given time. There's no fundamental principle to say what "best" means, but operationally it's usually in the end defined in terms of what maximizes user clicks, or revenue from clicks.

The major innovation that has made modern ACS systems possible is the idea of automatically extrapolating from large numbers of examples. The techniques have evolved, but the basic idea is to effectively deduce a model of the examples and then to use this model to make predictions, for example about what ranking of items will be best for a given user.

Because it will be relevant for the suggestions I'm going to make later, let me explain here a little more about how most current ACS systems work in practice. The starting point is normally to extract a collection of perhaps hundreds or thousands of features (or "signals") for each item. If a human were doing it, they might use features like: "How long is the video? Is it entertainment or education? Is it happy or sad?" But these days—with the volume of data that's involved—it's a machine doing it, and often it's also a machine figuring out what features to extract. Typically the machine will optimize for features that make its ultimate task easiest—whether or not (and it's almost always not) there's a human-understandable interpretation of what the features represent.

As an example, here are the letters of the alphabet automatically laid out by a machine in a "feature space" in which letters that "look similar" appear nearby:




How does the machine know what features to extract to determine whether things will "look similar"? A typical approach is to give it millions of images that have been tagged with what they are of ("elephant", "teacup", etc.). And then from seeing which images are tagged the same (even though in detail they look different), the machine is able—using the methods of modern machine learning—to identify features that could be used to determine how similar images of anything should be considered to be.

OK, so let's imagine that instead of letters of the alphabet laid out in a 2D feature space, we've got a million videos laid out in a 200-dimensional feature space. If we've got the features right, then videos that are somehow similar should be nearby in this feature space.

But given a particular person, what videos are they likely to want to watch? Well, we can do the same kind of thing with people as with videos: we can take the data we know about each person, and extract some set of features. "Similar people" would then be nearby in "people feature space", and so on.

But now there's a "final ranking" problem. Given features of videos, and features of people, which videos should be ranked "best" for which people? Often in practice, there's an initial coarse ranking. But then, as soon as we have a specific definition of "best"—or enough examples of what we mean by "best"—we can use machine learning to learn a program that will look at the features of videos and people, and will effectively see how to use them to optimize the final ranking.

The setup is a bit different in different cases, and there are many details, most of which are proprietary to particular companies. However, modern ACS systems—dealing as they do with immense amounts of data at very high speed—are a triumph of engineering, and an outstanding example of the power of artificial intelligence techniques.

**Is It "Just an Algorithm"?**

When one hears the term "algorithm" one tends to think of a procedure that will operate in a precise and logical way, always giving a correct answer, not influenced by human input. One also tends to think of something that consists of well-defined steps, that a human could, if needed, readily trace through.

But this is pretty far from how modern ACS systems work. They don't deal with the same kind of precise questions ("What video should I watch next?" just isn't something with a precise, well-defined answer). And the actual methods involved make fundamental use of machine learning, which doesn't have the kind of well-defined structure or explainable step-by-step character that's associated with what people traditionally think of as an "algorithm". There's another thing too: while traditional algorithms tend to be small and self-contained, machine learning inevitably requires large amounts of externally supplied data.

In the past, computer programs were almost exclusively written directly by humans (with some notable exceptions in my own scientific work). But the key idea of machine learning is instead to create programs automatically, by "learning the program" from large numbers of examples. The most common type of program on which to apply machine learning is a so-called neural network. Although originally inspired by the brain, neural networks are purely computational constructs that are typically defined by large arrays of numbers called weights.

Imagine you're trying to build a program that recognizes pictures of cats versus dogs. You start with lots of specific pictures that have been identified—normally by humans—as being either of cats or dogs. Then you "train" a neural network by showing it these pictures and gradually adjusting its weights to make it give the correct identification for these pictures. But then the crucial point is that the neural network generalizes. Feed it another picture of a cat, and even if it's never seen that picture before, it'll still (almost certainly) say it's a cat.

What will it do if you feed it a picture of a cat dressed as a dog? It's not clear what the answer is supposed to be. But the neural network will still confidently give some result—that's derived in some way from the training data it was given.

So in a case like this, how would one tell why the neural network did what it did? Well, it's difficult. All those weights inside the network were learned automatically; no human explicitly set them up. It's very much like the case of extracting features from images of letters above. One can use these features to tell which letters are similar, but there's no "human explanation" (like "count the number of loops in the letter") of what each of the features are.

Would it be possible to make an explainable cat vs. dog program? For 50 years most people thought that a problem like cat vs. dog just wasn't the kind of thing computers would be able to do. But modern machine learning made it possible—by learning the program rather than having humans explicitly write it. And there are fundamental reasons to expect that there can't in general be an explainable version—and that if one's going to do the level of automated content selection that people have become used to, then one cannot expect it to be broadly explainable.

Sometimes one hears it said that automated content selection is just "being done by an algorithm", with the implication that it's somehow fair and unbiased, and not subject to human manipulation. As I've explained, what's actually being used are machine learning methods that aren't like traditional precise algorithms.

And a crucial point about machine learning methods is that by their nature they're based on learning from examples. And inevitably the results they give depend on what examples were used.

And this is where things get tricky. Imagine we're training the cat vs. dog program. But let's say that, for whatever reason, among our examples there are spotted dogs but no spotted cats. What will the program do if it's shown a spotted cat? It might successfully recognize the shape of the cat, but quite likely it will conclude—based on the spots—that it must be seeing a dog.

So is there any way to guarantee that there are no problems like this, that were introduced either knowingly or unknowingly? Ultimately the answer is no—because one can't know everything about the world. Is the lack of spotted cats in the training set an error, or are there simply no spotted cats in the world?

One can do one's best to find correct and complete training data. But one will never be able to prove that one has succeeded.

But let's say that we want to ensure some property of our results. In almost all cases, that'll be perfectly possible—either by modifying the training set, or the neural network. For example, if we want to make sure that spotted cats aren't left out, we can just insist, say, that our training set has an equal number of spotted and unspotted cats. That might not be a correct representation of what's actually true in the world, but we can still choose to train our neural network on that basis.

As a different example, let's say we're selecting pictures of pets. How many cats should be there, versus dogs? Should we base it on the number of cat vs. dog images on the web? Or how often people search for cats vs. dogs? Or how many cats and dogs are registered in America? There's no ultimate "right answer". But if we want to, we can give a constraint that says what should happen.

This isn't really an "algorithm" in the traditional sense either—not least because it's not about abstract things; it's about real things in the world, like cats and dogs. But an important development (that I happen to have been personally much involved in for 30+ years) is the construction of a computational language that lets one talk about things in the world in a precise way that can immediately be run on a computer.

In the past, things like legal contracts had to be written in English (or "legalese"). Somewhat inspired by blockchain smart contracts, we are now getting to the point where we can write automatically executable computational contracts not in human language but in computational language. And if we want to define constraints on the training sets or results of automated content selection, this is how we can do it.

**Issues from Basic Science**

Why is it difficult to find solutions to problems associated with automated content selection? In addition to all the business, societal and political issues, there are also some deep issues of basic science involved. Here's a list of some of those issues. The precursors of these issues date back nearly a century, though it's only quite recently (in part through my own work) that they've become clarified. And although they're not enunciated (or named) as I have here, I don't believe any of them are at this point controversial—though to come to terms with them requires a significant shift in intuition from what exists without modern computational thinking.

**Data Deducibility**

*Even if you don't explicitly know something (say about someone), it can almost always be statistically deduced if there's enough other related data available*

What is a particular person's gender identity, ethnicity, political persuasion, etc.? Even if one's not allowed to explicitly ask these questions, it's basically inevitable that with enough other data about the person, one will be able to deduce what the best answers must be.

Everyone is different in detail. But the point is that there are enough commonalities and correlations between people that it's basically inevitable that with enough data, one can figure out almost any attribute of a person.

The basic mathematical methods for doing this were already known from classical statistics. But what's made this now a reality is the availability of vastly more data about people in digital form—as well as the ability of modern machine learning to readily work not just with numerical data, but also with things like textual and image data.

What is the consequence of ubiquitous data deducibility? It means that it's not useful to block particular pieces of data—say in an attempt to avoid bias—because it'll essentially always be possible to deduce what that blocked data was. And it's not just that this can be done intentionally; inside a machine learning system, it'll often just happen automatically and invisibly.

**Computational Irreducibility**

*Even given every detail of a program, it can be arbitrarily hard to predict what it will or won't do*

One might think that if one had the complete code for a program, one would readily be able to deduce everything about what the program would do. But it's a fundamental fact that in general one can't do this. Given a particular input, one can always just run the program and see what it does. But even if the program is simple, its behavior may be very complicated, and computational irreducibility implies that there won't be a way to "jump ahead" and immediately find out what the program will do, without explicitly running it.

One consequence of this is that if one wants to know, for example, whether with any input a program can do such-and-such, then there may be no finite way to determine this—because one might have to check an infinite number of possible in-

puts. As a practical matter, this is why bugs in programs can be so hard to detect. But as a matter of principle, it means that it can ultimately be impossible to completely verify that a program is "correct", or has some specific property.

Software engineering has in the past often tried to constrain the programs it deals with so as to minimize such effects. But with methods like machine learning, this is basically impossible to do. And the result is that even if it had a complete automated content selection program, one wouldn't in general be able to verify that, for example, it could never show some particular bad behavior.

**Non-explainability**

*For a well-optimized computation, there's not likely to be a human-understandable narrative about how it works inside*

Should we expect to understand how our technological systems work inside? When things like donkeys were routinely part of such systems, people didn't expect to. But once the systems began to be "completely engineered" with cogs and levers and so on, there developed an assumption that at least in principle one could explain what was going on inside. The same was true with at least simpler software systems. But with things like machine learning systems, it absolutely isn't.

Yes, one can in principle trace what happens to every bit of data in the program. But can one create a human-understandable narrative about it? It's a bit like imagining we could trace the firing of every neuron in a person's brain. We might be able to predict what a person would do in a particular case, but it's a different thing to get a high-level "psychological narrative" about why they did it.

Inside a machine learning system—say the cats vs. dogs program—one can think of it as extracting all sorts of features, and making all sorts of distinctions. And occasionally one of these features or distinctions might be something we have a word for ("pointedness", say). But most of the time they'll be things the machine learning system discovered, and they won't have any connection to concepts we're familiar with.

And in fact—as a consequence of computational irreducibility—it's basically inevitable that with things like the finiteness of human language and human knowledge, in any well-optimized computation we're not going to be able to give a high-level narrative to explain what it's doing. And the result of this is that it's impossible to expect any useful form of general "explainability" for automated content selection systems.

**Ethical Incompleteness**

*There's no finite set of principles that can completely define any reasonable, practical system of ethics*

Let's say one's trying to teach ethics to a computer, or an artificial intelligence. Is there some simple set of principles—like Asimov's Laws of Robotics—that will capture a viable complete system of ethics? Looking at the complexity of human systems of laws one might suspect that the answer is no. And in fact this is presumably a fundamental result—essentially another consequence of computational irreducibility.

Imagine that we're trying to define constraints (or "laws") for an artificial intelligence, in order to ensure that the AI behaves in some particular "globally ethical" way. We set up a few constraints, and we find that many things the AI does follow our ethics. But computational irreducibility essentially guarantees that eventually there'll always be something unexpected that's possible. And the only way to deal with that is to add a "patch"—essentially to introduce another constraint for that new case. And the issue is that this will never end: there'll be no way to give a finite set of constraints that will achieve our global objectives. (There's a somewhat technical analogy of this in mathematics, in which Gödel's theorem shows that no finite set of axiomatic constraints can give one only ordinary integers and nothing else.)

So for our purposes here, the main consequence of this is that we can't expect to have some finite set of computational principles (or, for that matter, laws) that will constrain automated content selection systems to always behave according to some reasonable, global system of ethics—because they'll always be generating unexpected new cases that we have to define a new principle to handle.

**The Path Forward**

I've described some of the complexities of handling issues with automated content selection systems. But what in practice can be done?

One obvious idea would be just to somehow "look inside" the systems, auditing their internal operation and examining their construction. But for both fundamental and practical reasons, I don't think this can usefully be done. As I've discussed, to achieve the kind of functionality that users have become accustomed to, modern

automated content selection systems make use of methods such as machine learning that are not amenable to human-level explainability or systematic predictability.

What about checking whether a system is, for example, biased in some way? Again, this is a fundamentally difficult thing to determine. Given a particular definition of bias, one could look at the internal training data used for the system—but this won't usually give more information than just studying how the system behaves.

What about seeing if the system has somehow intentionally been made to do this or that? It's conceivable that the source code could have explicit "if " statements that would reveal intention. But the bulk of the system will tend to consist of trained neural networks and so on—and as in most other complex systems, it'll typically be impossible to tell what features might have been inserted "on purpose" and what are just accidental or emergent properties.

So if it's not going to work to "look inside" the system, what about restricting how the system can be set up? For example, one approach that's been suggested is to limit the inputs that the system can have, in an extreme case preventing it from getting any personal information about the user and their history. The problem with this is that it negates what's been achieved over the course of many years in content selection systems—both in terms of user experience and economic success. And for example, knowing nothing about a user, if one has to recommend a video, one's just going to have to suggest whatever video is generically most popular—which is very unlikely to be what most users want most of the time.

As a variant of the idea of blocking all personal information, one can imagine blocking just some information—or, say, allowing a third party to broker what information is provided. But if one wants to get the advantages of modern content selection methods, one's going to have to leave a significant amount of information—and then there's no point in blocking anything, because it'll almost certainly be reproducible through the phenomenon of data deducibility.

Here's another approach: what about just defining rules (in the form of computational contracts) that specify constraints on the results content selection systems can produce? One day, we're going to have to have such computational contracts to define what we want AIs in general to do. And because of ethical incompleteness—like with human laws—we're going to have to have an expanding collection of such contracts.

But even though (particularly through my own efforts) we're beginning to have the kind of computational language necessary to specify a broad range of computational contracts, we realistically have to get much more experience with computational contracts in standard business and other situations before it makes sense to try setting them up for something as complex as global constraints on content selection systems.

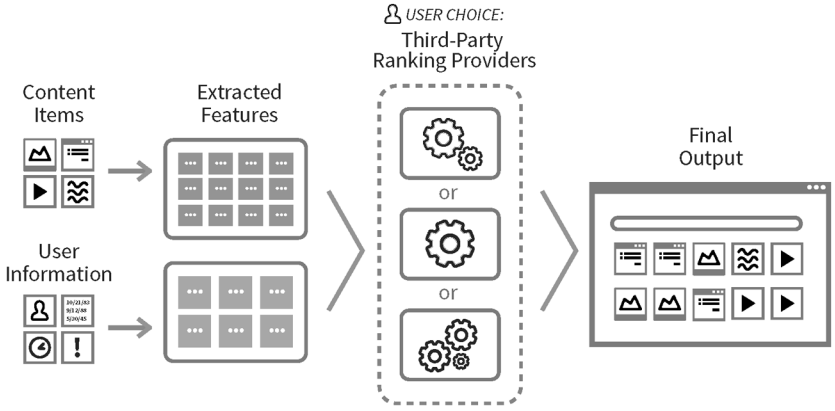
So, what can we do? I've not been able to see a viable, purely technical solution. But I have formulated two possible suggestions based on mixing technical ideas with what amount to market mechanisms.

The basic principle of both suggestions is to give users a choice about who to trust, and to let the final results they see not necessarily be completely determined by the underlying ACS business.

There's been debate about whether ACS businesses are operating as "platforms" that more or less blindly deliver content, or whether they're operating as "publishers" who take responsibility for content they deliver. Part of this debate can be seen as being about what responsibility should be taken for an AI. But my suggestions sidestep this issue, and in different ways tease apart the "platform" and "publisher" roles.

It's worth saying that the whole content platform infrastructure that's been built by the large ACS businesses is an impressive and very valuable piece of engineering—managing huge amounts of content, efficiently delivering ads against it, and so on. What's really at issue is whether the fine details of the ACS systems need to be handled by the same businesses, or whether they can be opened up. (This is relevant only for ACS businesses whose network effects have allowed them to serve a large fraction of a population. Small ACS businesses don't have the same kind of lock-in.)

*Suggestion A: Allow Users to Choose among Final Ranking Providers*




As I discussed earlier, the rough (and oversimplified) outline of how a typical ACS system works is that first features are extracted for each content item and each user. Then, based on these features, there's a final ranking done that determines what will actually be shown to the user, in what order, etc.

What I'm suggesting is that this final ranking doesn't have to be done by the same entity that sets up the infrastructure and extracts the features. Instead, there could be a single content platform but a variety of "final ranking providers", who take the features, and then use their own programs to actually deliver a final ranking.

Different final ranking providers might use different methods, and emphasize different kinds of content. But the point is to let users be free to choose among different providers. Some users might prefer (or trust more) some particular provider—that might or might not be associated with some existing brand. Other users might prefer another provider, or choose to see results from multiple providers.

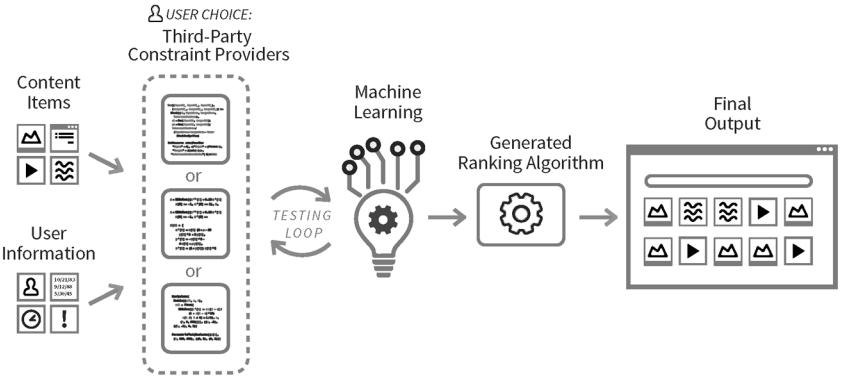
How technically would all this be implemented? The underlying content platform (presumably associated with an existing ACS business) would take on the large-scale information-handling task of deriving extracted features. The content platform would provide sufficient examples of underlying content (and user information) and its extracted features to allow the final ranking provider's systems to "learn the meaning" of the features.

When the system is running, the content platform would in real time deliver extracted features to the final ranking provider, which would then feed this into whatever system they have developed (which could use whatever automated or human selection methods they choose). This system would generate a ranking of content items, which would then be fed back to the content platform for final display to the user.

To avoid revealing private user information to lots of different providers, the final ranking provider's system should probably run on the content platform's infrastructure. The content platform would be responsible for the overall user experience, presumably providing some kind of selector to pick among final ranking providers. The content platform would also be responsible for delivering ads against the selected content.

Presumably the content platform would give a commission to the final ranking provider. If properly set up, competition among final ranking providers could actually increase total revenue to the whole ACS business, by achieving automated content selection that serves users and advertisers better.

*Suggestion B: Allow Users to Choose among Constraint Providers*




One feature of Suggestion A is that it breaks up ACS businesses into a content platform component, and a final ranking component. (One could still imagine, however, that a quasi-independent part of an ACS business could be one of the competing final ranking providers.) An alternative suggestion is to keep ACS businesses intact, but to put constraints on the results that they generate, for example forcing certain kinds of balance, etc.

Much like final ranking providers, there would be constraint providers who define sets of constraints. For example, a constraint provider could require that there be on average an equal number of items delivered to a user that are classified (say, by a particular machine learning system) as politically left-leaning or politically right-leaning.

Constraint providers would effectively define computational contracts about properties they want results delivered to users to have. Different constraint providers would define different computational contracts. Some might want balance; others might want to promote particular types of content, and so on. But the idea is that users could decide what constraint provider they wish to use.

How would constraint providers interact with ACS businesses? It's more complicated than for final ranking providers in Suggestion A, because effectively the constraints from constraint providers have to be woven deeply into the basic operation of the ACS system.

One possible approach is to use the machine learning character of ACS systems, and to insert the constraints as part of the "learning objectives" (or, technically, "loss functions") for the system. Of course, there could be constraints that just can't be successfully learned (for example, they might call for types of content that simply don't exist). But there will be a wide range of acceptable constraints, and in effect, for each one, a different ACS system would be built.

All these ACS systems would then be operated by the underlying ACS business, with users selecting which constraint provider—and therefore which overall ACS system—they want to use.

As with Suggestion A, the underlying ACS business would be responsible for delivering advertising, and would pay a commission to the constraint provider.

Although their detailed mechanisms are different, both Suggestions A and B attempt to leverage the exceptional engineering and commercial achievements of the ACS businesses, while diffusing current trust issues about content selection, providing greater freedom for users, and inserting new opportunities for market growth.

The suggestions also help with some other issues. One example is the banning of content providers. At present, with ACS businesses feeling responsible for content on their platforms, there is considerable pressure, not least from within the ACS businesses themselves, to ban content providers that they feel are providing inappropriate content. The suggestions diffuse the responsibility for content, potentially allowing the underlying ACS businesses not to ban anything but explicitly illegal content.

It would then be up to the final ranking providers, or the constraint providers, to choose whether or not to deliver or allow content of a particular character, or from a particular content provider. In any given case, some might deliver or allow

it, and some might not, removing the difficult all-or-none nature of the banning that's currently done by ACS businesses.

One feature of my suggestions is that they allow fragmentation of users into groups with different preferences. At present, all users of a particular ACS business have content that is basically selected in the same way. With my suggestions, users of different persuasions could potentially receive completely different content, selected in different ways.

While fragmentation like this appears to be an almost universal tendency in human society, some might argue that having people routinely be exposed to other people's points of view is important for the cohesiveness of society. And technically some version of this would not be difficult to achieve. For eexample, one could take the final ranking or constraint providers, and effectively generate a feature space plot of what they do.

Some would be clustered close together, because they lead to similar results. Others would be far apart in feature space—in effect representing very different points of view. Then if someone wanted to, say, see their typical content 80 percent of the time, but see different points of view 20 percent of the time, the system could combine different providers from different parts of feature space with a certain probability.

Of course, in all these matters, the full technical story is much more complex. But I am confident that if they are considered desirable, either of the suggestions I have made can be implemented in practice. (Suggestion A is likely to be somewhat easier to implement than Suggestion B.) The result, I believe, will be richer, more trusted, and even more widely used automated content selection. In effect both my suggestions mix the capabilities of humans and AIs—to help get the best of both of them—and to navigate through the complex practical and fundamental problems with the use of automated content selection.

Senator THUNE. Thank you, Mr. Wolfram.

Ms. Richardson.

**STATEMENT OF RASHIDA RICHARDSON, DIRECTOR OF POLICY RESEARCH, AI NOW INSTITUTE, NEW YORK UNIVERSITY**

Ms. RICHARDSON. Chairman Thune, Ranking Member Schatz, and members of the Subcommittee, thank you for inviting me to speak today.

My name is Rashida Richardson, and I'm the Director of Policy Research at the AI Now Institute at New York University, which is the first university research institute dedicated to understanding the social implications of artificial intelligence.

Part of my role includes researching the increasing reliance on AI and algorithmic systems and crafting policy and legal recommendations to address and mitigate the problems we identify in our research.

The use of data-driven technologies, like recommendation algorithms, predictive analytics, and inferential systems, are rapidly expanding in both consumer and government sectors. They determine where our children go to school, whether someone will receive Medicaid benefits, who is sent to jail before trial, which news articles we see, and which job seekers are offered an interview.

Thus, they have a profound impact on our lives and require immediate attention and action by Congress.

Though these technologies affect every American, they are primarily developed and deployed by a few powerful companies and therefore shaped by these companies' incentives, values, and interests. These companies have demonstrated limited insight into whether their products will harm consumers and even less experience in mitigating those harms.

So while most technology companies promise that their products will lead to broad societal benefits, there is little evidence to sup-

port these claims and, in fact, mounting evidence points to the contrary.

For example, IBM's Watson Super Computer was designed to improve patient outcomes but recently internal IBM documents showed it actually provided unsafe and erroneous cancer treatment recommendations. This is just one of numerous examples that have come to light in the last year showing the difference between the marketing companies used to sell these technologies and the stark reality of how these technologies ultimately perform.

While many powerful industries pose potential harms to consumers with new products, the industry-producing algorithmic and AI systems pose three particular risks that current laws and incentive structures fail to adequately address.

The first risk is that AI systems are based on compiled data that reflect historical and existing social and economic conditions. This data is neither neutral or objective. Thus, AI systems tend to reflect and amplify cultural biases, value judgments, and social inequities.

Meanwhile, most existing laws and regulations struggle to account for or adequately remedy these disparate outcomes as they tend to focus on individual acts of discrimination and less on systemic bias or bias encoded in the development process.

The second risk is that many AI systems and Internet platforms are optimization systems that prioritize technology companies' monetary interests and results in products being designed to keep users engaged while often ignoring social costs, like how the product may affect non-users environment to market.

A non-AI example of this logic and model is a slot machine. While a recent AI-based example is the navigation system Waze which was subject to public scrutiny following many incidents across the U.S. where the application redirected highway traffic through residential neighborhoods unequipped for the influx of vehicles which increased accidents and risks to pedestrians.

The third risk is that most of these technologies are black boxes, both technologically and legally. Technologically, they're black boxes because most of the internal workings are hidden away inside the companies. Legally, technology companies have struck accountability efforts through claims of proprietary or trade secret legal protections, even though there is no evidence that legitimate inspection, auditing, or oversight poses any competitive risk.

Controversies regarding emerging technologies are becoming increasingly common and show the harm caused by technologies optimized for narrow goals, like engagement, speed, and profit, at the expense of social and ethical considerations, like safety and accuracy.

We are at a critical moment where Congress is in a position to act on some of the most pressing issues and by doing so paving the way for a technological future that is safe, accountable, and equitable.

With these concerns in mind, I offer the following recommendations which are detailed in my written statement: require technology companies to waive trade secrecy and other legal claims that hinder oversight and accountability mechanisms, require public disclosure of technologies that are involved in any decision

about consumers by name and vendor, empower consumer protection agencies to apply truth-in-advertising laws, revive the Congressional Office of Technology Assessment to perform pre-market review and post-market monitoring of technologies, enhance whistle-blower protections for technology company employees that identify unethical and unlawful uses of AI or algorithms, require any transparency or accountability mechanisms to include detailed reporting of the full supply chain, and require companies to perform and publish algorithmic impact assessments prior to public use of products and services.

Thank you.

[The prepared statement of Ms. Richardson follows:]

PREPARED STATEMENT OF RASHIDA RICHARDSON, DIRECTOR OF POLICY RESEARCH, AI NOW INSTITUTE, NEW YORK UNIVERSITY

Chairman Thune, Ranking Member Schatz, and members of the Subcommittee, thank you for inviting me to speak today. My name is Rashida Richardson and I am the Director of Policy Research at the AI Now Institute at New York University. AI Now is the first university research institute dedicated to understanding the social implications of artificial intelligence ("AI"). Part of my role includes researching the increasing use of and reliance on data-driven technologies, including algorithmic systems and AI, and then designing and implementing policy and legal frameworks to address and mitigate problems identified in our research.

The use of data-driven technologies like recommendation algorithms, predictive analytics, and inferential systems, are rapidly expanding in both consumer and government sectors. These technologies impact consumers across many core domains—from health care to education to employment to the news and media landscape—and they affect the distribution of goods, services, and opportunities. Thus, they have a profound impact on people's lives and livelihoods. Though these technologies affect hundreds of millions of Americans, they are primarily developed and deployed by a few powerful private sector companies, and are therefore shaped by the incentives, values, and interests of these companies. Companies that arguably have limited insight into whether their products will harm consumers, and even less experience mitigating those harms or determining how to ensure that their technology products reflect the broader public interest. So while most technology companies promise that their products will lead to broad societal benefit, there is little evidence to support these claims. In fact, mounting evidence points to the contrary.[1] A recent notable example emerged when internal IBM documents showed its Watson supercomputer, which was designed to improve patient outcomes, provided unsafe and erroneous cancer treatment recommendations.[2] This is just one of numerous examples that have come to light in the last year, showing the difference between the marketing used to sell these technologies, and the reality of how these technologies ultimately perform.[3]

While many powerful industries pose potential harms to consumers with new products, the industry producing algorithmic and AI systems poses three particular risks that current laws and incentive structures fail to adequately address: (1) harm from biased training data, algorithms, or other system flaws that tend to reproduce historical and existing social inequities; (2) harm from optimization systems that prioritizes technology companies' interests often at the expense of broader societal interests; and (3) the use of 'black box' technologies that prevent public transparency, accountability, and oversight.

First, AI systems are trained on data sets that reflect historical and existing social and economic conditions. Thus, this data is neither neutral or objective, which

---

[1] *See, e.g.,* SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2013); LATONYA SWEENEY, *Discrimination in Online Ad Delivery,* 56 COMM. OF THE ACM 5, 44–45 (2013); Muhammad Ali *et al., Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes,* ARXIV (Apr. 19, 2019), *https://arxiv.org/pdf/1904.02095.pdf.*

[2] Casey Ross & Ike Swetlitz, *IBM's Watson Supercomputer Recommended 'unsafe and incorrect' Cancer Treatments, Internal Documents Show,* STAT (July 25, 2018), *https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/.*

[3] *See* AI Now Inst., *AI in 2018: A Year in Review* (Oct. 24, 2018), *https://medium.com/@AINowInstitute/ai-in-2018-a-year-in-review-8b161ead2b4e.*

leads to AI systems reflecting and amplifying cultural biases, value judgements, and social inequities. For instance, a recent study found that mechanisms in Facebook's ad targeting and delivery systems led to certain demographic segments of users being shown ads for housing and employment in a manner that aligns with gender and racial stereotypes.[4] Similarly, in 2018 Amazon chose to abandon an experimental hiring tool designed to help rank job candidates based on resumes. The tool turned out to be biased against women candidates because it learned from past gender-biased hiring preferences, and based on this, downgraded resumes from candidates who attended two all-women's colleges—along with those that contained even the word women's.[5] This outcome is particularly noteworthy because as one of the most well-resourced AI companies globally, Amazon was unable to mitigate or remedy this bias issue; yet, start-ups and other companies offering similar resume screening services proliferate.[6]

The use of flawed datasets and their biased outcomes create feedback loops that reverberate throughout society and are very difficult, if not impossible, to mitigate through traditional mathematical or technological techniques and audits.[7] Indeed, most existing laws and regulations struggle to account for or adequately remedy these challenges, as they tend to focus on individual acts of discrimination and less on systemic or computational forms of bias. For example, in recently-settled litigation against Facebook, the company tried to evade liability for the aforementioned discriminatory outcomes produced by its ad targeting and delivery platform. Facebook claimed it was simply a "neutral" platform under Section 230 of the Communications Decency Act's content safe harbors, despite recent research that demonstrated that the discriminatory outcomes were also attributed to Facebook's own, independent actions.[8]

Second, many consumer facing products are optimization systems, which are designed to prioritize technology companies' monetary interests and focus on scaling ideal outcomes rather than understanding potential flaws and adversarial behaviors in the design process. These skewed priorities in the absence of stringent design standards pose several social risks such as, optimizing Internet platforms for engagement, which can lead to profiling and mass manipulation, while also ignoring 'externalities,' like design tradeoffs that harm non-users and affected environments or markets.[9] For example, the navigation application, Waze,[10] has been subject to public and government scrutiny for instances where these consequences of optimization have actualized, including directing drivers towards forest fires during emergency evacuations, and redirecting highway commuters to residential streets, resulting in more accidents since these areas were unequipped to handle an influx of cars.[11] These outcomes are common, and rarely properly addressed, because tech-

---

[4] Muhammad Ali *et al., supra* note 1.

[5] Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women,* REUTERS (Oct. 9, 2018), *https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G;* Dave Gershgorn, *Companies are on the Hook If their Hiring Algorithms are Biased,* QUARTZ (Oct. 22, 2018), *https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased/.*

[6] *See, e.g.,* HIREVUE PLATFORM, (last visited June 17, 2019), *https://www.hirevue.com/products/hirevue-platform;* PYMETRICS, EMPLOYERS, (last visited June 17, 2019), *https://www.pymetrics.com/employers/;* APPLIED, APPLIED RECRUITMENT PLATFORM, (last visited June 17, 2019), *https://www.beapplied.com/features; See also* UPTURN, HELP WANTED: AN EXAMINATION OF HIRING ALGORITHMS, EQUITY, AND BIAS, 26–36 (Dec. 2019), *https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20—%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf.*

[7] *See* Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice,* 94 N.Y.U. L. REV. ONLINE 192 (2019).

[8] *Compare* Defendant's Motion to Dismiss, Onuoha v. Facebook, Inc., No. 16 Civ. 6440 (N.D. Cal. filed Apr. 3, 2017) *with* Muhammad Ali *et al., supra* note 1.

[9] Rebekah Overdorf *et al.,* Position Paper from NeurIPS 2018 Workshop in Montréal, Canada, *Questioning the Assumptions Behind Fairness Solutions,* ARXIV (Nov 27, 2018), *https://arxiv.org/pdf/1811.11293.pdf.*

[10] Waze is a subsidiary of Google. Google purchased the application in 2013.

[11] Samantha Raphelson, *New Jersey Town Restricts Streets from Commuters to Stop Waze Traffic Nightmare,* NPR (May 8, 2018), *https://www.npr.org/2018/05/08/609437180/new-jersey-town-restricts-streets-from-commuters-to-stop-waze-traffic-nightmare;* Christopher Weber, *Waze Causing LA Traffic Headaches, City Council Members Say,* ASSOCIATED PRESS (Apr. 17, 2018), *https://www.apnews.com/8a7e0b7b151c403a8d0089f9ed866863;* Jefferson Graham & Brett Molina, *Waze Sent Commuters Toward California Wildfires, Drivers Say,* USA TODAY (Dec. 7, 2017), *https://www.usatoday.com/story/tech/news/2017/12/07/california-fires-navigation-apps-like-waze-sent-commuters-into-flames-drivers/930904001/.*

nology companies lack incentives to comprehensively assess the negative effects of optimization within and outside a given technology, remedy their failures, and prioritize societal benefits (e.g.-incorporating the needs of all relevant stakeholders and environments).

Third, most of these technologies are "black boxes," both technologically and legally. Technologically, they are black boxes because most of the internal workings are hidden away inside the companies, hosted on their internal computer servers, without any regular means of public oversight, audit, or inspection to address consumer harm concerns. Legally, technology companies obstruct efforts of algorithmic accountability through claims of proprietary or "trade secret" legal protections, even though there is often no evidence that legitimate inspection, auditing, or oversight poses any competitive risks.[12] This means that neither government nor consumers are able to meaningfully assess or validate the claims made by companies. Some technology companies have suggested that the risks of emerging data driven technologies will eventually be mitigated by more technological innovation.[13] Conveniently, all of these remediations rely on us to trust the technology industry, which has few incentives or requirements to be accountable for the harms they produce or exacerbate.

Yet, history and current research demonstrates that there are significant limitations to relying solely on technical fixes and "self regulation" to address these urgent concerns.[14] Neither of these approaches allow room for public oversight and other accountability measures since technology companies remain the gatekeepers of important information that government and consumers would need to validate the utility, safety, and risks of these technologies. Ultimately, we are being asked to take technology companies' claims at face value, despite evidence from investigative journalists, researchers, and emerging litigation that demonstrate that these systems can, and do, fail in significant and dangerous ways.[15] To cite a few examples:

- Cambridge Analytica's exfiltration of Facebook user data exposed extreme breaches of consumer data privacy.
- Facebook's Ad-Targeting lawsuits and settlements highlighted ways the platform helped facilitate and possibly conceal discrimination.[16]

---

[12] AI Now Inst., Litigating Algorithms Workshop, June 2018, *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems* (Sept. 2018), *https://ainowinstitute.org/litigatingalgorithms.pdf* (highlighting lawsuits where vendors made improper trade secrecy claims); David S. Levine, *Can We Trust Voting Machines? Trade-Secret Law Makes it Impossible to Independently Verify that the Devices are Working Properly,* SLATE (Oct. 24, 2012), *https://slate.com/technology/2012/10/trade-secret-law-makes-it-impossible-to-independently-verify-that-voting-machines-work-properly.html* (describing how the application of trade secret law to e-voting machines threatens election integrity); Frank Pasquale, *Secret Algorithms Threaten the Rule of Law,* MIT TECHNOLOGY REVIEW (June 1, 2017), *https://www.technologyreview.com/s/608011/secret-algorithms-threaten-the-rule-of-law/*.

[13] Tom Simonite, *How Artificial Intelligence Can—and Can't—Fix Facebook,* WIRED (May 3, 2018), *https://www.wired.com/story/how-artificial-intelligence-canand-cantfix-facebook/;* F8 2018 Day 2 Keynote, Facebook for Developers (May 2, 2018), *https://www.facebook.com/FacebookforDevelopers/videos/10155609688618553/UzpfSTc0MTk2ODkwNzg6MTAxNTU4ODExNzI4MzQwNzk/;* Drew Harwell, *AI Will Solve Facebook's Most Vexing Problems, Mark Zuckerberg Says. Just Don't Ask When or How,* WASH. POST (Apr. 11, 2018), *https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how* ("he said, artificial intelligence would prove a champion for the world's largest social network in resolving its most pressing crises on a global scale") Stephen Shankland, *Google Working to Fix AI Bias Problems,* CNET (May 7, 2019), *https://www.cnet.com/news/google-working-to-fix-ai-bias-problems/*.

[14] Roy F. Baumeister & Todd F. Heatherton, *Self-Regulation Failure: An Overview,* 7 PSYCHOL. INQUIRY, no. 1, 1996 at 1; Stephanie Armour, *Food Sickens Millions as Company-Paid Checks Find It Safe,* BLOOMBERG (Oct. 11, 2012), *https://www.bloomberg.com/news/articles/2012-10-11/food-sickens-millions-as-industry-paid-inspectors-find-it-safe;* Andrew D. Selbst *et al., Fairness and Abstraction in Sociotechnical Systems,* 2019 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 59, *https://dl.acm.org/citation.cfm?id=3287598*

[15] *See* AI Now Inst., *supra* note 11; MEREDITH WHITTAKER ET AL., THE AI NOW REPORT 2018 (2018), *https://ainowinstitute.org/AI_Now_2018_Report.pdf;* Julia Angwin *et al., Machine Bias,* PROPUBLICA (May 23, 2016), *https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing;* Jaden Urbi, *Some Transgender Drivers are Being Kicked Off Uber's App,* CNBC (Aug. 13, 2018), *https://www.cnbc.com/2018/08/08/transgender-uber-driver-suspended-tech-oversight-facial-recognition.html;* U.N. EDUC., SCIENTIFIC, AND CULTURAL ORG., I'D BLUSH IF I COULD: CLOSING GENDER DIVIDES IN DIGITAL SKILLS THROUGH EDUCATION, U.N. Doc GEN/2019/EQUALS/1 REV 2 (2019); Paul Berger, *MTA's Initial Foray Into Facial Recognition at High Speed Is a Bust,* WALL ST. J. (Apr. 7, 2019), *https://www.wsj.com/articles/mtas-initial-foray-into-facial-recognition-at-high-speed-is-a-bust-11554642000*

[16] Brakkton Booker, *Housing Department Slaps Facebook With Discrimination Charge,* NPR (Mar. 28, 2019), *https://www.npr.org/2019/03/28/707614254/hud-slaps-facebook-with-hous-*

- The aftermath of the Christchurch Massacre and other deplorable terrorist attacks revealed how the engagement-driven design of Facebook, Youtube and other platforms have amplified misinformation, incited more violence, and increased radicalization.[17]
- Google's Dragonfly project demonstrated the intense secrecy around socially significant and ethically questionable corporate decisions.[18]

These types of controversies are increasingly common, and show the harm that technologies optimized for narrow goals like engagement, speed, or profit, at the expense of social and ethical considerations like safety or accuracy, can cause. And unlike other important and complex domains like health, education, criminal justice, and welfare, that each have their own histories, hazards, and regulatory frameworks, the technology sector has continued to expand and evolve without adequate governance, transparency, accountability, or oversight regimes.[19]

We are at a critical moment where Congress is in a position to act on some of the most pressing issues facing our social and economic institutions, and by doing so pave the way for a technological future that is safe, accountable, and equitable. Local, state and other national governments are taking action by performing domain specific inquiries to independently assess the actual benefits and risks of certain technologies. In some cases, they are creating transparency requirements or limitations on the use of technologies they deem too risky.[20]

Congress can build on this work and take actions that can help create necessary transparency, accountability and oversight mechanisms that empower relevant government agencies and even consumers to assess the utility and risks of certain technological platforms. The remainder of this testimony will highlight actions Congress can take to address specific concerns of data driven technologies.

## AI NOW'S POLICY RECOMMENDATIONS FOR CONGRESS

### 1. Require Technology Companies to Waive Trade Secrecy and Other Legal Claims That Hinder Oversight and Accountability Mechanisms

Corporate secrecy laws are a barrier to due process when technologies are used in the public sector. They can inhibit necessary government oversight and enforcement of consumer protection laws,[21] which contribute to the "black box effect," making it hard to assess bias, contest decisions, or remedy errors. Anyone procuring these technologies for use in the public sector should demand that vendors waive these claims before entering into any agreements. Additionally, limiting the use of these legal claims can help facilitate better oversight by state and Federal consumer protection agencies and enforcement of false and deceptive practice laws.

---

*ing-discrimination-charge;* Kenneth Terrell, *Facebook Reaches Settlement in Age Discrimination Lawsuits,* AARP (Mar. 20, 2019), *https://www.aarp.org/work/working-at-50-plus/info-2019/facebook-settles-discrimination-lawsuits.html*

[17] *Compare* Issie Lapowsky, *Why Tech Didn't Stop the New Zealand Attack from Going Viral,* WIRED (Mar. 15, 2019), *https://www.wired.com/story/new-zealand-shooting-video-social-media/ with* Natasha Lomas, *YouTube: More AI Can Fix AI-generated 'bubbles of hate',* TECHCRUNCH (Dec. 19, 2017), *https://techcrunch.com/2017/12/19/youtube-more-ai-can-fix-ai-generated-bubbles-of-hate/*

[18] Hamza Shaban, *Google Employees Go Public to Protest China Search Engine Dragonfly,* WASH. POST (Nov. 27, 2018), *https://www.washingtonpost.com/technology/2018/11/27/google-employees-go-public-protest-china-search-engine-dragonfly/*

[19] *See* WHITTAKER ET AL., *supra* note 14.

[20] Kate Conger *et al., San Francisco Bans Facial Recognition Technology,* N.Y. TIMES (May 14, 2019), *https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html;* 2018 N.Y.C LOCAL LAW NO. 49, *https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0;* H.B. 378, 91st Leg., Reg. Sess. (Vt. 2018), *https://legislature.vermont.gov/Documents/2018/Docs/ACTS/ACT137/ACT137%20As%20Enacted.pdf;* H.B. 2701, 191st Leg., Reg. Sess. (Ma. 2019), *https://malegislature.gov/Bills/191/HD951;* H.B.1655, 66th Leg., Reg. Sess. (Wa. 2019), *https://app.leg.wa.gov/billsummary?BillNumber=1655&Initiative=false&Year=2019;* Treasury Board of Canada Secretariat, Algorithmic Impact Assessment, (Mar. 8, 2019), available at: *https://open.canada.ca/data/en/dataset/748a97fb-6714-41ef-9fb8-637a0b8e0da1;* Mark Puente, *LAPD Ends Another Data-Driven Crime Program Touted to Target Violent Offenders,* L.A. TIMES (Apr. 12, 2019), *https://www.latimes.com/local/lanow/la-me-laser-lapd-crime-data-program-20190412-story.html;* Sam Schechner & Parmy Olson, *Facebook, Google in Crosshairs of New U.K. Policy to Control Tech Giants,* WALL ST. J. (Apr. 8, 2019), *https://www.wsj.com/articles/u-k-moves-to-end-self-regulation-for-tech-firms-11554678060,*

[21] Houston Fed'n of Teachers, Local 2415 v. Houston Indep. Sch. Dist., 251 F.Supp.3d 1168 (S.D. Tex. 2017).

**2. Require Public Disclosure of Technologies That Are Involved in Any Decisions About Consumers by Name and Vendor**

The need for meaningful insight and transparency is clear when you examine the way in which infrastructure owned by the major technology companies is repurposed by other businesses. Technology companies license AI application program interfaces (APIs), or "AI as a service" to third parties, who apply them to one or another purpose.[22] These business relationships, in which one organization repurposes potentially flawed and biased AI systems created by large technology companies, are rarely disclosed to the public, and are often protected under nondisclosure agreements. Even knowing that a given company is using an AI model created by Facebook, Google, or Amazon is currently hard, if not impossible, to ascertain. Thus, understanding the implications of bad, biased, or misused models is not currently possible. Consumers deserve to know about which data-based technologies are used to make decisions about them or affect the types of services, resources, or opportunities made available to them. Requiring disclosure of the type of technology used and which vendors it originates from will provide consumers with the kind of notice necessary to enforce their due process rights.

**3. Empower Consumer Protection Agencies to Apply "Truth in Advertising Laws" to Algorithmic Technology Providers**

Some technology companies and platforms serve as vendors to other companies or governments, often advertising their systems as capable of "objective" predictions, determinations, and decision-making without disclosing the risks and concerns, which include bias, discrimination, manipulation, and privacy harms. An example of this is the previously mentioned gender-biased hiring algorithm created by Amazon. Amazon shelved that project but imagine if they had instead sold it 'as a service' for other employers to use, such as companies like HireVue and Applied, who currently sell similar AI-enabled automated hiring and recruitment services. There are currently no legal mechanisms or requirements for companies who want to innovate their HR processes to determine whether these problems exist.

Though the Federal Trade Commission (FTC) does currently have jurisdiction to look for fraud and deception in advertising,[23] it has not yet looked at or tested many of these artificial intelligence, machine learning, or automated decision systems. Empowering the FTC to investigate and pursue enforcement through its existing authority is an urgent priority that Congress should support.

**4. Revitalize the Congressional Office of Technology Assessment to Perform Pre-Market Review and Post-Market Monitoring of Technologies**

Data driven technologies can pose significant risks to an individual's rights, liberties, opportunities and life; therefore, technologies that are likely to pose such risk should be subject to greater scrutiny before and after they are made available to consumers or government institutions. The Office of Technology Assessment existed from 1972 to 1995 to analyze these types of complex scientific and technical issues, and should be refunded to perform this function for Congress.[24] The Office could convene both technical and domain-specific experts (*e.g.,* practitioners and individuals likely to be affected by the technology) to assess whether certain technologies meet the claims made by technology companies, or whether they pose ethical risks warranting the imposition of technical or external restrictions before the technologies are publicly released. Once a product is made public, the Office should be empowered to perform periodic monitoring to ensure it continues to meet pre-market standards, and does not pose serious risks to the public.

---

[22] MICROSOFT AZURE, COGNITIVE SERVICES, *https://azure.microsoft.com/en-us/services/cognitive-services/* (last visited June 16, 2019); GOOGLE CLOUD, AI PRODUCTS, *https://cloud.google.com/products/ai/* (last visited June 16, 2019); FACEBOOK ARTIFICIAL INTELLIGENCE, TOOLS, *https://ai.facebook.com/tools/* (last visited June 16, 2019); AMAZON WEB SERVICES, MACHINE LEARNING *https://aws.amazon.com/machine-learning/* (last visited June 16, 2019); Matt Murphy & Steve Sloane, *The Rise of APIs,* TECHCRUNCH (May 21, 2016), *https://techcrunch.com/2016/05/21/the-rise-of-apis/*.

[23] FED. TRADE COMM'N, TRUTH IN ADVERTISING, *https://www.ftc.gov/news-events/media-resources/truth-advertising* (last visited June 16, 2019);

[24] U.S. GOVERNMENTAL ACCOUNTABILITY ORG., THE OFFICE OF TECHNOLOGY ASSESSMENT (Oct. 13, 1977), available at *https://www.gao.gov/products/103962;* Mike Masnick, *Broad Coalition Tells Congress to Bring Back the Office of Technology Assessment,* TECHDIRT (May 10, 2019), *https://www.techdirt.com/articles/20190510/14433442180/broad-coalition-tells-congress-to-bring-back-office-technology-assessment.shtml.*

**5. Enhanced Whistleblower Protections for Technology Company Employees That Identify Unethical or Unlawful Uses of AI or Algorithms**

Organizing and resistance by technology workers has emerged as a force for accountability and ethical decision making.[25] Many technology companies workforce are organized in silos, which can also contribute to opacity during product development. Thus whistleblowers can serve a crucial role in revealing problems that may not otherwise visible to relevant oversight bodies, or even to all of the workforce at a given firm. Whistleblowers in the technology industry can be a crucial component to government oversight and should have enhanced protections as they serve the public interest.

**6. Require Any Transparency or Accountability Mechanism To Include A Detailed Account and Reporting of The "Full Stack Supply Chain"**

For meaningful accountability, we need to better understand and track the component parts of an AI system and the full supply chain on which it relies: that means accounting for the origins and use of training data, test data, models, application program interfaces (APIs), and other infrastructural components over a product life cycle. This type of accounting for the "full stack supply chain" of AI systems is a necessary condition for a more responsible form of auditing. The full stack supply chain also includes understanding the true environmental and labor costs of AI systems, as well as understanding risks to non-users. This incorporates energy use, the use of labor in the developing world for content moderation and training data creation, and the reliance on clickworkers to develop and maintain AI systems.[26] This type of accounting may also incentivize companies to develop more inclusive product design that engages different teams and expertise earlier to better assess the implications throughout the product life cycle. Companies can submit these reports to the appropriate executive agency that regulates AI in the sector where the technology is being used.

**7. Require Companies to Perform and Publish Algorithmic Impact Assessments Prior to Public Use of Products and Services**

In 2018, AI Now published an Algorithmic Impact Assessment (AIA) framework, which offers a practical transparency and accountability framework for assessing the use and impact of algorithmic systems in government, including AI based systems.[27] AIAs draw directly from impact assessment frameworks in environmental protection, human rights, privacy, and data protection policy domains by combining public agency review and public input.[28] When implemented in government, AIAs provides both the agency and the public the opportunity to evaluate the potential impacts of the adoption of an algorithmic system before the agency has committed to its use. AIAs also require ongoing monitoring and review, recognizing that the dynamic contexts within which such systems are applied.

The framework has been adopted in Canada, and is being considered by local, state, and national governments globally.[29] Though it was originally proposed to address concerns associated with government use of automated decisions systems, the framework can also be integrated into private companies before a product or service is used by the public. This can provide companies opportunities to assess and pos-

---

[25] Daisuke Wakabayashi & Scott Shane, *Google Will Not Renew Pentagon Contract that Upset Employees,* N.Y. TIMES (June 1, 2018), *https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html;* Avie Schneider, *Microsoft Workers Protest Army Contract With Tech 'Designed to Help People Kill',* NPR (Feb. 22, 2019), *https://www.npr.org/2019/02/22/697110641/microsoft-workers-protest-army-contract-with-tech-designed-to-help-people-kill;* Mark Bergen & Nico Grant, *Salesforce Staff Ask CEO to Revisit Ties with Border Agency,* BLOOMBERG (June 25, 2018), *https://www.bloomberg.com/news/articles/2018-06-25/salesforce-employees-ask-ceo-to-revisit-ties-with-border-agency.*

[26] KATE CRAWFORD & VLADAN JOLER, AI NOW INST. & SHARE LAB, ANATOMY OF AN AI SYSTEM: THE AMAZON ECHO AS AN ANATOMICAL MAP OF HUMAN LABOR, DATA AND PLANETARY RESOURCES (Sept. 7, 2018), *https://anatomyof.ai.*

[27] AI NOW INST., ALGORITHMIC ACCOUNTABILITY POLICY TOOLKIT (Oct. 2018), *https://ainowinstitute.org/aap-toolkit.pdf*

[28] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact,* 104 CALIF. L. REV. 671 (2016).

[29] EUROPEAN PARLIAMENT PANEL FOR THE FUTURE OF SCI. AND TECH., A GOVERNANCE FRAMEWORK FOR ALGORITHMIC ACCOUNTABILITY AND TRANSPARENCY: STUDY (Apr. 4, 2019), *http://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2019)624262;* Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong., (1st Sess. 2019), *https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf;* AUTONOMISATION DES ACTEURS JUDICIAIRES PAR LA CYBERJUSTICE, CANADA TREASURY BOARD'S DIRECTED AUTOMATED DECISION-MAKING (Nov. 25, 2018), *https://www.ajcact.org/2018/11/25/canada-treasury-boards-directive-on-automated-decision-making/*

sibly mitigate adverse or unanticipated outcomes during the development process. It can also provide the government and public with greater transparency and strengthen existing consumer accountability mechanisms. It can encourage the development of safer and more ethical technologies by requiring companies to engage external stakeholders in review, who are likely to identify technical mistakes, design oversights, or even less obvious adverse outcomes.

Senator THUNE. Thank you, Ms. Richardson.

Let me start, Mr. Harris, with you. As we go about crafting consumer data privacy legislation in this Committee, we know that Internet platforms, like Google and Facebook, have vast quantities of data about each user.

What can these companies predict about users based on that data?

Mr. HARRIS. Thank you for the question.

So I think there's an important connection to make between privacy and persuasion that I think often isn't linked and maybe it's helpful to link that. You know, with Cambridge Analytica, that was an event in which, based on your Facebook likes, based on 150 of your Facebook likes, I could predict your political personality and then I could do things with that.

The reason I described in my opening statement that this is about an increasing asymmetry of power is that without any of your data, I can predict increasing features about you using AI.

There's a paper recently that with 80 percent accuracy, I can predict your same big five personality traits that Cambridge Analytica got from you without any of your data. All I have to do is look at your mouse movements and click patterns.

So, in other words, at the end of the poker face, your behavior is your signature and we can know your political personality. Based on tweet texts alone, we can actually know your political affiliation with about 80 percent accuracy.

Computers can calculate probably that you're homosexual before you might know that you're homosexual. They can predict with 95 percent accuracy that you're going to quit your job according to an IBM study. They can predict that you're pregnant. They can predict your micro expressions on your face better than a human being can. Micro expressions are your soft like reactions to things that are not very visible or are invisible. Computers can predict that.

As you keep going, you realize that you can start to deep fake things. You can actually generate a new synthetic piece of media, a new synthetic face or synthetic message that is perfectly tuned to these characteristics, and the reason why I opened the statement by saying we have to recognize that what this is all about is a growing asymmetry of power between technology and the limits of the human mind.

My favorite socio-biologist, E.O. Wilson, said, "The fundamental problem of humanity is that we have Paleolithic ancient emotions, we have medieval institutions, and we have god-like technology."

So we're chimpanzees with nukes and our Paleolithic brains are limited against the increasing exponential power of technology at predicting things about us. The reason why it's so important to migrate this relationship from being extractive, to get things out of you, to being a fiduciary is you can't have asymmetric power that is specifically designed to extract things from you, just like you

can't have again lawyers or doctors whose entire business model is to take everything they learned and sell it to someone else, except in this case the level of things that we can predict about you is far greater than actually each of those fields combined when you actually add up all the data that assembles more and more accurate voodoo doll of each of us and there are two billion voodoo dolls, by the way.

For one out of every four people on earth with YouTube and Facebook are more than two billion people.

Senator THUNE. Ms. Stanphill, in your prepared testimony, you note that companies like Google have a responsibility to ensure that product support users of digital well-being.

Does Google use persuasive technology, meaning technology that is designed to change people's attitudes and behaviors, and, if so, how do you use it, and do you believe that persuasive technology supports a user's digital well-being?

Ms. STANPHILL. Thank you, Senator.

No, we do not use persuasive technology at Google. In fact, our foremost principles are built around transparency, security, and control of our users' data. Those are the principles through which we design products at Google.

Senator THUNE. Dr. Wolfram, in your prepared testimony, you write that "It's impossible to expect any useful form of general explainability for automated content selection systems." If this is the case, what should policymakers require/expect of Internet platforms with respect to algorithm explanation or transparency?

Dr. WOLFRAM. I don't think that explaining how algorithms work is a great direction and the basic issue is if the algorithm's doing something really interesting, then you aren't going to be able to explain it because if you could explain it, it would be like saying you can jump ahead and say what it's going to do without letting it just do what it's going to do.

So it's kind of a scientific issue that if you're going to have something that is explainable, then it isn't getting to use the sort of full power of computation to do what it does.

So my own view, which is sort of disappointing for me at a technologist, is that you actually have to put humans in the loop and in a sense, the thing to understand about AI is we can automate many things about how things get done. What we don't get to automate is the goals of what we want to do.

The goals of what we want to do are not something that is sort of definable as an automatic thing. The goals of what we want to do is something that humans have to come up with and so I think the most promising direction is to think about breaking kind of the AI pipeline and figuring out where you can put into that AI pipeline the right level of kind of human input and my own feeling is the most promising possibility is to kind of insert—to leave the great value that's been produced by the current automatic content selection companies ingesting large amounts of data, being able to monetize large amounts of content, et cetera, but to insert a way for users to be able to choose who they trust about what finally shows up and then use the search results or whatever else.

I think that there are technological ways to make that kind of insertion that will actually, if anything, adds to the richness of po-

tential experience of the users and possibly even the financial returns for the market. complicated.

Senator THUNE. Very quickly, Ms. Richardson, what are your views about whether algorithm explanation or algorithm transparency are appropriate policy responses in counterresponse to Dr. Wolfram?

Ms. RICHARDSON. I think they're an interim step in that transparency is almost necessary to understand what these technologies are doing and to assess the benefits and risks, but I don't think transparency or even explainability is an end goal because I still think you're going to need some level of legal regulation to impose liability to bad or negligent actors who act in an improper manner, but also to incentivize companies to do the right thing or apply due diligence because in a lot of cases that I cited in my written testimony, there are sort of public relations disasters that happen on the back end and many of them could have been assessed or interpreted during the development process but companies aren't incentivized to do that.

So in some ways, transparency and explainability can give both legislators and the public more insight into these choices that companies are making to assess whether or not liability should be attached or different regulatory enforcement needs to be pursued.

Senator THUNE. Thank you.

Senator Schatz.

Senator SCHATZ. Thank you, Chairman. Thank you to the testifiers.

First, a yes or no question. Do we need more human supervision of algorithms on online platforms, Mr. Harris?

Mr. HARRIS. Yes.

Ms. STANPHILL. Yes.

Dr. WOLFRAM. Yes, though I would put some footnotes.

Senator SCHATZ. Sure.

Ms. RICHARDSON. Yes, with footnotes.

Senator SCHATZ. So I want to follow up on what Dr. Wolfram said in terms of the unbreakability of these algorithms and the lack of transparency that is sort of built into what they are foundationally, and the reason I think that's an important point that you're making, which is that you need a human circuit breaker at some point to say no, I choose not to be fed things by an algorithm, I choose to jump off of this platform. That's one aspect of humans acting as a circuit breaker.

I'm a little more interested in the human employee either at the line level or the human employee at the supervisory level who takes some responsibility for how these algorithms evolve over time.

Ms. Richardson, I want you to maybe speak to that question because it seems to me as policymakers that's where the sweet spot is, is to find an incentive or a requirement where these companies will not allow these algorithms to run essentially unsupervised and not even understood by the highest echelons of the company, except in their output, and so, Ms. Richardson, can you help me to flesh out what that would look like in terms of enabling human supervision?

Ms. RICHARDSON. So I think it's important to understand some of the points about the power asymmetry that Mr. Harris mentioned because I definitely do think we need a human in the loop, but we also need to be cognizant of who actually has power in those dynamics and that you don't necessarily want a front line employee taking full liability for a decision or a system that they had no input in the design or even in their current sort of position in using it.

So I think it needs to go all the way up in that if you're thinking about liability or responsibility in any form, it needs to attach at those who are actually making decisions about the goals, the designs, and ultimately the implementation and use of these technologies, and then figuring out what are the right pressure points or incentive dynamics to encourage companies or those making those decisions to make the right choice that benefits society.

Senator SCHATZ. Yes, I think that's right. I think that none of this ends up coming to much unless the executive level of these companies feel a legal and financial responsibility to supervise these algorithms.

Ms. Stanphill, I was a little confused by one thing you said. Did you say Google doesn't use persuasive technology?

Ms. STANPHILL. That is correct, sir.

Senator SCHATZ. Mr. Harris, is that true?

Mr. HARRIS. It's complicated, persuasion is happening all throughout the ecosystem. In my mind, by the way, this is less about accusing one company, Google or Facebook. It's about understanding that every company——

Senator SCHATZ. I get that, but she's here and she just said that they don't use persuasive technology, and I'm trying to figure out are you talking about just the Google suite of products? You're not talking about YouTube or are you saying in the whole Alphabet pantheon of companies, you don't use persuasive technology because either I misunderstand your company or I misunderstand the definition of persuasive technology. Can you help me to understand what's going on here?

Ms. STANPHILL. Sure. With respect to my response, Mr. Senator, it is related to the fact that dark patterns and persuasive technology is not core to how we design our products at Google, which are built around transparency.

Senator SCHATZ. But you're talking about YouTube or the whole family of companies?

Ms. STANPHILL. The whole family of companies, including YouTube.

Senator SCHATZ. You don't want to clarify that a little further?

Ms. STANPHILL. We build our products with privacy, security, and control for the users. That is what we build for, and ultimately this builds a lifelong relationship with the user which is primary. That's our——

Senator SCHATZ. I don't know what any of that meant.

Ms. Richardson, can you help me?

Ms. RICHARDSON. I think part of the challenges, Mr. Harris mentioned it, is how you're defining persuasive in that both of us mentioned that a lot of these systems and Internet platforms are a form of an optimization system which is optimizing for certain

goals and there you could say that is a persuasive technology which is not accounting for a certain social risk, but I think there's a business incentive to not—to take a more narrow view of that definition.

So it's like I can't speak for Google because I don't work for them, but I think the reason you're confused is because you may need to clarify definitions of what is actually persuasive in the way that you're asking the question and what is Google suggesting doesn't have persuasive characteristics in their technologies.

Senator SCHATZ. Thank you.

Senator THUNE. Thank you, Senator Schatz.

Senator Fischer.

**STATEMENT OF HON. DEB FISCHER,**
**U.S. SENATOR FROM NEBRASKA**

Senator FISCHER. Thank you, Mr. Chairman.

Mr. Harris, as you know, I've introduced the DETOUR Act with Senator Warner to curb some of the manipulative user interfaces. We want to be able to increase transparency, especially when it comes to behavioral experimentation online. Obviously we want to make sure children are not targeted with some of the dark patterns that are out there.

In your perspective, how do dark patterns thwart that user autonomy online?

Mr. HARRIS. Yes, so persuasion is so invisible and so subtle. In fact, oftentimes we're criticized on the use of language. We say we're crawling down the brain stem. People think that you're overreacting, but it's a design choice.

So my background, I studied with a lab called the Persuasive Technology Lab at Stanford that taught engineering students essentially about this whole field and my friends in the class were the founders of Instagram and Instagram is a product invented—well, copied Twitter actually in the technique of, well, you could call it dark pattern of the number of followers that you have to get people to follow each other. So there's a follow button on each profile and that's meant—I mean that doesn't seem so dark. That's what so insidious about it. You're giving people a way to follow each other's behavior.

But what it actually is doing is an attempt to cause you to come back every day because now you want to see do I have more followers now than I did yesterday.

Senator FISCHER. And how are these platforms then getting our personal information? How much choice do we really have? I thought the doctor's comment about that the goals we want to do as humans, that, you know, we have to get involved in this, but then your introductory comments are basically, I think, telling us that everything about us is already known.

So it wouldn't be really hard to manipulate where our goals—what they even want to be at this point, right?

Mr. HARRIS. The goal is to subvert your goals. I'll give you an example. If you say I want to delete my Facebook account, if you hit delete, it puts up a screen that says are you sure you want to delete your Facebook account? The following friends will miss you and it puts up faces of certain friends.

Now am I asking to know which friends will miss me? No. Does Facebook ask those friends are they going to miss me if I leave? No. They're calculating which of the five faces would be most likely to get you to cancel and not delete your Facebook account. So that's a subtle and invisible dark pattern that's meant to persuade behavior.

I think another example you're trying to get at in your opening question is if you consent to giving your data to Facebook or your location, and oftentimes, you know, there will be a big blue button, which they have a hundred engineers behind the screen split testing all the different colors and variations and arrangements on where that button should be, and then a very, very, very small gray link that people don't even know is there and so what we're calling a free human choice is a manipulated choice and again it's just like a magician. They're saying pick a card, any card, but in fact there's an asymmetry of power.

Senator FISCHER. When you're on the Internet and you're trying to look something up and you have this deal pop up on your screen, this is so irritating, and you have to hit OK to get out of it because you don't see the other choice on the screen. As you said, it's very light. It's gray. But now I know if I hit OK, this is going to go on and on and whoever is going to get more and more information about me. They're really invading my privacy, but I can't get rid of this screen otherwise, unless you turn off your computer and start over, right?

Mr. HARRIS. There are all sorts of ways to do this. If I'm a persuader and I really want you to hit okay on my dialogues, so I can get your data. I'll wait until the day that you're in a rush to get the address to that place you're looking for and that's the day that I'll put up the dialogue that says, hey, were you willing to give me your information and now of course you're going to say fine, yes, whatever, because in persuasion there's something called hot states and cold states.

When you're in a hot state or an immediate impulsive state, it's very easy to persuade someone than versus when they're in a cold, calm, and reflective state, and technology can actually either manufacture or wait till you're in those hot states.

Senator FISCHER. So how do we protect ourselves and our privacy and what role does the Federal Government have to play in this, besides getting our bill passed?

Mr. HARRIS. I mean, at the end of the day, the reason why we go back to the business model is it is about alignment of interests. You don't want a system of asymmetric power that is designed to manipulate people. You're always going to have that insofar as the business model is one of manipulation as opposed to regenerative, meaning you have a subscription-style relationship.

So I would say Netflix probably has many fewer dark patterns because it's in a subscription relationship with its users. When Facebook says that, you know, how else could we give people this free service, well, it's like a priest whose entire business model is to manipulate people, saying, well, how else can I serve so many people?

Senator FISCHER. Yes. How do we keep our kids safe?

Mr. HARRIS. There's so much to that. I think what we need is a mass public awareness campaign so people understand what's going on. One thing I have learned is that if you tell people this is bad for you, they won't listen. If you tell people this is how you're being manipulated—no one wants to feel manipulated.

Senator FISCHER. Thank you.

Senator THUNE. Thank you, Senator Fischer.

Senator Blumenthal.

**STATEMENT OF HON. RICHARD BLUMENTHAL, U.S. SENATOR FROM CONNECTICUT**

Senator BLUMENTHAL. Thank you, Mr. Chairman, and thank you to all of you for being here today.

You know, I was struck by what Senator Schatz said in his opening statement. Algorithms are not only running wild but they are running wild in secrecy. They are cloaked in secrecy in many respects from the people who are supposed to know about them.

Ms. Richardson referred to the black box here. That black box is one of our greatest challenges today and I think that we are at a time when algorithms, AI, and the exploding use of them is almost comparable to the time of the beginnings of atomic energy in this country.

We now have an Atomic Energy Commission. Nobody can build bombs, nuclear bombs in their backyard because of the dangers of nuclear fission and fusion, which is comparable, I think, to what we have here, systems that are in many respects beyond our human control and affecting our lives in very direct extraordinarily consequential terms beyond the control of the user and maybe the builder.

So on the issue of persuasive technology, I find, Ms. Stanphill, your contention that Google does not build systems with the idea of persuasive technology in mind is somewhat difficult to believe because I think Google tries to keep people glued to its screens at the very least. That persuasive technology is operative. It's part of your business model, keep the eyeballs.

It may not be persuasive technology to convert them to the far left or the far right. Some of the content may do it, but at the very least, the technology is designed to promote usage.

YouTube's recommendation system has a notorious history of pushing dangerous messages and content promoting radicalization, disinformation, and conspiracy theories.

Earlier this month, Senator Blackburn and I wrote to YouTube on reports that its recommendation system was promoting videos that sexualized children, effectively acting as a shepherd for pedophiles across its platform.

Now you say in your remarks that you've made changes to reduce the recommendation of content that "violates our policies or spreads harmful misinformation," and according to your account, the number of views from recommendations for these videos has dropped by over 50 percent in the United States. I take those numbers as you provided them.

Can you tell me what specific steps you have taken to end your recommendation system's practice of promoting content that sexualizes children?

Ms. STANPHILL. Thank you, Senator.

We take our responsibility to supporting child safety online extremely seriously. Therefore, these changes are in effect and as you stated, these have had a significant impact,——

Senator BLUMENTHAL. But what specifically?

Ms. STANPHILL.—resulting in actually changing which content appears in the recommendations. So this is now classified as borderline content. That includes misinformation and child exploitation content.

Senator BLUMENTHAL. You know, I am running out of time. I have so many questions. But I would like each of the witnesses to respond to the recommendations that Ms. Richardson has made, which I think are extraordinarily promising and important.

I'm not going to have time to ask you about them here, but I would like the witnesses to respond in writing, if you would, please, and, second, let me just observe on the topic of human supervision, I think that human supervision has to be also independent supervision.

On the topic of arms control, we have a situation here where we need some kind of independent supervision, some kind of oversight and, yes, regulation. I know it's a dirty word these days in some circles, but protection will require intervention from some independent source here. I don't think trust me can work anymore.

Thank you, Mr. Chairman.

Senator THUNE. Thank you, Senator Blumenthal.

Senator Blackburn.

## STATEMENT OF HON. MARSHA BLACKBURN, U.S. SENATOR FROM TENNESSEE

Senator BLACKBURN. Thank you, Mr. Chairman, and thank you to our witnesses. We appreciate that you are here and I enjoyed visiting with you for a few minutes before the hearing began.

Mr. Wolfram, I want to pick up where we had discussed. In your testimony, computational irreducibility and look at that for just a moment. As we talk about this, does it make algorithmic transparency sound increasingly elusive and would you consider that moving toward that transparency is a worthy goal or should we be asking another question?

Dr. WOLFRAM. Yes, I think, you know, there are different meanings to transparency. You know, if you are asking tell me why the algorithm did this, versus that, that's really hard, and if we really want to be able to answer that, we're not going to be able to have algorithms that do anything very powerful because in a sense by being able to say this is why it did that, well, we might as well just follow the path that we used to explain it rather than have it do what it needed to do itself.

Senator BLACKBURN. So the transparency, what we can't do is try to get a pragmatic result?

Dr. WOLFRAM. No, we can't go inside. We can't open up the hood and say why did this happen, and that's why I think the other problem is knowing what you want to have happen, like you say this algorithm is bad, this algorithm gives bad recommendations, what do you mean by bad recommendations?

We have to be able to define something that says, oh, the thing is biased in this way, the thing is producing content we didn't like. You know, you have to be able to give a way to define those bad things.

Senator BLACKBURN. All right. Ms. Richardson, I can see you're making notes and want to weigh in on this, but you also talked about compiled data and encoded bias and getting the algorithm to yield a certain result.

So let's say you build this algorithm and you build this box to contain this dataset or to make certain that it is moving this direction. Then as that algorithm self-replicates and moves forward, does it move further that direction or does data inform it and pull it a separate direction if you're building it to get it to yield a specific result?

Ms. RICHARDSON. So it depends on what type of technical system we're talking about, too, but to unpack what I was saying is the problem with a lot of these systems is they're based on datasets which reflect all of our current conditions which also means any imbalances in our conditions.

So one of the examples that I gave in my written testimony referenced Amazon hiring algorithm which was found to have gender disparate outcomes and that's because it was learning from prior hiring practices and there are also examples of other similar hiring algorithms, one of which found that if you have the name Gerard and you played lacrosse, you had a better chance of getting a job interview and there, it's not necessarily that the correlation between your name being Gerard and playing lacrosse means that you're necessarily a better employee than anyone else, it's simply looking at patterns and the underlying data, but it doesn't necessarily mean that the patterns that the system is seeing actually reflects reality or in some cases it does and it's not necessarily how we want to view reality and instead shows the skew that we have in society.

Senator BLACKBURN. Got it. OK. Mr. Wolfram, you mentioned in your testimony there could be a single content platform but a variety of final ranking providers.

Are you suggesting that it would be wise to prohibit companies from using cross-business data flows?

Dr. WOLFRAM. I'm not sure how that relates to—I mean, you know, the thing that I think is the case is it is not necessary to have the final ranking of content. There's a lot of work that has to be done to get content ready to be finally ranked for a newsfeed or for search results and so on. That's a lot of heavy lifting.

The choice which is made often separately for each user about how to finally rank content I don't think has to be made by the same entity and I think if you break that apart, you kind of change the balance between what is controllable by users and what is not.

I don't think it's realistic to—I think—yes, I mean, I would like to say that one of the questions about, you know, a dataset implies certain things. We don't like what that implies and so on.

One of the challenges is to define what we actually want and one of the things that's happening here is that because these are AI systems, computational systems, we have to define much more precisely what we want than we've had to do before. So it's necessary

to kind of write these computational rules and that's a tough thing to do and it's something which cannot be done by a computer and it can't be even be necessarily done from prior data. It's something people like you guys have to decide what to do about it.

Senator BLACKBURN. Thank you.

Mr. Chairman, I would like unanimous consent to enter the letter that Senator Blumenthal and I sent earlier this month and thank you.

[The letter referred to follows:]

UNITED STATES SENATE
*Washington, DC, June 6, 2019*

Ms. SUSAN WOJCICKI,
CEO,
YouTube,
San Bruno, CA.

Dear Ms. Wojcicki:

We write with concern that YouTube has repeatedly failed to address child sexual exploitation and predatory behavior on its platform. Since February, bloggers, journalists, and child safety organizations have raised alarm over a chilling pattern of pedophiles and child predators using YouTube to sexualize and exploit minors.[1] Despite promises of change, the *New York Times* now reports that YouTube's recommendation mechanism continues to actively and automatically push sensitive videos involving children. The sexualization of children through YouTube's recommendation engine represents the development of a dangerous new kind of illicit content meant to avoid law enforcement detection. Action is overdue; YouTube must act forceful and swiftly to end this disturbing risk to children and society.

In February, video blogger Mark Watson published a series of videos demonstrating that the platform is being used for "facilitating pedophiles ' ability to connect with each-other, trade contact info, and link to actual child pornography in the comments."[2] At that time, YouTube's video recommendation system was found to promote increasingly sexualized content involving minors. Below those videos were often comments attempting to contact and groom children.[3]

Shockingly, those comments also concealed a network of predators, providing each other timestamps and links to sensitive and revealing moments within videos—such as those of children wearing bathing suits or dressing. Effectively, YouTube's comments have fostered a ring of predators trafficking in the sexualization and exploitation of innocent videos of minors. In response, YouTube disabled comments for videos involving children.[4]

Recent research has found that even without the pedophilic comments on videos involving minors, YouTube's recommendation system is guiding child predators to find sensitive and at-risk videos of children. Researchers from the Berkman Klein Center for Internet and Society found when users started with one risky video, the recommendation system would start "showing the video to users who watched other videos of prepubescent, partially clothed children."[5] Researchers found that YouTube viewers would be provided increasingly extreme recommendations over time—more sexualized content and younger women, including partially clothed children. This pattern appears to be related to how YouTube learns recommendations: if a subset of viewers are using the platform to seek suggestive videos of children,

---

[1] MattsWhatItIs. "Youtube Is Facilitating the Sexual Exploitation of Children, and It's Being Monetized (2019)." YouTube. February 17, 2019. *https://www.youtube.com/watch?time_continue=4&v=O13G5A5w5P0.*

[2] Alexander, Julia. "YouTube Still Can't Stop Child Predators in Its Comments." The Verge. February 19, 2019. *https://www.theverge.com/2019/2/19/18229938/youtube-child-exploitation-recommendation-algorithm-predators.*

[3] Orphanides, K.G. "On YouTube, a Network of Paedophiles Is Hiding in Plain Sight." WIRED. June 03, 2019. *https://www.wired.co.uk/article/youtube-pedophile-videos-advertising.*

[4] Wakabayashi, Daisuke. "YouTube Bans Comments on Videos of Young Children in Bid to Block Predators." The New York Times. February 28, 2019. *https://www.nytimes.com/2019/02/28/technology/youtube-pedophile-comments.html?module=inline.*

[5] Fisher, Max, and Amanda Taub. "On YouTube's Digital Playground, an Open Gate for Pedophiles." The New York Times. June 03, 2019. *https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html.*

it will begin to reproduce that pattern to find and recommend other suggestive videos. With YouTube asleep at the wheel, predators have taken the reins.

As members of the Senate Judiciary Committee and the Committee on Commerce, Science, and Transportation, we are dismayed at YouTube's slow and inadequate response to repeated stories about child exploitation of its platform. Once again, YouTube has promised change, including to reduce the risk from its recommendation system.[6] However, despite past promises to address its recommendation system, YouTube has continued steering of users into exploitative content.[7] This is not merely an issue with algorithms, YouTube has failed to take down videos from child abusers, even highly-visible cases and after repeated reports.[8] YouTube must do all it can to prevent the exploitation of children, starting with the design of its algorithms and administration of its products.

Given the sensitivity and seriousness of the matter, we request a written response to the following questions by June 25, 2019:

1. Who at YouTube is in charge of coordinating its efforts to combat child sexual exploitation and protect the safety of minors on the platform? How is that individual included in design decisions and the product lifecycle?
2. What specific criteria (such as the correlation of previously watched or liked videos) does YouTube's content recommendation system use in order to recommend videos involving children? Does it take any measures to prevent content involving minors from being recommended after sexualized videos or based on patterns from predatory users?
3. In its June 3, 2019 announcement, YouTube offers that it will reduce recommendations for "videos featuring minors in risky situations." How will YouTube deem whether a video puts a child at risk? What steps will be taken when it identifies such a video?
4. Will YouTube disable recommendations for videos involving minors until it can ensure its systems no longer facilitates the sexualization and exploitation of children?
5. Will YouTube commit to an independent audit of how its content recommendation systems and other functions of its platform addresses and prevents predatory practices against children?
6. What policies does YouTube have or is considering to proactively address videos involving known child sexual predators or individuals on publicly available sex-offender databases?

Thank you for your attention to these important issues. We look forward to your response.

Sincerely,

/s/ *Richard Blumenthal*
RICHARD BLUMENTHAL
United States Senate

/s/ *Marsha Blackburn*
MARSHA BLACKBURN
United States Senate

Senator BLACKBURN. I know my time has expired, but I will just simply say to Ms. Stanphill that the evasiveness on answering Senator Blumenthal's question about what they are doing is inadequate when you look at the safety of children online just to say that you're changing the content that appears in the recommended list is inadequate.

Mr. Harris, I will submit a question to you about what we can look at on platforms for combating some of this bad behavior.

Senator THUNE. Thank you, Senator Blackburn.

Senator Peters.

---

[6] "An Update on Our Efforts to Protect Minors and Families." Official YouTube Blog. June 03, 2019. *https://youtube.googleblog.com/2019/06/an-update-on-our-efforts-to-protect.html.*

[7] "Continuing Our Work to Improve Recommendations on YouTube." Official YouTube Blog. January 25, 2019. *https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html.*

[8] Pilon, Mary. "Larry Nassar's Digital Ghosts." The Cut. May 29, 2019. *https://www.thecut.com/2019/05/why-wouldnt-youtube-remove-a-disturbing-larry-nassar-video.html.*

**STATEMENT OF HON. GARY PETERS,**
**U.S. SENATOR FROM MICHIGAN**

Senator PETERS. Thank you, Mr. Chairman, and thank you to our witnesses for a very fascinating discussion.

I'd like to address an issue that I think is of profound importance to our democratic republic and that's the fact that in order to have a vibrant democracy, you need to have an exchange of ideas and an open platform and certainly part of the promise of the Internet as it was first conceived is that we'd have this incredible universal commons where a wide range of ideas would be discussed and debated. It would be robust and yet it seems as if we're not getting that. We're actually getting more and more siloed.

Dr. Wolfram, you mentioned how people can make choices and they can live in a bubble but at least it would be their bubble that they get to live in, but that's what we're seeing throughout our society. As polarization increases, more and more folks are reverting to tribal-type behavior.

Mr. Harris, you talked about our medieval institutions and Stone Age minds. Tribalism was alive and well in the past and we're seeing advances in technology in a lot of ways bring us back into that kind of tribal behavior.

So my question is to what extent is this technology actually accelerating that and is there a way out? Yes, Mr. Harris.

Mr. HARRIS. Yes, thank you. I love this question. There's a tendency to think here that this is just human nature. Now that's just people are polarized and this is just playing out. It's a mirror. It's holding up a mirror to society.

But what it's really doing is it's an amplifier for the worst parts of us. So in the race to the bottom of the brain stem to get attention, let's take an example like Twitter, it's calculating what is the thing that I can show you that will get the most engagement and it turns out that outrage, moral outrage gets the most engagement. So it was found in a study that for every word of moral outrage that you add to a tweet, it increases your retweet rate by 17 percent.

So, in other words, you know, the polarization of our society is actually part of the business model. Another example of this is that shorter, briefer things work better in an attention economy than long complex nuanced ideas that take a long time to talk about and so that's why you get 140 characters dominating our social discourse but reality and the most important topics to us are increasingly complex, while we can say increasingly simple things about them.

That automatically creates polarization because you can't say something simple about something complicated and have everybody agree with you. People will by definition misinterpret and hate you for it and then it has never been easier to retweet that and generate a mob that will come after you and this has created call-out culture and chilling effects and a whole bunch of other subsequent effects in polarization that are amplified by the fact that these platforms are rewarded to give you the most sensational stuff.

One last example of this is on YouTube, let's say we actually equalize—I know there are people here concerned about equal representation on the left and the right in media.

Let's say we get that perfectly right. As recently as just a month ago on YouTube, if you did a map of the top 15 most frequently mentioned verbs or keywords in the recommended videos, they were "hate, debunks, obliterates, destroys." In other words, Gordon Peterson destroys social justice warrior in video.

So that kind of thing is the background radiation that we're dosing two billion people with and you can hire content moderators in English and start to handle the problem, as Ms. Stanphill said, but the problem is that two billion people in hundreds of languages are using these products. How many engineers at YouTube speak the 22 languages in India where there's an election coming up? So that's some context on that.

Senator PETERS. Well, that was a lot of context. Fascinating. I'm running out of time, but I took particular note in your testimony when you talked about how technology will eat up elections and you were referencing, I think, another writer on that issue.

In the remaining brief time I have, what's your biggest concern about the 2020 elections and how technology may eat up this election coming up?

Mr. HARRIS. Yes, that comment was another example of we used to have protections that technology took away. We used to have equal price campaign ads so that it cost the same amount on Tuesday night at 7 p.m. for any candidate to run an election.

When Facebook gobbles up that part of the media, it just takes away those protections. So there's now no equal pricing.

Here's what I'm worried about. I'm mostly worried about the fact that none of these problems have been solved. The business model hasn?t changed and the reason why you see a Christchurch event happen and the video just show up everywhere or, you know, any of these examples, fundamentally there's no easy way for these platforms to address this problem because the problem is their business model.

I do think there are some small interventions, like fast lanes for researchers, accelerated access for people who are spotting disinformation, but the real problem, another example of software eating the world, is that instead of NATO or the Department of Defense protecting us in a global information warfare, we have a handful of 10 or 15 security engineers at Facebook and Twitter and they were woefully unprepared, especially in the last election, and I'm worried that they still might be.

Senator PETERS. Thank you.

Senator THUNE. Thank you, Senator Peters.

Senator Johnson.

**STATEMENT OF HON. RON JOHNSON,**
**U.S. SENATOR FROM WISCONSIN**

Senator JOHNSON. Thank you, Mr. Chairman.

Mr. Harris, I agree with you when you say that our best line of defense as individuals is exposure. People need to understand that they are being manipulated and a lot of this hearing has been talking about manipulation algorithms, artificial intelligence.

I want to talk about the manipulation by human intervention, human bias. You know, we don't allow or we certainly put restrictions through the FCC on an individual owning their ownership of

TV stations, radio stations, newspapers, because we don't want that monopoly of content in the community, much less, you know, Facebook, Google accessing billions of people, hundreds of millions of Americans.

So I had staff on Instagram go to the Politico account and, by the way, I have a video of this, so I'd like to enter that into the record.

They hit follow and this is what the list they are given and this is in exact order and I'd asked the audience and the witnesses to just see if there's a conservative in here, how many there are. Here's the list, Elizabeth Warren, Kamala Harris, *New York Times, Huffington Post,* Bernie Sanders, *CNN Politics, New York Times Opinion, NPR Economist,* Nancy Pelosi, *The Daily Show, Washington Post Covering POTUS, NBC, Wall Street Journal,* Pete Buttigieg, *Time New Yorker, Reuters,* Southern Poverty Law Center, Kirsten Gillibrand, *The Guardian, BBC News,* ACLU, Hillary Clinton, Joe Biden, Beto O'Rourke, *Real Time with Bill Maher, C-SPAN, SNL,* Pete Souza, United Nations, Guardian, *HuffPost Women's, Late Show with Steven Colbert,* Moveon.org, *Washington Post Opinion, USAToday, New Yorker,* Williams Marsh, *Late Night with Seth Meyers, The Hill, CBS,* Justin Trudeau. It goes on.

These are five conservative staff members. If they're really algorithms shuffling the content that they might actually want or they would agree with, you'd expect you'd see maybe *Fox News, Breitbart, Newsmax.* You might even see like a really big name like Donald Trump and there wasn't.

So my question is who's producing that list? Is that Instagram? Is that the *Politico* site? How is that being generated? I have a hard time feeling that's generated or being manipulated by an algorithm or by AI.

Mr. HARRIS. I don't know any—I'd be really curious to know what the click pattern was that—in other words, you open up an Instagram account and it's blank and you're saying that if you just ask who do I follow——

Senator JOHNSON. You hit follow and you're given suggestions for you to follow.

Mr. HARRIS. Yes, I mean, I honestly have no idea how Instagram ranks those things, but I'd be very curious to know what the original clicks were that produced that list.

Senator JOHNSON. Can anybody else explain that? I mean, I don't believe that's AI trying to give content to a conservative staff member of things they may want to read. I mean, this to me looks like Instagram, if they're actually the ones producing that list, trying to push a political bias.

Mr. Wolfram, you seem to want to weigh in.

Dr. WOLFRAM. You know, the thing that will happen is if there's no other information, it will tend to be just where there is the most content or where the most people on the platform in general have clicked. So it may simply be a statement in that particular case, and I'm really speculating, but that the users of that platform tend to like those things and so there's——

Senator JOHNSON. So you have to assume then that the vast majority of users of Instagram are liberal progressives?

Dr. WOLFRAM. That might be evidence of that.

Senator JOHNSON. Ms. Stanphill, is that what your understanding would be?

Ms. STANPHILL. Thank you, Senator.

Senator JOHNSON. If I were to do it on Google, too, it'd be interesting.

Ms. STANPHILL. I can't speak for Twitter. I can speak for Google just generally with respect to AI, which is we build products for everyone. So we've got systems in place to ensure no bias is introduced.

Senator JOHNSON. But we have—I mean, you won't deny the fact that there are plenty of instances of content being pulled off of conservative websites and having to repair the damage of that, correct? I mean, what's happening here?

Ms. STANPHILL. Thank you, Senator.

I wanted to quickly remind everyone that I am a user experience director and I work on digital well-being, which is a program to ensure that users have a balanced relationship with tech so that is a bit out of scope.

Senator JOHNSON. Mr. Harris, what's happening on here because again I think conservatives have legitimate concern that content is being pushed from the liberal progressive standpoint to the vast majority of users of these social sites?

Mr. HARRIS. Yes, I mean, I really wish I could comment, but I don't know much about where that's happening.

Senator JOHNSON. Ms. Richardson?

Ms. RICHARDSON. So there has been some research on this and it showed that when you're looking at engagement levels, there is no partisan disparity. In fact, it's equal. So I agree with Dr. Wolfram in that what you may have saw was just what was trending. Like even in the list you have Southern Poverty Law Center and they were simply trending because their Executive Director was fired. So that may just be a result of the news, not necessarily the organization.

But it's also important to understand that research has also shown that when there is any type of disparity along partisan lines, it's usually dealing with the veracity of the underlying content and that's more of a content moderation issue rather than what you're shown.

Senator JOHNSON. OK. I'd like to get that video entered in the record and we'll keep looking into this.

Senator THUNE. Without objection.

[The video referred to follows]

Senator THUNE. To the Senator from Wisconsin's point, I think if you Google yourself, you'll find most of the things that pop up right away are going to be from news organizations that tend to be hot. I mean, I have had that experience, as well, and it seems like if that actually was based upon a neutral algorithm or some other form of artificial intelligence, that since you're the user and since they know your habits and patterns, you might see something instead of from the *New York Times* pop up from *Fox News* or the *Wall Street Journal.* That to me has always been hard to explain.

Senator JOHNSON. Well, let's work together to try and get that explanation because it's a valid concern.

Senator THUNE. Senator Tester.

**STATEMENT OF HON. JON TESTER,**
**U.S. SENATOR FROM MONTANA**

Senator TESTER. Thank you, Mr. Chairman. Thanks to all the folks who have testified here today.

Ms. Stanphill, does YouTube have access to personal data on a user's Gmail account?

Ms. STANPHILL. Thank you, Senator.

I am an expert in digital well-being at Google. So, I'm sorry, I don't know that with depth and I don't want to get out of my depth. So I can take that back for folks to answer.

Senator TESTER. OK. So when it comes to Google search history, you wouldn't know that either?

Ms. STANPHILL. I'm sorry, Senator. I'm not an expert in search and I don't want to get out of my depth, but I can take it back.

Senator TESTER. OK. All right. So let me see if I can ask a question that you can answer.

Do you know if YouTube uses personal data in shaping recommendations?

Ms. STANPHILL. Thank you, Senator.

I can tell you that I know that YouTube has done a lot of work to ensure that they are improving recommendations. I do not know about privacy and data because that is not necessarily core to digital well-being. I focus on helping provide users with balanced technology usage. So in YouTube, that includes time watch profiles. It includes a reminder where if you want to set a time limit you'll get a reminder.

Senator TESTER. I got it.

Ms. STANPHILL. Ultimately, we give folks power to basically control their usage.

Senator TESTER. I understand what you're saying. I think that what I'm concerned about is that if—it doesn't matter if you're talking Google or Facebook or Twitter, whoever it is, has access to personal information, which I believe they do.

Mr. Harris, do you think they do?

Mr. HARRIS. I wish that I really knew the exact answer to the question.

Senator TESTER. Does anybody know the answer to that question?

Mr. HARRIS. The general premise is that with more personal access to information that Google has, they can provide better recommendations is usually the talking point——

Senator TESTER. So it's correct.

Mr. HARRIS.—and the business model, because they're competing for who can predict better what will keep your attention,——

Senator TESTER. My eyes on that website?

Mr. HARRIS. Yes, they would use as much information as they can and usually the way that they get around this is by giving you an option to opt out but, of course, the default is usually to opt in and that's what I think is leading to what you're talking about.

Senator TESTER. Yes, so I am 62 years old, getting older every minute the longer this conversation goes on, but I will tell you that it never ceases to amaze me that my grandkids, the oldest one is about 15 or 16, goes down to about eight, when we're on the farm is absolutely glued to this, absolutely glued to it, to the point where

if I want to get any work out of him, I have to threaten him, OK, because they're riveted.

So, Ms. Stanphill, do you guys, when you're in your leadership meetings, do you actually talk about addictive nature of this because it's as addictive as a cigarette or more and do you talk about the addictive nature? Do you talk about what you can do to stop it?

I will tell you that I'm probably going to be dead and gone and I'll probably be thankful for it when all this shit comes to fruition because I think that this scares me to death.

Senator Johnson can talk about the conservative websites. You guys could literally sit down at your board meeting, I believe, and determine who's going to be the next president of the United States. I personally believe you have that capacity. Now I could be wrong and I hope I'm wrong.

And so do any of the other folks that are here—I'll go with Ms. Richardson. Do you see it the same way or am I overreacting to a situation that I don't know enough about?

Ms. RICHARDSON. No, I think your concerns are real in that the business model that most of these companies are using and most of the optimization systems are built to keep us engaged keep us engaged with provocative material that can skew in the direction that you're concerned about.

Senator TESTER. And I don't know your history, but do you think that the board of directors for any of these companies actually sit down and talk about impacts that I'm concerned about or are they talking about how they continue to use what they've been doing to maximize their profit margin?

Ms. RICHARDSON. I don't think they're talking about the risk you're concerned about and I don't even think that's happening in the product development level and that's in part because a lot of teams are siloed. So I doubt these conversations are happening in a holistic way to sort of address your concern, which is——

Senator TESTER. Well, listen, I don't want to get in a fist fight on this panel.

Ms. Stanphill, the conversations you have, since you couldn't answer the previous ones, indicate that she's right, the conversations are siloed, is that correct?

Ms. STANPHILL. No, that's not correct, sir.

Senator TESTER. So why can't you answer my questions?

Ms. STANPHILL. I can answer the question with respect to how we think about digital well-being at Google. It's across the company. So it's actually a goal that we work on across the company. So I have the novel duty of connecting those dots, but we are doing that and we have incentives to make sure that we make progress.

Senator TESTER. OK. Well, I just want to thank you all for being here and hopefully you all leave friends because I know that there are certain Senators, including myself, who have tried to pit you against one another. That's not intentional.

I think that this is really serious. I have exactly the opposite opinion of Senator Johnson has in that I think there's a lot of driving to the conservative side. So it shows you that when humans get involved in this, we're going to screw it up, but by the same token,

there needs to be those circuit breakers that Senator Schatz talked about.

Thank you very, very much.

Senator THUNE. Thank you to the old geezer from Montana.

[Laughter.]

Senator THUNE. Senator Rosen.

**STATEMENT OF HON. JACKY ROSEN,**
**U.S. SENATOR FROM NEVADA**

Senator ROSEN. Thank you, Mr. Chairman. Thank all of you for being here today.

I have so many questions as a former software developer and systems analyst and so I see this really as I have three issues and one question.

So Issue 1 really is going to be there's a combination happening of machine language, artificial intelligence, and quantum computing all coming together that exponentially increases the capacity of predictive analytics. It grows on itself. This is what it's meant to do.

Issue 2, the monetization, the data brokering of these analytics, and the bias in all areas in regards to the monetization of this data, and then as you spoke earlier, where does the ultimate liability lie? With the scientists that craft the algorithm, the computer that potentiates the data and the algorithm, or the company or the persons who monetize the end use of the data for whatever means, right?

So three big issues, many more but on its face. My question today is on transparency. So, in many sectors we require transparency, we're used to it every day. Think about this for potential harms.

So every day, you go to the grocery store, the market, the convenience store. In the food industry, we have required nutrition labeling on every single item that clearly discloses our nutrition content. We even have it on menus now, calorie count. Oh, my, maybe I won't have that alfredo, right? You'll go for the salad.

And so we've accepted this. All of our companies have done this. It's the state of—there isn't any food that doesn't have a label. Maybe there's some food but basically we have it.

So to empower consumers, how do you think we could address some of this transparency that maybe at the end of the day we're all talking about in regards to these algorithms of data, what happens to it, how we deal with it? It's overwhelming.

Dr. WOLFRAM. I think with respect to things like nutrition labels, we have the advantage that we're using 150-year-old science to say what the chemistry of what is contained in the food is.

Things like computation and AI are a bit of a different kind of science and they have this feature that this phenomenon of computational reducibility happens and it's not possible to just give a quick summary of what the effect of this computation is going to be.

Senator ROSEN. But we know, I know having written algorithms for myself, I have kind of an expected outcome. I have a goal in there. You talk about no goal. There is a goal. Whether you meet it or not, whether you exceed it or not, whether you fail or not,

there is a goal when you write an algorithm to give somebody who's asking you for this data.

Dr. WOLFRAM. The confusing thing is that the practice of software development has changed and that it's changed in machine learning and AI.

Senator ROSEN. They can create their own goals. Machine learning——

Dr. WOLFRAM. It's not quite its own goals. It's, rather, that when you write an algorithm, you know, I expect, you know, when I started using computers a ridiculously long time ago, also, you know, you would write a small program and you would know what every line of code was supposed to do.

Senator ROSEN. With quantum computing you don't, but you still should have some ability to control the outcome.

Dr. WOLFRAM. Well, I think my feeling is that rather than saying—yes, you could put constraints on the outcome. The question is how do you describe those constraints and you have to essentially have something like a program to describe those constraints.

Let's say you want to say we want to have balanced treatment. We want to have——

Senator ROSEN. Well, let's take it out of technology and just talk about transparency in a way we can all understand. Can we put it in English terms that we're going to make your data well-being, how you use it, do you sleep, don't you sleep, how many hours a day, think about your Fitbit, who's it going to? We can bring it down to those English language parameters that people understand.

Dr. WOLFRAM. Well, I think some parts of it you could. I think the part that you cannot is when you say we're going to make this give unbiased treatment of, you know, let's say, political direction to something.

Senator ROSEN. I'm not even talking unbiased in political direction. There's going to be bias in age, in sex, in race and ethnicity. There's inherent bias in everything. So that given, you can still have other conversations.

Dr. WOLFRAM. My feeling is that rather than labeling—rather than saying we'll have a nutrition label like thing that says what this algorithm is doing, I think the better strategy is to say let's give some third party the ability to be the brand that finally decides what you see, just like with different newspapers. You can decide to see your news through the *Wall Street Journal* or through the *New York Times* or whatever.

Senator ROSEN. Who's ultimately liable if people get hurt——

Dr. WOLFRAM. Well,——

Senator ROSEN.—by the monetization of this data or the data brokering of some of it?

Dr. WOLFRAM.—that's a good question. I mean, I think that it will help to break apart the underlying platform. Something like Facebook, for example, you kind of have to use it. There's a network effect and it's not the case that, you know, you can't say let's break Facebook into a thousand different Facebooks and you could pick which one you want to use. That's not really an option.

But what you can do is to say when there's a newsfeed that's being delivered, is everybody seeing a newsfeed with the same set

of values or the same brand or not, and I think the realistic thing is to say have separate providers for that final newsfeed, for example. I think that's a possible direction, there are a few other possibilities, and that's a way, and so your sort of label says this is such and such branded newsfeed and people then get a sense of is that the one I like, is that the one that's doing something reasonable? If it's not, they'll just as a market matter reject it. That's my thought.

Senator ROSEN. I think I'm way over my time. We can all have a big conversation here. I'll submit more questions for the record. Thank you.

Senator THUNE. Thank you, Senator Rosen.

And my apologies to the Senator from New Mexico, who I missed. You were up actually before the Senator from Nevada.

Senator Udall is recognized.

**STATEMENT OF HON. TOM UDALL,**
**U.S. SENATOR FROM NEW MEXICO**

Senator UDALL. Thank you, Mr. Chairman, and thank you to the panel on a very, very important topic here.

Mr. Harris, I'm particularly concerned about the radicalizing effect that algorithms can have on young children and it has been mentioned here today in several questions. I'd like to drill down a little deeper on that.

Children can inadvertently stumble on extremist material in a number of ways, by searching for terms they don't know are loaded with subtexts, by clicking on shocking content designed to catch the eye, by getting unsolicited recommendations on content designed to engage their attention and maximize their viewing time.

It's a story told over and over by parents who don't understand how their children have suddenly become engaged with the alt-right and white nationalist groups or other extremist organizations.

Can you provide more detail how young people are uniquely impacted by these persuasive technologies and the consequences if we don't address this issue promptly and effectively?

Mr. HARRIS. Thank you, Senator.

Yes, this is one of the issues that most concerns me. As I think Senator Schatz mentioned at the beginning, there's evidence that in the last month, even as recently as that, keeping in mind that these issues have been reported on for years now, there was a pattern identified by YouTube that young girls who had taken videos of themselves dancing in front of cameras were linked in usage patterns to other videos like that that went further and further into that realm and that was just identified by YouTube, you know, a super computer, as a pattern. It's a pattern of this is a kind of pathway that tends to be highly engaging.

The way that we tend to describe this, if you imagine a spectrum on YouTube, on my left side there's the calm Walter Cronkite section of YouTube, on the right-hand side there's crazy town, UFOs, conspiracy theories, Big Foot, you know, whatever, and if you take this human being and you drop them anywhere. You could drop them in the calm section or you could drop them in crazy town, but if I'm YouTube and I want you to watch more, which direction from there am I going to send you?

I'm never going to send you to the calm section. I'm always going to send you toward crazy town. So now you imagine two billion people, like an ant colony of humanity, and it's tilting the playing field toward the crazy stuff, and the specific examples of this, a year ago a teen girl who looked at a dieting video on YouTube would be recommended anorexia videos because that was the more extreme thing to show the voodoo doll that looks like a teen girl. There are all these voodoo dolls that look like that and the next thing that shows is anorexia.

If you looked at a NASA moon landing, it would show flat earth conspiracy theories, which were recommended hundreds and hundreds of millions of times before being taken down recently.

Another example, 50 percent of white nationalists in a Belling Catch study had said that it was YouTube that had red pilled them. Red pilling is the term for, you know, the opening of the mind.

The best predictor of whether you'll believe in a conspiracy theory is whether I can get you to believe in one conspiracy theory because one conspiracy sort of opens up the mind and makes you doubt and question things and, say, get really paranoid and the problem is that YouTube is doing this en masse and it's created sort of two billion personalized *Truman Shows.*

Each channel had that radicalizing direction and if you think about it from the accountability perspective, back when we had Janet Jackson on one side of the TV screen at the Super Bowl and you had 60 million Americans on the other, we had a five-second TV delay and a bunch of humans in the loop for a reason.

But what happens when you have two billion *Truman Shows,* two billion possible Janet Jacksons, and two billion people on the other end? It's a digital *Frankenstein* that's really hard to control and so that's, I think, the way that we need to see it. From there, we talk about how to regulate it.

Senator UDALL. Yes, and, Ms. Stanphill, you've heard him just describe what Google does with young people.

What responsibility does Google have if the algorithms are recommending harmful videos to a child or a young adult that they otherwise would not have viewed?

Ms. STANPHILL. Thank you, Senator.

Unfortunately, the research and information cited by Mr. Harris is not accurate. It does not reflect current policies nor the current algorithm. So what the team has done in an effort to make sure these advancements are made, they have taken such content out of the recommendations, for instance. That limits the views by more than 50 percent.

Senator UDALL. So are you saying you don't have any responsibility?

Ms. STANPHILL. Thank you, Senator.

Senator UDALL. Because clearly young people are being directed toward this kind of material. There's no doubt about it.

Ms. STANPHILL. Thank you, Senator.

YouTube is doing everything that they can to ensure child safety online and works with a number of organizations to do so and will continue to do so.

Senator UDALL. Do you agree with that, Mr. Harris?

Mr. HARRIS. I don't because I know the researchers who are unpaid and stay up till 3 in the morning trying to scrape the datasets to show what these actual results are and it's only through huge amounts of public pressure that incrementally they tackle bit by bit, issue by issue, bits and pieces of it, and if they were truly acting with responsibility, they would be doing so preemptively without the unpaid researchers staying up till 3 in the morning doing that work.

Senator UDALL. Yes, thank you, Mr. Chairman.

Senator THUNE. Thank you, Senator Udall.

Senator Sullivan.

**STATEMENT OF HON. DAN SULLIVAN,**
**U.S. SENATOR FROM ALASKA**

Senator SULLIVAN. Thank you, Mr. Chairman, and I appreciate the witnesses being here today, very important issue that we're all struggling with.

Let me ask Ms. Stanphill. I had the opportunity to engage in a couple rounds of questions with Mr. Zuckerberg from Facebook when he was here. One of the questions I asked, which I think we're all trying to struggle with, is this issue of what you, when I say you, Google or Facebook, what you are, right.

You think there's this notion that you're a tech company, but some of us think you might be the world's biggest publisher. I think about a 140 million people get their news from Facebook. When it combines Google and Facebook, I think it's about somewhere north of 80 percent of Americans get theirs news.

So what are you? Are you a publisher? Are you a tech company? Are you responsible for your content? I think that's another really important issue. Mark Zuckerberg did say he was responsible for their content but at the same time, he said that they're a tech company, not a publisher, and as you know, whether you are one or the other, it is really critical, almost the threshold issue in terms of how and to what degree you would be regulated by Federal law.

So which one are you?

Ms. STANPHILL. Thank you, Senator.

As I might remind everybody, I am a user experience director for Google and so I support our Digital Well-Being Initiative.

With that said, I know we're a tech company. That's the extent to which I know the definition that you're speaking of.

Senator SULLIVAN. So do you feel you're responsible for the content that comes from Google on your websites when people do searches?

Ms. STANPHILL. Thank you, Senator.

As I mentioned, this is a bit out of my area of expertise as the digital well-being expert. I would defer to my colleagues to answer that specific question.

Senator SULLIVAN. Well, maybe we can take those questions for the record.

Ms. STANPHILL. Of course.

Senator SULLIVAN. Anyone else have a thought on that pretty important threshold question?

Mr. HARRIS. Yes, I think——

Senator SULLIVAN. Mr. Harris?

Mr. HARRIS. Is it okay if I jump in, Senator?

Senator SULLIVAN. Yes.

Mr. HARRIS. The issue here is that Section 230 of the Communications Act——

Senator SULLIVAN. It's all about Section 230.

Mr. HARRIS. It's all about Section 230, has obviously made it so that the platforms are not responsible for any content that is on them which freed them up to do what they've created today.

The problem is if, you know, is YouTube a publisher? Well, they're not generating the content. They're not paying journalists. They're not doing that, but they are recommending things, and I think that we need a new class between, you know, the *New York Times* is responsible if they say something that defames someone else that reaches a certain hundred million or so people.

When YouTube recommends flat earth conspiracy theories hundreds of millions of times and if you consider that 70 percent of YouTube's traffic is driven by recommendations, meaning driven by what they are recommending, when the algorithm is choosing to put in front of the eyeballs of a person, if you were to backward derive a motto, it would be with great power comes no responsibility.

Senator SULLIVAN. Let me follow up on that, two things real quick because I want to make sure I don't run out of time here. It's a good line of questioning.

You know, when I asked Mr. Zuckerberg, he actually said they were responsible for their content. That was in a hearing like this. Now that actually starts to get close to being a publisher from my perspective. So I don't know what Google's answer is or others, but I think it's an important question.

Mr. Harris, you just mentioned something that I actually think is a really important question and I don't know if some of you saw Tim Cook's commencement speech at Stanford a couple weeks ago. I happened to be there and saw it. I thought it was quite interesting.

But he was talking about all the great innovations from Silicon Valley, but then he said, "Lately, it seems this industry is becoming better known for a less noble innovation, the belief that you can claim credit without accepting responsibility."

Then he talked about a lot of the challenges and then he said, "It feels a bit crazy that anyone should have to say this but if you built a chaos factory, you can't dodge responsibility for the chaos. Taking responsibility means having the courage to think things through."

So I'm going to open this up, kind of final question, and maybe we start with you, Mr. Harris. What do you think he was getting at? It was a little bit generalized, but he obviously put a lot of thought into his commencement speech at Stanford, this notion of building things, creating things and then going whoa, whoa, I'm not responsible for that. What's he getting at? I'll open that to any other witnesses. I thought it was a good speech, but I'd like your views on it.

Mr. HARRIS. Yes, and I think it's exactly what everyone's been saying on this panel, that these things have become digital Frankensteins that are terror-forming the world in their image, whether

it's the mental health of children or our politics and our political discourse, and without taking responsibility for taking over the public square.

So again it comes back to——

Senator SULLIVAN. Who do you think's responsible?

Mr. HARRIS. I think we have to have the platforms be responsible for when they take over election advertising, they're responsible for protecting elections. When they take over mental health of kids on Saturday morning, they're responsible for protecting Saturday morning.

Senator SULLIVAN. Anyone else have a view on the quotes I gave from Tim Cook's speech? Mr. Wolfram?

Dr. WOLFRAM. I think one of the questions is what do you want to have happen? That is, you know, when you say something bad is happening, it's giving the wrong recommendations. By what definition of wrong? What is the—you know, who is deciding? Who is kind of the moral auditor? If I was running one of these automated content selection companies, my company does something different, I would not want to be kind of a moral arbiter for the world, which is what effectively having to happen when there are some decisions being made about what content will be delivered, what will not be being delivered.

My feeling is the right thing to have happen is to break that apart, to have a more market-based approach, to have third parties be the ones who are responsible for sort of that final decision about what content is delivered to what users, so that the platforms can do what they do very well, which is the kind of large-scale engineering, large-scale monetization of content, but somebody else gets to be—somebody that users can choose from. The third party gets to be the one who is deciding sort of the final ranking of content shown to particular users, so users can get, you know, brand allegiance to the particular content providers that they want and not to other ones.

Senator SULLIVAN. Thank you, Mr. Chairman.

Senator THUNE. Thank you, Senator Sullivan.

Senator Markey.

**STATEMENT OF HON. EDWARD MARKEY,**
**U.S. SENATOR FROM MASSACHUSETTS**

Senator MARKEY. Thank you, Mr. Chairman, very much.

YouTube is far and away the top website for kids today. Research shows that a whopping 80 percent of six-through-12-year-olds, six-through-12-year-olds use YouTube on a daily basis, but when kids go on YouTube, far too often they encounter inappropriate and disturbing video clips that no child should ever see.

In some instances, when kids click to view cartoons and characters in their favorite games, they find themselves watching material promoting self-harm and even suicide. In other cases, kids have opened videos featuring beloved Disney princesses and all of a sudden see a sexually explicit scene.

Videos like this shouldn't be accessible to children at all, let alone systematically served to children.

Mr. Harris, can you explain how, once a child consumes one inappropriate YouTube video, the website's algorithms begin to prompt the child to watch more harmful content of that sort?

Mr. HARRIS. Yes, thank you, Senator.

So if you watch a video about a topic, let's say it's that cartoon character *The Hulk* or something like that, YouTube picks up some pattern that maybe *Hulk* videos are interesting to you.

The problem is there's a dark market of people who you're referencing in that long article that's very famous who actually generate content that's based on the most viewed videos. They'll look at the thumbnails and say, oh, there's a *Hulk* in that video, there's a *Spiderman* in that video, and then they have machines actually manufacture free-generated content and then upload it to YouTube machines and tag it in such a way that it gets recommended near those content items and YouTube is trying to maximize traffic for each of these publishers.

So when these machines upload the content, it tries to dose them with some views and saying, well, maybe this video's really good, and it ends up gathering millions and millions of views because kids, quote unquote, like them, and I think the key thing going on here is that, as I said in the opening statement, this is about an asymmetry of power being masked in an equal relationship because technology companies claim we're giving you what you want as opposed to——

Senator MARKEY. So the six-to-12-year-olds, they just keep getting fed the next video, the next video, the next video,——

Mr. HARRIS. Correct.

Senator MARKEY.—and there's no way that that can be a good thing for our country over a long period of time.

Mr. HARRIS. Especially when you realize the asymmetry that YouTube's pointing a super computer at that child's brain in a calculated——

Senator MARKEY. That is a six-year-old, an eight-year-old, ten-year-old. It's wrong. So clearly the way the websites are designed impose serious harm to children and that's why in the coming weeks, I will be introducing the KIDS Internet Design and Safety Act, the KIDS Act.

Specifically, my bill will combat amplification of inappropriate and harmful content on the internet, online design features, like auto-play, that coerce children and create bad habits, and commercialization and marketing that manipulates kids and push them into consumer culture.

So to each of today's witnesses, will you commit to working with me to enact strong rules that tackle the design features and underlying issues that make the Internet unsafe for kids? Mr. Harris?

Mr. HARRIS. Yes.

Senator MARKEY. Ms. Stanphill?

Ms. STANPHILL. Yes.

Dr. WOLFRAM. It's a terrific goal but it's not particularly my expertise.

Senator MARKEY. OK.

Ms. RICHARDSON. Yes.

Senator MARKEY. OK. Thank you.

Ms. Stanphill, recent reporting suggests that YouTube is considering significant changes to its platform, including ending auto-play for children's videos, so that when one video ends, another doesn't immediately begin, hooking the child on to long viewing sessions. I've called for an end to auto-play for kids.

Can you confirm to this Committee that YouTube is getting rid of that feature?

Ms. STANPHILL. Thank you, Senator.

I cannot confirm that as a representative from Digital Well-Being. Thank you. I can get back to you, though.

Senator MARKEY. I think it's important and I think it's very important that that happen voluntarily or through Federal legislation to make sure that the Internet is a healthier place for kids.

Senators Blunt and Schatz and myself, Senator Sasse, Senator Collins, Senator Bennett, are working on a bipartisan Children and Media Research Advancement Act that will commission a 5-year $95 million research initiative at the National Institutes of Health to investigate the impact of tech on kids. It will produce research to shed light on the cognitive, physical, and socio-emotional impacts of technology on kids.

I look forward on that legislation to working with everyone at this table, as well, so that we can design legislation and ultimately a program.

I know that Google has endorsed the CAMERA Act. Ms. Stanphill, can you talk to this issue?

Ms. STANPHILL. Yes, thank you, Senator.

I can speak to the fact that we have endorsed the CAMERA Act and look forward to working with you on further regulation.

Senator MARKEY. OK. Same thing for you, Mr. Harris.

Mr. HARRIS. We've also endorsed it at the Center for Humane Technology.

Senator MARKEY. Thank you. So I just think we're late as a nation to this subject, but I don't think that we have an option. We have to make sure that there are enforceable protections for the children of our country.

Thank you, Mr. Chairman.

Senator THUNE. Thank you, Senator Markey.

Senator Young.

**STATEMENT OF HON. TODD YOUNG,**
**U.S. SENATOR FROM INDIANA**

Senator YOUNG. I thank our panel for being here.

I thought I'd ask a question about concerns that many have and I expect concerns will grow about AI becoming a black box where it's unclear exactly how certain platforms make decisions.

In recent years, deep learning has proved very powerful at solving problems and has been widely deployed for tasks, like image captioning, voice recognition, and language translation. As the technology advances, there is great hope for AI to diagnose deadly diseases, calculate multimillion dollar trading decisions, and implement successful autonomous innovations for transportation and other sectors.

Nonetheless, the intellectual power of AI has received public scrutiny and has become unsettling for some futurists. Eventually,

society might cross a threshold in which using AI requires a leap of faith.

In other words, AI might become, as they say, a black box where it might be impossible to tell how in AI that has internalized massive amounts of data is making its decisions through its neural network and, by extension, it might be impossible to tell how those decisions impact the psyche, the perceptions, the human understanding, and perhaps even the behavior of an individual.

In early April, the European Union released final ethical guidelines calling for what it calls trustworthy AI. The guidelines aren't meant to be or intended to interfere with policies or regulations but instead offer a loose framework for stakeholders to implement their recommendations.

One of the key guidelines relates to transparency in the ability for AI systems to explain their capabilities, limitations, and decisionmaking. However, with the improvement of AI requires, for example, more complexity, imposing transparency requirements will be equivalent to a prohibition on innovation.

So I will open this question to the entire panel but my hope is that Dr. Wolfram, I'm sorry, sir, you can begin.

Can you tell this Committee the best ways for Congress to collaborate with the tech industry to ensure AI system accountability without hindering innovation and specifically should Congress implement industry requirements or guidelines for best practices?

Dr. WOLFRAM. It's a complicated issue. I think that it varies from industry to industry. I think in the case of what we're talking about here, Internet automated content selection, I think that the right thing to do is to insert a kind of level of human control into what is being delivered but not in the sense of taking apart the details of an AI algorithm but making the structure of the industry be such that there is some human choice injected into what's being delivered to people.

I think the biggest story is we need to understand how we're going to make laws that can be specified in computational form and applied to AIs. We're used to writing laws in English basically and we're used to being able to say, you know, write down some words and then have people discuss whether they're following those words or not.

When it comes to computational systems that won't work. Things are happening too quickly. They're happening too often. You need something where you're specifying computationally this is what you want to have happen and then the system can perfectly well be set up to automatically follow those computational rules or computational laws.

The challenge is to create those computational rules and that's something we're just not yet experienced with. It's something that we're starting to see computational contracts as a practical thing in the world of block chain, and so on, but we don't yet know how you'd specify some of the things that we want to specify as rules for how systems work. We don't yet know how to do that computationally.

Senator YOUNG. Are you familiar with the EU's approach to develop ethical guidelines for trustworthy AI?

Dr. WOLFRAM. I'm not familiar with those guidelines.

Senator YOUNG. OK. Are any of the other panelists?

[Negative responses.]

Senator YOUNG. OK. Well, then perhaps that's a model we could look at. Perhaps that would be ill-advised. So for stakeholders that may be watching these proceedings or listening to them, they can tell me. Do others have thoughts?

Ms. RICHARDSON. So in my written comments, I outlined a number of transparency mechanisms that could help address some of your concerns and some of the recommendations, one specifically, which was the last one, is we suggested that companies create an algorithmic impact assessment and that framework, which we initially wrote for government use, can actually be applied in the private sector and we built the framework from learning from different assessments.

So in the U.S., we used environmental impact assessments, which allows for robust conversation about developmental projects and their impact on the environment but also in the EU, which is one of the reference points that we used, they have a data protection impact assessment and that's something that's done both in government and in the private sector, but the difference here and why I think it's important for Congress to take action is what we're suggesting is something that's actually public, so we can have a discourse about whether this is a technological tool that has a net benefit for society or it's something that's too risky that shouldn't be available.

Senator YOUNG. I'll be attentive to your proposal. Do you mind if we work with you, a dialogue, if we have any questions about it?

Ms. RICHARDSON. Yes, very much.

Senator YOUNG. All right. Thank you. Others have any thoughts? It's OK if you don't.

[No response.]

Senator YOUNG. OK. It sounds like we have a lot of work to do, industry working with other stakeholders, to make sure that we don't act impulsively, but we also don't neglect this area of public policy.

Thank you.

Senator THUNE. Thank you, Senator Young.

Senator Cruz.

## STATEMENT OF HON. TED CRUZ, U.S. SENATOR FROM TEXAS

Senator CRUZ. Ms. Stanphill, a lot of Americans have concerns that big tech media companies and Google in particular are engaged in political censorship and bias. As you know, Google enjoys a special immunity from liability under Section 230 of the Communications Decency Act. The predicate for that immunity was that Google and other big tech media companies would be neutral public fora.

Does Google consider itself a neutral public forum?

Ms. STANPHILL. Thank you, Senator.

Yes, it does.

Senator CRUZ. OK. Are you familiar with the report that was released yesterday from Veritas that included a whistleblower from within Google, that included videos from a senior executive at

Google, that included documents that are purportedly internal PowerPoint documents from Google?

Ms. STANPHILL. Yes, I heard about that report in industry news.

Senator CRUZ. Have you seen the report?

Ms. STANPHILL. No, I have not.

Senator CRUZ. So you didn't review the report to prepare for this hearing?

Ms. STANPHILL. It has been a busy day and I have a day job which is Digital Well-Being at Google. So I'm trying to make sure I keep the——

Senator CRUZ. Well, I'm sorry that this hearing is infringing on your day job.

Ms. STANPHILL. It's a great opportunity. Thank you.

Senator CRUZ. Well, one of the things in that report, and I would recommend people interested in political bias at Google watch the entire report and judge for yourself, there's a video from a woman, Jen Gennai—it's a secret video that was recorded. Jen Gennai, as I understand it, is the Head of "Responsible Innovation for Google." Are you familiar with Ms. Gennai?

Ms. STANPHILL. I work in User Experience and I believe that AI Group is somebody we worked with on the AI Principles, but it's a big company, and I don't work directly with them.

Senator CRUZ. Do you know her or no?

Ms. STANPHILL. I do not know Jen.

Senator CRUZ. OK. As I understand it, she is shown in the video saying, and this is a quote, "Elizabeth Warren is saying that we should break up Google and like I love her but she's very misguided, like that will not make it better. It will make it worse because all these all these smaller companies who don't have the same resources that we do will be charged with preventing the next Trump situation. It's like a small company cannot do that."

Do you think it's Google's job to "prevent the next Trump situation?"

Ms. STANPHILL. Thank you, Senator.

I don't agree with that. No, sir.

Senator CRUZ. So a different individual, a whistleblower identified simply as an insider at Google with knowledge of the algorithm is quoted on the same report as saying Google "is bent on never letting somebody like Donald Trump come to power again."

Do you think it's Google's job to make sure "somebody like Donald Trump" never comes to power again?

Ms. STANPHILL. No, sir, I don't think that is Google's job, and we build for everyone, including every single religious belief, every single demographic, every single region, and certainly every political affiliation.

Senator CRUZ. Well, I have to say that certainly does not appear to be the case.

Of the senior executives at Google, do you know of a single one who voted for Donald Trump?

Ms. STANPHILL. Thank you, Senator.

I'm a user experience director, and I work on Google Digital Well-Being, and I can tell you we have diverse views, but I can't——

Senator CRUZ. Do you know of anyone who voted for Trump of the senior executives?

Ms. STANPHILL. I definitely know of people who voted for Trump.

Senator CRUZ. Of the senior executives at Google?

Ms. STANPHILL. I don't talk politics with my workmates.

Senator CRUZ. Is that a no?

Ms. STANPHILL. Sorry. Is that a no to what?

Senator CRUZ. Do you know of any senior executives, even a single senior executive at the company who voted for Donald Trump?

Ms. STANPHILL. As the digital well-being expert, I don't think this is in my purview to comment on people——

Senator CRUZ. Do you know of—that's all right. You don't have to know.

Ms. STANPHILL. I definitely don't know.

Senator CRUZ. I can tell you what the public records show. The public records show that in 2016 Google employees gave to Hillary Clinton Campaign $1.315 million. That's a lot of money. Care to venture how much they gave to the Trump Campaign?

Ms. STANPHILL. I would have no idea, sir.

Senator CRUZ. Well, the nice thing is it's a round number, zero dollars and zero cents, not a penny, according to the public reports.

Let's talk about one of the PowerPoints that was leaked. The Veritas report has Google internally saying, "I propose we make machine learning intentionally human-centered and intervene for fairness."

Is this document accurate?

Ms. STANPHILL. Thank you, sir.

I don't know about this document, so I don't know.

Senator CRUZ. OK. I'm going to ask you to respond to the Committee in writing afterwards as to whether this PowerPoint and the other documents that are included in the Veritas report, whether those documents are accurate, and I recognize that your lawyers may want the right explanation. You're welcome to write all the explanation that you want, but I also want a simple clear answer. Is this an accurate document that was generated by Google?

Do you agree with the sentiment expressed in this document?

Ms. STANPHILL. No, sir, I do not.

Senator CRUZ. Let me read you another also in this report. It indicates that Google, according to this whistleblower, "deliberately makes recommendations if someone is searching for conservative commentators, deliberately shifts the recommendations so instead of recommending other conservative commentators, it recommends organizations, like CNN or MSNBC or left-leaning political outlets." Is that occurring?

Ms. STANPHILL. Thank you, sir.

I can't comment on search algorithms or recommendations, given my purview as the digital well-being lead. I can take that back to my team, though.

Senator CRUZ. So is it part of digital well-being for search recommendations to reflect where the user wants to go rather than deliberately shifting where they want to go?

Ms. STANPHILL. Thank you, sir.

As the user experience professional, we focus on delivering on user goals. So we try to get out of the way and get them on the task at hand.

Senator CRUZ. So a final question. One of these documents that was leaked explains what Google is doing and it has a series of steps, "Training data are collected and classified, algorithms are programmed, media are filtered, ranked, aggregated, and guaranteed and that ends with people (like us) are programmed."

Does Google view its job as programming people with search results?

Ms. STANPHILL. Thank you, Senator.

I can't speak for the whole entire company, but I can tell you that we make sure that we put our users first in our design.

Senator CRUZ. Well, I think these questions raise very serious—these documents raise very serious questions about political bias at the company.

Senator THUNE. Thank you, Senator Cruz.

Senator Schatz, anything to wrap up with?

Senator SCHATZ. Just a quick statement and then a question.

I don't want the working of the refs to be left unresponded to and I won't go into great detail, except to say that there are Members of Congress who use the working of the refs to terrify Google and Facebook and Twitter executives so that they don't take action in taking down extreme content, false content, polarizing content, contra their own rules of engagement, and so I don't want the fact that the Democratic side of the aisle is trying to engage in good faith on this public policy matter and not work the refs allow the message to be sent to the leadership of these companies that they have to respond to this bad faith accusation every time we have any conversation about what to do in tech policy.

My final question for you, and this will be the last time I leap to your defense, Ms. Stanphill, did you say privacy and data is not core to digital well-being?

Ms. STANPHILL. Thank you, sir.

I might have misstated how that's being phrased. So what I meant——

Senator SCHATZ. What do you mean to say?

Ms. STANPHILL. Oh, I mean to say that there is a team that focuses day-in/day-out on privacy, security, control as it relates to user data. That's outside of my area.

Senator SCHATZ. But so in your talking sort of bureaucratically and I don't mean that as a pejorative, you're talking about the way the company is organized.

I'm saying aren't privacy and data core to digital well-being?

Ms. STANPHILL. I see. Sorry I didn't understand that point, Senator. In retrospect, what I believe is that it is inherent in our digital well-being principles that we focus on the user and that requires that we focus on privacy, security, control of their data.

Senator SCHATZ. Thank you.

Senator THUNE. Thank you, Senator Schatz.

And to be fair, I think both sides work the refs, but let me just ask a follow-on question. I appreciate Senator Blackburn's line of questioning from earlier which may highlight some of the limits on transparency.

As we have sort of started, I think, in our opening statements today by trying to look at ways that in this new world we can provide a level of transparency, you said it's going to be very difficult in terms of explainability of AI, but just understanding a little bit better how to provide users the information they need to make educated decisions about how they interact with the platform services.

So the question is, might it make sense to let users effectively flip a switch to see the difference between a filtered algorithm-based presentation and an unfiltered presentation?

Dr. WOLFRAM. I mean, there are already, for example, search services that aggregate user searches and feed them en masse to search engines, like Bing, so that you're effectively seeing the results of a generic search, independent of specific information about you works okay.

There are things for which it doesn't work well. I think that this idea of, you know, you flip a switch, I think that is probably not going to have great results because I think there will be unfortunately great motivation to have the case where the switch is flipped to not give user information give bad results. I'm not sure how you would motivate giving good results in that case.

I think that it's also—it's sort of when you think about that switch, you can think about a whole array of other kinds of switches and I think pretty soon it gets pretty confusing for users to decide, you know, which switches do they flip for what. Do they give location information but not this information? Do they give that information, not that information?

I mean, my own feeling is the most promising direction is to let some third party be inserted who will develop a brand. There might be 20 of these third parties. It might be like newspapers where people can pick, you know, do they want news from this place, that place, another place? To insert third parties and have more of a market situation where you are relying on the trust that you have in that third party to determine what you're seeing rather than saying the user will have precise detailed control.

I mean, as much as I would like to see more users be more engaged in kind of computational thinking and understanding what's happening to their computational systems, I don't think this is a case where that's going to work in practice.

Senator THUNE. Anybody else? Ms. Richardson.

Ms. RICHARDSON. So I think the issue with the flip the switch hypothetical, users need to be aware of the tradeoffs and currently so many users are used to the conveniences of existing platforms. So there's currently a privacy-preserving platform called DuckDuckGo which doesn't take your information and it gives you search results.

But if you're used to seeing the most immediate result at the top, DuckDuckGo, even though it's privacy-preserving, may not be the choice that all users would use but they're not hyper-aware of what are the tradeoffs of them giving that information to the provider.

So I think it's—while I understand the reason you're giving that metaphor, it's important for users to understand both the practices of a platform and also to understand the tradeoffs where if they want a more privacy-preserving service, what are they losing or gaining from that.

Senator THUNE. Mr. Harris.

Mr. HARRIS. Yes, the issue is also that users, I think it's already been mentioned, will quote unquote "prefer" because it saves them time and energy the summarized feed that's algorithmically filtering it down for them.

Even Jack Dorsey at Twitter has said that when you show people the reverse chronological feed versus the algorithmic one, people, they just save some time and it's more relevant to do the algorithmic one.

So even if there's a switch, most people will, quote unquote, prefer that one, and I think we have to be aware of the tradeoffs and we have to have a notion of what fair really means there.

What I'm most concerned about is the fact that this is still fairness with respect to the increasingly fragmented truth that debases the information environment that democracy depends on of shared truth or shared narrative.

Senator THUNE. OK.

Dr. WOLFRAM. I'd like to comment on that issue. I mean, I think the challenge is when you want to sort of have a single shared truth, the question is who gets to decide what that truth is, and I think that's—you know, the question is, is that decided within a single company, you know, implemented using AI algorithms? Is that decided in some more, you know, market kind of way by a collection of companies?

I think it makes more sense in kind of the American way of doing things to imagine that it's decided by a whole selection of companies rather than being something that is burnt into a platform that, for example, has sort of become universal through network effects and so on.

Senator THUNE. All right. Well, thank you all very much. This is a very complicated subject but one I think that your testimony and responses have helped shed some light on and certainly will shape our thinking in terms of how we proceed, but there's definitely a lot of food for thought there. So thank you very much for your time and for your input today.

We'll leave the hearing record open for a couple of weeks and we'll ask Senators if they have questions for the record to submit those, and we would ask all of you, if you can, to get those responses back as quickly as possible so that we can include them in the final hearing record.

I think with that, we are adjourned.

[Whereupon, at 12:05 p.m., the hearing was adjourned.]

# A P P E N D I X

ELECTRONIC PRIVACY INFORMATION CENTER
*Washington, DC, June 24, 2019*

Senator JOHN THUNE, Chairman,
Senator BRIAN SCHATZ, Ranking Member,
Committee on Commerce, Science, and Transportation,
Subcommittee on Communications, Technology, Innovation, and the Internet,
Washington, DC.

Dear Chairman Thune and Ranking Member Schatz:

We write to you regarding the hearing this week on "Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms."[1] We appreciate your interest in this important issue.

EPIC has been at the forefront of efforts to promote Algorithmic Transparency.[2] We also helped draft Universal Guidelines for AI,[3] which received support from 60 associations (including the AAAS) and 250 experts from more than 40 countries.[4] We also helped draft the OECD AI Principles, which were endorsed by 42 countries, including the United States.[5]

We would be pleased to provide more information to the Committee about this work.

Sincerely,

/s/*Marc Rotenberg*
Marc Rotenberg
EPIC President

/s/*Caitriona Fitzgerald*
Caitriona Fitzgerald
EPIC Policy Director

---

RESPONSE TO WRITTEN QUESTION SUBMITTED BY HON. JOHN THUNE TO TRISTAN HARRIS

*Question.* Innovation cannot be focused on building new capabilities alone. It has to be paired with forward thinking design that promotes safety and user trust. Do companies have a social responsibility to design technology that is optimized for consumers' digital wellbeing?

Answer. Yes, companies absolutely have a social responsibility to design technology that is optimized for consumers' digital wellbeing. Today, they are acting as if they have none—they assume their impact is good. But now that the world has woken up to the harms intrinsic to their business model, which is to extract attention and data through mass behavior modification, that must change. More than do no harm, technology platforms should have a responsibility to get clear about the goods they aim to achieve, while avoiding the many harms and externalities to mental health, civic health and the social fabric.

There is a precedent for this kind of responsibility. The asymmetry between technology's power over those it impacts is comparable to that of a lawyer, doctor or psychotherapist. These occupations are governed under fiduciary law, due to the

---

[1] *Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms: Hearing Before the S. Comm. on Commerce, Science, & Transportation, Subcomm. on Communications, Technology, Innovation, and the Internet,* 116th Cong. (2019), *https://www.commerce.senate.gov/public/index.cfm/2019/6/optimizing-for-engagement-understanding-the-use-of-persuasive-technology-on-internet-platforms* (June 25, 2019).

[2] EPIC, *Algorithmic Transparency, https://epic.org/algorithmic-transparency/*.

[3] The Public Voice, *Universal Guidelines for Artificial Intelligence, https://thepublicvoice.org/AI-universal-guidelines.*

[4] A full list of endorsers is available at The Public Voice, *Universal Guidelines for Artificial Intelligence: Endorsement, https://thepublicvoice.org/AI-universal-guidelines/endorsement.*

[5] *OECD Privacy Guidelines, https://www.oecd.org/internet/ieconomy/privacy-guidelines.htm.*

level of compromising and vulnerable information they hold over their clients. Because the level of compromising information technology platforms hold over their users exceeds that asymmetry, they should also be governed under fiduciary law.

This would make their advertising and behavior modification business model illegal, much like it would be illegal for a doctor, psychotherapist or lawyer to operate under a business model of extracting as much value from their clients by manipulating them into outcomes only possible because of their knowledge of their clients' vulnerabilities. This means it is critical to ensure that this asymmetric power is governed by a relationship of responsibility, not of extraction—very similar to the responsible practices and standards that the FCC created to protect children and children's television programming.

As technology eats the public square, companies have a social responsibility to protect both consumers' digital wellbeing and the social fabric in which they operate.

---

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO TRISTAN HARRIS

*A.I. Accountability and Civil Rights.* One tech company, Facebook, announced that it is conducting an audit to identify and address discrimination. It has also formed Social Science One, which provides external researchers with data to study the platform's effects of social media on democracy and elections.

*Question 1.* What specific datasets and information would you need to scrutinize Facebook and Google's systems on civil rights and disinformation?

Answer. While an incredibly important question, I'm not an expert on Facebook's and Google's existing datasets on civil rights and discrimination.

*Loot Boxes.* One of the most prolific manipulative practices in the digital economy is "loot boxes." Loot boxes are, in effect, gambling—selling gamers randomly-selected virtual prizes. The games do everything they can to coax people to taking chances on loot boxes. There is increasing scientific evidence that loot boxes share the same addictive qualities as gambling.

*Question 2.* Do you agree with me that loot boxes in video games share the same addictive qualities as gambling, particularly when targeting children?

Answer. Yes, I agree that loot boxes in video games share the same addictive qualities as gambling in that they operate on intermittent variable reward schedules, which mirror the mechanics of casinos in Las Vegas.[1]

*Question 3.* Would you support legislation like the Protecting Children from Abusive Games Act, which would prohibit the sale of loot boxes in games catering to children?

Answer. Yes.

*Data Privacy and Manipulative Technologies.* Google and Facebook have an intimate understanding of the private lives of their users. They know about our family relationships, our financial affairs, and our health. This rich profile of our lives is intensively mined to exploit our attention and target us with ever-more manipulative advertising. However, while persuasive technologies take advantage of information about users, their users know little about them.

*Question 4.* Would Google and Facebook, if they wanted to, be able to specifically single out and target people when they are emotionally vulnerable or in desperate situations based on the data they collect?

Answer. Yes, Facebook's own documents demonstrate that in one Facebook marketer's account in Australia, they knew when teenagers were feeling low self-esteem and could predict it based on usage patterns. Additionally, they were able to deduce whether people are feeling lonely or isolated based on the kinds of usage that they are demonstrating.[2] Google, as another example, knows when people are typing in search queries like "how to commit suicide."[3]

---

[1] Bailey, J. M. (2018, April 24). A Video Game 'Loot Box' Offers Coveted Rewards, but Is It Gambling? *The New York Times.* Retrieved August 2, 2019, from *https://www.nytimes.com/2018/04/24/business/loot-boxes-video-games.html*

[2] Davidson, D. (2017, May 1). Facebook targets "insecure" young people. *The Australian.* Retrieved August 2, 2019, from *http://www.theaustralian.com.au/business/media/digital/facebook-targets-insecure-young-people-to-sell-ads/news-story/a89949ad016eee7d7a61c3c30c909fa6.* Facebook responded (*https://newsroom.fb.com/news/h/comments-on-research-and-ad-targeting/*) denying that they were targeting the teens for ads, but not denying that they had the ability to do so.

[3] Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural Language Processing of Social Media as Screening for Suicide Risk. *Biomedical Informatics Insights, 10,* 117822261879286. doi:10.1177/1178222618792860; Ma-Kellams, C., Or, F., Baek, J. H., & Kawachi, I. (2015). Rethinking Suicide Surveillance. *Clinical Psychological Science,4*(3), 480–

Technology companies are *already* aware of desperate situations—based not just on the data that they collect, but also based on predictions they can make by their consumers' usage patterns.

*Question 5.* Currently, would it be against the law to do so—for example, were Facebook to target teenagers that it predicts feel like "a failure" with ads?

Answer. I'm not a legal expert, but this is an area that certainly seems worthy of deep legal scrutiny.

*Question 6.* How can we ensure data privacy laws prevent the use of personal data to manipulate people based on their emotional state and vulnerabilities?

Answer. Even with good data privacy laws, we need to regulate the channels by which this data can be used to manipulate users' psychological vulnerabilities. We also must recognize the fact that the genie is out of the bottle—much of the data that is sensitive is already available on the dark web for purchase by any malicious actor. Therefore, more important than just regulating the collection of sensitive information, we need to regulate the sensitive channels that permit targeting these vulnerable individuals—specifically, micro-targeting features on advertising platforms that include Facebook Custom Audiences and Facebook Lookalike models. Without assurances that would ensure only ethical use, these features should be banned altogether.

I'm not a legal expert but in general, I am supportive of data privacy laws that prevent the use of personal data to manipulate people based on their emotional state and vulnerabilities. It would be beneficial for lawmakers to consider extending privacy laws to capture the differential vulnerability of users based on their emotional state and vulnerable qualities of their situation.

Much like doctors, psychotherapists and lawyers are in a relationship with clients who are vulnerable by their sharing of highly sensitive information that could impact their health, financial or psychological outcomes, there are special governing laws of fiduciary or duty of care that protect those relationships. We believe it is worth extending the application of fiduciary to technology platforms.

*Recommendations from Ms. Richardson.* Ms. Richardson provided a set of recommendations in her remark for Congress to act, including:

1.) Require Technology Companies to Waive Trade Secrecy and Other Legal Claims That Hinder Oversight and Accountability Mechanisms

2.) Require Public Disclosure of Technologies That Are Involved in Any Decisions About Consumers by Name and Vendor

3.) Empower Consumer Protection Agencies to Apply "Truth in Advertising Laws" to Algorithmic Technology Providers

4.) Revitalize the Congressional Office of Technology Assessment to Perform Pre-Market Review and Post-Market Monitoring of Technologies

5.) Enhanced Whistleblower Protections for Technology Company Employees That Identify Unethical or Unlawful Uses of AI or Algorithms

6.) Require Any Transparency or Accountability Mechanism To Include A Detailed Account and Reporting of The "Full Stack Supply Chain"

7.) Require Companies to Perform and Publish Algorithmic Impact Assessments Prior to Public Use of Products and Services

During the hearing, I requested for you to respond in writing if possible.

*Question 7.* Please provide feedback to Ms. Richardson's suggestions for Congressional Action.

Answer. I'm not an expert on all of these suggestions, but fully support recommendations 3, 4, 5 and 7, especially for companies who serve such large bases of users they have effectively become the infrastructure our society depends on.

*Question 8.* What other steps or actions should Congress consider in regulating the use or consumer protection regarding persuasive technologies or artificial intelligence?

Answer. The central problem is the need to decouple the relationship between profit and the frequency and duration of use of products. The current model incentivizes companies to encourage frequent and long duration of use. This model results in many of the harms we're seeing today.

---

484. doi:10.1177/2167702615593475. Mark Zuckerberg has also written about how Facebook uses AI tools to detect when users are expressing suicidal thoughts. *https://www.facebook.com/zuck/posts/10104242660091961*

It's most important to go after the incentives that create addictive, infinite-use technologies, rather than regulating the use or consumer protection regarding persuasive technologies or artificial intelligence.

- What if technologies that maximize engagement through time-on-screen were regulated like a utility?
- What if employee performance, incentive packages, and bonuses were decoupled from 'engagement' (time-on-site, daily active user) metrics?

For an example of a successful decoupling, we can look to the energy utility industry. At one time, the energy utility model revolved around usage: they made more money when consumers consumed. In short, they were incentivized to encourage high and frequent usage.[4]

Once regulated, utilities decoupled the relationship between profit and energy use beyond a certain point—now referred to utility rate decoupling.[5] Energy utilities, use tiered pricing to disincentivize usage past a certain point. They do not profit directly from that heightened pricing. Instead those profits are allocated to renewable energy infrastructure and to accelerate the transition from extractive energy to regenerative energy.[6]

Imagine a world that respected users' attention and engagement!

Technology companies would be allowed to profit from a low threshold of basic usage, but beyond a certain point, the profits made would disincentivize addiction and the race to the bottom of the brain stem dynamics. The profits made beyond basic usage could then be invested in diverse media ecosystems as well as research for better humane technologies.

---

## Response to Written Questions Submitted by Hon. John Thune to Maggie Stanphill

*Question 1.* Dr. Stephen Wolfram, a witness at this hearing who has spent his life working on the science and technology of artificial intelligence, described Google and other Internet platforms as "automated content selection businesses," which he defined as entities that "work by getting large amounts of content they didn't themselves generate, then using what amounts to [artificial intelligence] to automatically select what content to deliver or to suggest to any particular user at any given time—based on data they've captured about the user." Does Google agree with these characterizations of its business by Dr. Wolfram? If not, please explain why not.

Answer. Our mission is to organize the world's information and make it universally accessible and useful.

We have many different products that are designed differently and serve this mission in different ways, including:

- Google Search organizes information about webpages in our Search index.
- YouTube provides a platform for people to upload videos to the open web with ease, and makes it easy for people to access those videos.
- Our advertising products allow businesses large and small to reach customers around the world and grow their businesses.

In many cases, we use automated processes to organize the vast array of information available on our platforms and the Web in order to provide relevant, useful information to users in a timely and accessible manner.

We believe in ensuring our users have choice, transparency, and control over how they engage with all of our products; for instance, Google Search and YouTube have options that allow users to operate them without any input from their personal data or browsing data, as well as the ability to turn off autoplay of videos suggested by YouTube's recommendation system.

*Question 2.* In Dr. Wolfram's prepared testimony, he formulates possible market-based suggestions for large Internet platforms to consider that would "leverage the

---

[4] Eto, Joseph, *et al.,* "The Theory and Practice of Decoupling Utility Revenues from Sales." Utilities Policy, vol. 6, no. 1, 1997, pp. 43–55., doi:10.1016/s0957–1787(96)00012–4

[5] Decoupling Policies: Options to Encourage Energy Efficiency Policies for Utilities, Clean Energy Policies in States and Communities, National Renewable Energy Laboratory (NREL) *https://www.energy.gov/eere/downloads/decoupling-policies-options-encourage-energy-efficiency-policies-utilities-clean*

[6] Decoupling Policies: Options to Encourage Energy Efficiency Policies for Utilities, Clean Energy Policies in States and Communities, National Renewable Energy Laboratory (NREL) *https://www.energy.gov/eere/downloads/decoupling-policies-options-encourage-energy-efficiency-policies-utilities-clean*

exceptional engineering and commercial achievements of the [automated content selection] businesses, while diffusing current trust issues about content selection, providing greater freedom for users, and inserting new opportunities for market growth." Specifically, Dr. Wolfram asked "Why does every aspect of automated content selection have to be done by a single business? Why not open up the pipeline, and create a market in which users can make choices for themselves?"

a. In what he labels "Suggestion A: Allow Users to Choose among Final Ranking Providers" Dr. Wolfram suggests that the final ranking of content a user sees doesn't have to be done by the same entity. Instead, there could be a single content platform but a variety of "final ranking providers", who use their own programs to actually deliver a final ranking to the user. Different final ranking providers might use different methods, and emphasize different kinds of content. But the point is to let users be free to choose among different providers.

Some users might prefer (or trust more) some particular provider—that might or might not be associated with some existing brand. Other users might prefer another provider, or choose to see results from multiple providers. Has Google considered Dr. Wolfram's suggestion to allow users to choose among final ranking providers? If so, please provide Google's reaction to Dr. Wolfram's proposal. If not, will Google commit to considering Dr. Wolfram's suggestion and providing a briefing to the Committee on its efforts to consider this suggestion?

Answer. Today, users have myriad choices when it comes to finding and accessing all types of content online. There are a variety of providers that organize information in different ways.

For general-purpose search engines, consumers can choose among a range of options: Bing, Yahoo, and many more. DuckDuckGo, for instance, a relatively new search engine provider, hit a record 1 billion monthly searches in January 2019, demonstrating that a new entrant can compete in this space.

There are many ways consumers find and access news content on the Internet. They navigate directly to sites and use dedicated mobile apps. They access news articles via social media services like Twitter and Facebook. And they use aggregators like News 360 and Drudge Report.

It has never been easier for a new entrant to build and become a new 'final ranking provider' for end users. Developers today can build on free repositories of web index data, like *Common Crawl,* to build new search engines. This is the kind of underlying, common content "platform" Dr. Wolfram seems to describe.

b. In what he labels "Suggestion B: Allow Users to Choose among Constraint Providers" Dr. Wolfram suggests putting constraints on results that automated content businesses generate, for example forcing certain kinds of balance. Much like final ranking providers in Suggestion A, there would be constraint providers who define sets of constraints. For example, a constraint provider could require that there be on average an equal number of items delivered to a user that are classified (say, by a particular machine learning system) as politically left-leaning or politically right-leaning. Constraint providers would effectively define computational contracts about properties they want results delivered to users to have. Different constraint providers would define different computational contracts. Some might want balance; others might want to promote particular types of content, and so on. But the idea is that users could decide what constraint provider they wish to use. Has Google considered Dr. Wolfram's suggestion to allow users to choose among constraint providers? If so, please provide Google's reaction to Dr. Wolfram's proposal. If not, will Google commit to considering Dr. Wolfram's suggestion and providing a briefing to the Committee on its efforts to consider this suggestion?

Answer. Google cares deeply about giving users transparency, choice and control in our products and services. We offer a number of resources to help users better understand the products and services we provide. For example, users can control what Google account activity is used to customize their experiences, including adjusting what data is saved to their Google account, at *myaccount.google.com.* If users wish to consume content in a different way, there are many other platforms and websites where they can do so, as discussed above.

*Question 3.* Does Google believe that algorithmic transparency is a policy option Congress should be considering? If not, please explain why not.

Answer. Transparency has long been a priority at Google to help our users understand how our products work. We must balance this transparency with the need to ensure that bad actors do not game our systems through manipulation, spam, fraud and other forms of abuse. Since Google launched our first *Transparency Report* in 2010, we've been sharing data that sheds light on how government actions and policies affect privacy, security, and access to information online. For Search, our *How Search Works* site provides extensive information to anyone interested in learning

more about how Google Search, our algorithms, and Search features operate. The site includes information on our approach to algorithmic *ranking.* We offer extensive *resources to all webmasters* to help them succeed in having their content discovered online. We also publish our 160 page *Search Quality Evaluator Guidelines,* which explain in great detail what our search engine is aiming to achieve, and which form a crucial part of the process by which we assess proposed changes to our algorithms.

It's important to note, however, that there are tradeoffs with different levels of transparency, and we aim to balance various sensitivities. For example, disclosing the full code powering our product algorithms would make it easier for malicious actors to manipulate or game our systems, and create vulnerabilities that would represent a risk to our users–while failing to provide meaningful, actionable information to well-meaning users or researchers, notably due to the scale and the pace of evolution of our systems. Extreme model openness can also risk exposing user or proprietary information, causing privacy breaches or threatening the security of our platforms.

Regarding transparency in AI algorithms more broadly, in our own consumer research, we've seen that access to underlying source code is not useful to users. Rather, we have found that algorithmic explanation is more useful. We've identified a few hallmarks of good explanations: it accurately conveys information regarding the system prediction or recommendation; is clear, specific, relatable, and/or actionable; boosts understanding of the overall system; and takes appropriate account of context. In our research we have been demonstrating progress in designing interpretable AI models, model understanding, and data and model cards for more transparent model reporting (see our *Responsible AI Practices* for a full list of technical recommendations and work). And we've outlined more details where government, in collaboration with civil society and AI practitioners, has a crucial role to play in AI explainability standards, among other areas, in our paper *Perspectives on Issues in AI Governance.*

*Question 4.* Does Google believe that algorithmic explanation is a policy option that Congress should be considering? If not, please explain why not.

Answer. Transparency has long been a priority at Google to help our users understand how our products work. We must balance this transparency with the need to ensure that bad actors do not game our systems through manipulation, spam, fraud and other forms of abuse. Since Google launched our first *Transparency Report* in 2010, we've been sharing data that sheds light on how government actions and policies affect privacy, security, and access to information online. For Search, our *How Search Works* site provides extensive information to anyone interested in learning more about how Google Search, our algorithms, and Search features operate. The site includes information on our approach to algorithmic *ranking.* We offer extensive resources to all webmasters to help them succeed in having their content discovered online. We also publish our 160 page *Search Quality Evaluator Guidelines,* which explain in great detail what our search engine is aiming to achieve, and which form a crucial part of the process by which we assess proposed changes to our algorithms.

It's important to note, however, that there are tradeoffs with different levels of transparency, and we aim to balance various sensitivities. For example, disclosing the full code powering our product algorithms would make it easier for malicious actors to manipulate or game our systems, and create vulnerabilities that would represent a risk to our users–while failing to provide meaningful, actionable information to well-meaning users or researchers, notably due to the scale and the pace of evolution of our systems. Extreme model openness can also risk exposing user or proprietary information, causing privacy breaches or threatening the security of our platforms.

Regarding transparency in AI algorithms more broadly, in our own consumer research, we've seen that access to underlying source code is not useful to users. Rather, we have found that algorithmic explanation is more useful. We've identified a few hallmarks of good explanations: it accurately conveys information regarding the system prediction or recommendation; is clear, specific, relatable, and/or actionable; boosts understanding of the overall system; and takes appropriate account of context. In our research we have been demonstrating progress in designing interpretable AI models, model understanding, and data and model cards for more transparent model reporting (see our *Responsible AI Practices* for a full list of technical recommendations and work). And we' outlined more details where government, in collaboration with civil society and AI practitioners, has a crucial role to play in AI explainability standards, among other areas, in our paper *Perspectives on Issues in AI Governance.*

*Question 5.* At the hearing, I noted that the artificial intelligence behind Internet platforms meant to enhance user engagement also has the ability, or at least the

potential, to influence the thoughts and behaviors of literally billions of people. Does Google agree with that statement? If not, please explain why not.

Answer. We strongly believe that AI can improve lives in a number of ways, though we also recognize that AI is a rapidly evolving technology that must be applied responsibly. For these reasons, we assess all of our AI applications in accordance with our *Google AI Principles:* be socially beneficial, avoid creating or reinforcing unfair bias, be built and tested for safety, be accountable to people, incorporate privacy design principles, uphold high standards of scientific excellence, and be made available for uses in accordance with these principles. Following these principles, we do not build AI products for the purpose of manipulating users. Furthermore, it would not be in our business interest to engage in activities that risk losing user trust.

Rather, like our other technologies, we are using AI to provide a better experience for our users, and our efforts are already proving invaluable in different ways. For example, nearly 1 billion unique users use Google Translate to communicate across language barriers, and more than 1 billion users use Google Maps to navigate roads, explore new places, and visualize places from the mountains to Mars. We also recently introduced an *AI-powered app called Bolo* to help improve childrens' reading skills; early results in India demonstrate that 64 percent of children showed an improvement in reading proficiency in just 3 months.

These opportunities to use AI for social good come with significant responsibility, and we have publicly outlined our commitment to responsible AI development—including algorithmic accountability and explainability—in the *Google AI Principles* (also see our *Responsible AI Practices* for a full list of technical recommendations and work).

*Question 6.* YouTube has offered an autoplay feature since 2015. The company also offers users the option of disabling autoplay. To date, what percentage of YouTube users have disabled autoplay?

Answer. Autoplay is an optional feature we added based on user feedback, as users wanted an option for a smoother YouTube experience, like listening to the radio or having a TV channel on in the background. We added an easy on/off toggle for the feature so that users can make a choice about whether they want to keep autoplay enabled, depending on how they are using the platform in a given session. Many users have chosen to disable autoplay in some situations and enable it in others. Our priority is to provide users with clear ways to use the product according to their specific needs.

*Question 7.* How many minutes per day do users in the United States spend, on average, watching content from YouTube? How has this number changed since YouTube added the autoplay feature in 2015?

Answer. YouTube is a global platform with over 2 billion monthly logged-in users. Every day people watch over a billion hours of video and generate billions of views. More than 500 hours of content are uploaded to YouTube every minute. We are constantly making improvements to YouTube's features and systems to improve the user experience, and would not attribute changes in user behavior over the course of four years to a single product change.

*Question 8.* What percentage of YouTube video views in the United States and worldwide are the result of clicks and embedded views from social media?

Answer. YouTube provides a number of ways for users to discover content, including through social media. The ways users choose to engage with the platform vary depending on their individual preferences, the type of content, and many other factors.

*Question 9.* What percentage of YouTube video views in the United States and worldwide are the result of YouTube automatically suggesting or playing another video after the user finishes watching a video?

Answer. A significant percentage of video views on YouTube come from recommendations. Overall, a majority of video views on YouTube come from recommendations. This includes what people see on their home feed, in search results and in Watch Next panels. Recommendations are a popular and useful tool that helps users discover new artists and creators and surface content to users that they might find interesting or relevant to watch next. The ways users choose to engage with recommendations and YouTube's autoplay feature vary depending on user preferences, the type of content they are watching, and many other factors. Many users like to browse the Watch Next panel and choose the next video they want to play. In some cases, users want to continue to watch videos without having to choose the next video, for example if they are using YouTube to listen to music or to follow a set playlist of content. To provide users with choices, YouTube has an easy toggle switch to turn autoplay off if users do not want to have videos automatically play.

*Question 10.* What percentage of YouTube video views in the United States and worldwide are the result of users searching YouTube.com?

Answer. When users first start using YouTube, they often begin by searching for a video. YouTube search works similar to Google search—users type a search query into the search box, and we present a list of videos or YouTube channels that are relevant to that search query. Videos are ranked based on a number of factors including how well the title and description match the query, what is in the video content, and how satisfied previous users were when they viewed these videos. The ways users choose to find content, including through the YouTube home page, searches, and recommendations vary depending on user preferences, the type of content, and many other factors.

*Question 11.* In 2018, YouTube started labeling videos from state-funded broadcasters. What impact, if any, have these labels had on the rate that videos from these channels are viewed, clicked on, and shared by users?

Answer. If a channel is owned by a news publisher that is funded by a government, or publicly funded, an information panel providing *publisher context* may be displayed on the watch page of the videos on its channel. YouTube also has other information panels, including to provide topical context for well-established historical and scientific topics that have often been subject to misinformation online, like the moon landing. We have delivered more than 2.5 billion impressions across all of our information panels since July 2018.

*Question 12.* During the hearing, I discussed my efforts to develop legislation that will require Internet platforms to give its users the option to engage with the platform without having the experience shaped by algorithms driven by user-specific data. In essence, the bill would require Internet platforms like Google to provide users with the option of a "filter bubble-free" view of services such as Google search results, and enabling users to toggle between the opaque artificial intelligence driven personalized search results and the "filter bubble-free" search results. Does Google support, at least in principle, providing its users with the option of a "filter bubble-free" experience of its search results?

Answer. There is very little personalization in organic Search results based on users' inferred interests or Search history before their current session. It doesn't take place often and generally doesn't significantly change organic Search results from one person to another. Most differences that users see between their organic Search results and those of another user typing the same Search query are better explained by other factors such as a user's location, the language used in the search, the distribution of Search index updates throughout our data centers, and more. One of the most common reasons results may differ between people involves localized organic search results, when listings are customized to be relevant for anyone in a particular area. Localization isn't personalization because everyone in the same location gets similar results. Localization makes our search results more relevant. For example, people in the U.S. searching for "football" do not generally want UK football results, and vice versa. People searching for "zoos" in one area often want locally-relevant listings.

Search does include some features that personalize results based on the activity in their Google account. For example, if a user searches for "events near me" Google may tailor some recommendations to event categories we think they may be interested in. These systems are designed to match a user's interests, but they are not designed to infer sensitive characteristics like race or religion. Overall, Google strives to make sure that our users continue to have access to a diversity of websites and perspectives.

Anyone who doesn't want personalization using account-based activity can disable it using the Web & App Activity *setting.* Users can also choose to keep their search history stored but exclude Chrome and app activity.

*Question 13.* In 2013, former Google Executive Chairman Eric Schmidt wrote that modern technology platforms like Google "are even more powerful than most people realize." Does Google agree that it is even more powerful than most people realize? If not, please explain why not.

Answer. We are committed to providing users with powerful tools, and our users look to us to provide relevant, authoritative information. We work hard to ensure the integrity of our products, and we've put a number of checks and balances in place to ensure they continue to live up to our standards. We also recognize the important role of governments in setting rules for the development and use of technology. To that end, we support Federal privacy legislation and proposed a legislative framework for privacy last year.

*Question 14.* Does Google believe it is important for the public to better understand how it uses artificial intelligence to make inferences from data about its users?

Answer. Automated predictions and decision making can improve lives in a number of ways, from recommending music to monitoring a patient's vital signs, and we believe public explainability is crucial to being able to question, understand, and trust machine learning systems. We've identified a few hallmarks of good explanations: they accurately convey information regarding the system prediction or recommendation; are clear, specific, relatable, and/or actionable; boost understanding of the overall system; and take appropriate account of context.

We've also been taken numerous steps in our technical research to make our algorithms more understandable and transparent (see our *Responsible AI Practices* for a full list of technical recommendations and work), including:

- We've developed a lot of research and tools to help people better understand their data and design more interpretable models.
- We're also working on *visualizing* what's going on inside deep neural nets.
- And explainability is built into some projects such as *predicting cardiovascular risk* from images of the retina—our model shows what parts of the image most contributed to the prediction.

Explainability when it comes to machine learning is something we take very seriously, and we'll continue to work with researchers, academics, and public policy groups to make sure we're getting this right. It's important to note that government, in collaboration with civil society and AI practitioners, also has a crucial role to play in AI explainability standards, and we've outlined more details in our paper *Perspectives on Issues in AI Governance.*

*Question 15.* Does Google believe that its users should have the option to engage with their platform without being manipulated by algorithms powered by its users' own personal data? If not, please explain why not.

Answer. Google cares deeply about giving users transparency, choice and control in our products and services. We offer a number of resources to help users better understand the products and services we provide. These resources include plain-English and easy-to-understand instructions about how users can make meaningful privacy and security choices on Google products and more generally, online. For example, Google's Privacy Policy (available at *https://policies.google.com/privacy*) includes short, educational videos about the type of data Google collects.

*Question 16.* Does Google design its algorithms to make predictions about each of its users?

Answer. There are indeed some places in our products where we endeavor to make predictions about users in order to be more helpful, for example in our Maps products we might suggest that a user plan to leave early for a trip to the airport depending on the user's settings and the data we have. Specifically, this might happen when the user has received an e-mail confirmation from an airline suggesting the user may be flying that day; combining this with traffic data that shows an accident has stalled traffic on a nearby road may trigger us to prompt the user to leave early to allow for additional traffic.

As described in response to other answers, we offer a number of resources to help users better understand the products and services we provide including our uses of data. These resources include plain-English and easy-to-understand instructions about how users can make meaningful privacy and security choices on Google products and more generally, online. For example, Google's Privacy Policy (available at *https://policies.google.com/privacy*) includes short, educational videos about the type of data Google collects.

*Question 17.* Does Google design its algorithms to select and display content on its Search service in a manner that seeks to optimize user engagement?

Answer. The purpose of Google Search is to help users find the information they are looking for on the web. Keeping them on the Google Search results page is not our objective.

*Question 18.* Does Google design its algorithms to select and display content on its YouTube service in a manner that seeks to optimize user engagement?

Answer. We built our YouTube recommendation system to help users find new content, discover their next favorite creator, or learn more about the world. We want to provide more value to our users, and we work hard to ensure that we only recommend videos that will create a satisfying and positive user experience.

We update our systems continuously, and have been focusing on information quality and authoritativeness, particularly in cases like breaking news, or around sen-

sitive or controversial topics. In January of this year, we announced the latest of our improvements to our recommendation system is to greatly reduce recommendations of borderline content and content that could misinform users in harmful ways. In June, 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, 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.

*Question 19.* Does Google design its algorithms to select and display content on its News service in a manner that seeks to optimize user engagement?

Answer. The algorithms used for our news experiences are designed to analyze hundreds of different factors to identify and organize the stories journalists are covering, in order to elevate diverse, trustworthy information.

*Question 20.* Tristan Harris, a witness at this hearing who was formerly an employee of Google, stated that what we're experiencing with technology is an increasing asymmetry of power between Internet platforms and users, and that Internet platforms like Google essentially have a supercomputer pointed at each user's brain that can predict things about the user that the user does not even know about themselves.

a. Does Google agree that there is an asymmetry of power between it and its users?

Answer. Users have transparency, choice, and control when it comes to how they use our platforms, and what information they choose to provide to us in order for us to customize their user experience. Users are in control of how they use our products, and if we do not earn their trust, they will go elsewhere.

b. What predictions does Google seek to make about each user?

Answer. There are indeed some places in our products where we endeavor to make predictions about users in order to be more helpful, for example in our Maps products we might suggest that a user plan to leave early for a trip to the airport depending on the user's settings and the data we have. Specifically, this might happen when the user has received an e-mail confirmation from an airline suggesting the user may be flying that day; combining this with traffic data that shows an accident has stalled traffic on a nearby road may trigger us to prompt the user to leave early to allow for additional traffic.

We offer a number of resources to help users better understand the products and services we provide including our uses of data. These resources include plain-English and easy-to-understand instructions about how users can make meaningful privacy and security choices on Google products and more generally, online. For example, Google's Privacy Policy (available at *https://policies.google.com/privacy*) includes short, educational videos about the type of data Google collects.

c. Does Google agree with Tristan Harris's characterization that Internet platforms like Google essentially have a supercomputer pointed at each user's brain?

Answer. No, we do not agree with that characterization. We work hard to provide search results that are relevant to the words in a user's search, and with some products, like YouTube, we are clear when we are offering recommendations based on a user's preferences, but users retain control through their settings and controls to optimize their own experience.

*Question 21.* Does Google seek to optimize user engagement?

Answer. We seek to optimize user experience. We have a multitude of tools and options to help our users interact with our products and platforms in ways that work best for them. We are committed to keeping our users safe online, and providing them with positive experiences. We do this through technological innovation, strong community guidelines, extensive education and outreach, and providing our users with choice, transparency and control over their experience. Our Digital Wellbeing Initiative focuses on these issues. More information about how we help our users find the balance with technology that feels right to them can be found on our *Digital Wellbeing site.*

*Question 22.* How does Google optimize for user engagement?

Answer. As mentioned above in question 21, we optimize for user experience rather than user engagement, and give our users a number of tools to control their use of our platforms through our Digital Wellbeing product features. We continue to invest in these efforts to help users find the balance with technology that is right for them.

*Question 23.* How does Google personalize search results for each of its users?

Answer. Search does not require personalization in order to provide useful organic search results to users' queries. In fact, there is very little personalization in organic

Search based on users' inferred interests or Search history before their current session. It doesn't take place often and generally doesn't significantly change organic Search results from one person to another. Most differences that users see between their organic Search results and those of another user typing the same Search query are better explained by other factors such as a user's location.

For instance, if a user in Chicago searches for "football", Google will most likely show results about American football first. Whereas if the user searches "football" in London, Google will rank results about soccer higher. Overall, Google strives to make sure that our users have access to a diversity of websites and perspectives.

Anyone who doesn't want personalization using account-based activity can disable it using the Web & App Activity setting. Users can also choose to keep their search history stored but exclude Chrome and app activity. "Incognito" search mode or a similar private browsing window can also allow users to conduct searches without having account-based activity inform their search results.

Search ads are ranked in a similar manner to organic Search results. The match between a user's search terms and the advertisers' selected keywords is the key factor underlying the selection of ads users see.

In relation to Google Ads, users can turn off personalized ads at *myaccount.google.com.* Once they've turned off personalization, Google will no longer use Account information to personalize the user's ads. Ads can still be targeted with info like the user's general location or the content of the website they are visiting.

*Question 24.* How does Google personalize what content it recommends for its users to see on YouTube?

Answer. A user's activity on YouTube, Google and Chrome may influence their YouTube search results, recommendations on the Home page, in-app notifications and suggested videos among other places.

There are several ways that users can influence these recommendations and search results. They can remove specific videos from their watch history and queries from their search history, pause their watch and search history, or start afresh by clearing their watch and search history.

*Question 25.* How does Google personalize what content its users see on its News service?

Answer. Whether our users are checking in to see the top news of the day or looking to dive deeper on an issue, we aim to connect them with the information they're seeking, in the places and formats that are right for them. To this end, Google provides three distinct but interconnected ways to find and experience the news across our products and devices: top news stories for everyone, personalized news, and additional context and perspectives.

1. Top News for everyone: For users who want to keep up with the news, they need to know what the important stories are at any point in time. With features such as Headlines in Google News and Breaking News on YouTube, we identify the major stories news sources are covering. This content is not personalized to individuals, but does vary depending on region and location settings. Google's technology analyzes news across the web to determine the top stories for users with the same language settings in a given country, based primarily on what publishers are writing about. Once these stories are identified, algorithms then select which specific articles or videos to surface and link to for each story, based on factors such as the prominence and freshness of the article or video, and authoritativeness of the source.

2. Personalized news: Several Google news experiences show results that are personalized for our users. These include Discover, For you in Google News, and the Latest tab of the YouTube app on TVs. Our aim is to help our users stay informed about the subjects that matter to them, including their interests and local community. Google relies on two main ways to determine what news may be interesting to our users. In the experiences mentioned above, users can specify the topics, locations, and publications they're interested in, and they will be shown news results that relate to these selections. Additionally, depending on *their account settings,* our algorithms may suggest content based on a user's past activity on Google products. Algorithms rank articles based on factors like relevance to their interests, prominence and freshness of the article, and authoritativeness of the source. Google's news algorithms do not attempt to personalize results based on the political beliefs or demographics of news sources or readers. Users can control what account activity is used to customize their news experiences, including adjusting what data is saved to their Google account, at *myaccount.google.com.* In some Google products, such as Google News and Discover, users can also follow topics of interest, follow or hide specific publishers, or tell us when they want to see similar articles more or less frequently.

3. Additional contexts and perspectives: A central goal of Google's news experiences is to provide access to context and diverse perspectives for stories in the news. By featuring unpersonalized news from a broad range of sources, Google empowers people to deepen their understanding of current events and offers an alternative to exclusively personalized news feeds and individual sources that might only represent a single perspective.

a. Search experiences: When users search for something on Google, they have access to information and perspectives from a broad range of publishers from across the web. If they search for a topic that's in the news, their results may include some news articles labeled "Top stories" at the top of the results, featuring articles related to the search and a link to more related articles on the News tab. Users can also search for news stories and see context and multiple perspectives in the results on news.google.com, news on the Assistant, and within the "Top News" section of search results on YouTube. These results are not personalized. Our algorithms surface and organize specific stories and articles based on factors like relevance to the query, prominence and freshness of the article, and authoritativeness of the publisher. Users can always refine the search terms to find additional information.

b. In-product experiences: In some news experiences, such as "Full coverage" in Google News, we show related articles from a variety of publishers alongside a given article. These results are not personalized. In providing additional context on a story, we sometimes surface videos, timelines, fact check articles, and other types of content. Algorithms determine which articles to show, and in which order, based on a variety of signals such as authoritativeness, relevance, and freshness.

*Question 26.* Does Google engage in any effort to change its user's attitudes? [response below]

*Question 27.* Does Google engage in any effort to change its user's behaviors? [response below]

*Question 28.* Does Google engage in any effort to influence its users in any way? [response below]

*Question 29.* Does Google engage in any effort to manipulate its users in any way? [response below]

*Question 30.* Do rankings of search results provided by Google have any impact on consumer attitudes, preferences, or behavior?

Answer. We answer questions 26, 27, 28, 29, and 31 together. When users come to Google Search, our goal is to connect them with useful information as quickly as possible. That information can take many forms, and over the years the search results page has evolved to include not only a list of blue links to pages across the web, but also useful features to help users find what they're looking for even faster. For our Knowledge Graph allows us to respond to queries like "Bessie Coleman" with a *Knowledge Panel* with facts about the famous aviator. Alternatively, in response to queries like "how to commit suicide", Google has worked with the National Suicide Prevention Hotline to surface a results box at the top of the search results page with the organization's phone number and website that can provide help and support. The goal of this type of result is to connect vulnerable people in unsafe situations to reliable and free support as quickly as possible.

For other questions, Search is a tool to explore many angles. We aim to make it easy to discover information from a wide variety of viewpoints so users can form their own understanding of a topic. We feel a deep sense of responsibility to help all people, of every background and belief, find the high-quality information they need to better understand the topics they care about and we try to make sure that our users have access to a diversity of websites and perspectives.

When it comes to the ranking of our search results—the familiar "blue links" of web page results—the results are determined algorithmically. We do not use human curation to collect or arrange the results on a page. Rather, we have automated systems that are able to quickly find content in our index—from the hundreds of billions of pages we have indexed by crawling the web—that are relevant to the words in the user's search. To rank these, our systems take into account a number of factors to determine what pages are likely to be the most helpful for what a user is looking for. We describe this in greater detail in our *How Search Works* site.

*Question 31.* The website *moz.com* tracks every confirmed and unconfirmed update Google makes to its search algorithm. In 2018, Google reported 3,234 updates. However, moz.com reported that there were also at least six unconfirmed algorithm updates in 2018. Does Google publicly report every change it makes to its search algorithm? If not, why not?

Answer. We report the number of changes we make to Google Search each year on our *How Search Works* website. To prevent bad actors from gaming our systems, we do not publicly report on the nature of each change.

*Question 32.* Does an item's position in a list of search results have a persuasive impact on a user's recollection and evaluation of that item?

Answer. We aim to make it easy to discover information from a wide variety of viewpoints so users can form their own understanding of a topic. We feel a deep sense of responsibility to help all people, of every background and belief, find the high-quality information they need to better understand the topics they care about and we try to make sure that our users have access to a diversity of websites and perspectives.

When it comes to the ranking of our search results—the familiar "blue links" of web page results—the results are determined algorithmically. We do not use human curation to collect or arrange the results on a page. Rather, we have automated systems that are able to quickly find content in our index—from the hundreds of billions of pages we have indexed by crawling the web—that are relevant to the words in the user's search. To rank these, our systems take into account a number of factors to determine what pages are likely to be the most helpful for what a user is looking for. We describe this in greater detail in our *How Search Works* site.

*Question 33.* A study published in 2015 in the Proceedings of the National Academy of Sciences entitled "The Search Engine Manipulation Effect (SEME) and its Possible Impact on the Outcomes of Elections" discussed an experiment where the study's authors (one of whom is a former editor in chief of Psychology Today) sought to manipulate the voting preferences of undecided eligible voters throughout India shortly before the country's 2014 national elections. The study concluded that the result of this and other experiments demonstrated that (i) biased search rankings can shift the voting preferences of undecided voters by 20 percent or more, (ii) the shift can be much higher in some demographic groups, and (iii) search ranking bias can be masked so that people show no awareness of the manipulation. This is a rigorously peer-reviewed study in the Proceedings of the National Academy of Sciences of the United States of America, one of the world's most-cited scientific journals, which strives to publish only the highest quality scientific research. Has Google carefully reviewed this study and taken steps to address the conclusions and concerns highlighted in this study? If so, please describe the steps taken to address this study. If Google has not taken steps to address this study, please explain why not?

Answer. Google takes these allegations very seriously. Elections are a critical part of the democratic process and Google is committed to helping voters find relevant, helpful, and accurate information. Our job—which we take very seriously—is to deliver to users the most relevant and authoritative information out there. And studies have shown that we do just that. It would undermine people's trust in our results, and our company, if we were to change course. There is absolutely no truth to Mr. Epstein's hypothesis. Google is not politically biased and Google has never re-ranked search results on any topic (including elections) to manipulate user sentiment. Indeed, we go to extraordinary lengths to build our products and enforce our policies in an analytically objective, apolitical way. We do so because we want to create tools that are useful to all Americans. Our search engine and our platforms reflect the online world that is out there.

We work with external Search Quality Evaluators from diverse backgrounds and locations to assess and measure the quality of search results. Any change made to our Search algorithm undergoes rigorous user testing and evaluation. The ratings provided by these Evaluators help us benchmark the quality of our results so that we can continue to meet a high bar for users of Google Search all around the world. We publish our Search Quality Evaluator Guidelines and make them publicly available on our *How Search Works* website.

On Google Search, we aim to make civic information more easily accessible and useful to people globally as they engage in the political process. We have been building products for over a decade that provide timely and authoritative information about elections around the world and help voters make decisions that affect their communities, their cities, their states, and their countries. In 2018, for example, we helped people in the U.S. access authoritative information about registering to vote, locations of polling places, and the mechanics of voting. We also provided information about all U.S. congressional candidates on the Search page in Knowledge Panels, and provided the opportunity for those candidates to make their own statements in those panels. On election day, we surfaced election results for U.S. congressional races directly in Search in over 30 languages. We have also partnered with organizations like the *Voting Information Project,* with whom we've worked since 2008 to help millions of voters get access to details on where to vote, when to vote, and who

will be on their ballots. This project has been a collaboration with the offices of 46 Secretaries of State to ensure that we are surfacing fresh and authoritative information to our users.

In addition to Search results about election information, we have made voting information freely available through the *Google Civic Information API,* which has allowed developers to create useful applications with a civic purpose. Over 400 sites have embedded tools built on the Civic Information API; these include sites of candidates, campaigns, government agencies, nonprofits, and others who encourage and make it easier for people to get to the polls.

---

RESPONSE TO WRITTEN QUESTION SUBMITTED BY HON. AMY KLOBUCHAR TO MAGGIE STANPHILL

*Question.* Recent news articles have reported that YouTube's automated video recommendation system—which drives 70 percent of the platform's video traffic—has been recommending home videos of children, including of children playing in a swimming pool, to users who have previously sought out sexually themed content.

Reports have also stated that YouTube has refused to turn off its recommendation system on videos of children—even though such videos can be identified automatically. Why has YouTube declined to take this measure?

What steps are being taken to identify these kinds of flaws in YouTube's recommendation system?

Answer. We are deeply committed to protecting children and families online, and we work very hard to ensure that our products, including YouTube, offer safe, age-appropriate content for children. We also enforce a strong set of policies to protect minors on YouTube, including those that prohibit exploiting minors, encouraging dangerous or inappropriate behaviors, and aggregating videos of minors in potentially exploitative ways. In the first quarter of 2019 alone, we removed more than 800,000 videos for violations of our child safety policies, the majority of these before they had ten views.

The vast majority of videos featuring minors on YouTube do not violate our policies and are innocently posted—a family creator providing educational tips, or a parent sharing a proud moment. But when it comes to kids, we take an extra cautious approach towards our enforcement and we're always making improvements to our protections. Earlier this year we made significant changes to our systems so we could limit recommendations of videos featuring minors in risky situations. We made this change recognizing the concern that minors could be at risk of online or offline exploitation if those types of videos were recommended. We have applied these recommendations changes to tens of millions of videos across YouTube.

We also recognize that a great deal of content on the platform that features children is not violative or of interest to bad actors, including child actors in mainstream-style content, kids in family vlogs, and more. Turning off recommendations for all videos with children would cut off the ability for these types of creators to reach audiences and build their businesses. We do not think that type of solution is necessary when we can adjust for the type of videos that are of concern. That said, we are always evaluating our policies and welcome further conversations about efforts to protect children online.

---

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO MAGGIE STANPHILL

*A.I. Accountability and Civil Rights.* A peer company, Facebook, announced that it is conducting an audit to identify and address discrimination. It has also formed Social Science One, which provides external researchers with data to study the platform's effects of social media on democracy and elections.

*Question 1.* What data has Google provided to outside researchers to scrutinize for discrimination and other harmful activities?

Answer. We have long invested in tools and reporting systems that enable outside researchers to form an understanding of our products and practices:

- Our Google Trends product, which has been freely available to the public since 2006, enables third party researchers to explore trending searches on Google and YouTube, at scale and over time.
- The open nature of our services makes it possible for researchers to seek and analyze content relating to these trends easily: every Search and YouTube user has access to the same content, with limited exceptions including age-gating, private videos, or legal restrictions across countries. Many researchers and aca-

demics have scrutinized our products and published extensive papers and analysis commenting on our practices.

- Google's Transparency Report, launched in 2010, sheds light on the many ways that the policies and actions of governments and companies impact user privacy, security, and access to information, and provide. Recent additions include our YouTube Community Guidelines Enforcement report, which provides information about how we enforce our content policies on YouTube via flags and automated systems, and our Political Advertising Transparency report, which we launched prior to the 2018 midterms in the U.S. and have since expanded to provide data for the 2019 India elections and the recent EU Parliament elections. We are always looking for ways to share new data and make our reports easy to use and interpret.

*Question 2.* Has Google initiated a civil rights audit to identify and mitigate discriminatory bias on its platform?

Answer. While we have not conducted a civil rights audit, we have long had senior staff assigned as our Liaison to the U.S. Civil and Human Rights community and work with a large number of the most recognized and subject matter relevant organizations in the US, including: the Leadership Conference on Civil and Human Rights, the NAACP, the League of United Latin American Citizens, Asian Americans Advancing Justice, the National Hispanic Media Coalition, Muslim Public Affairs Council, the National Congress of American Indians, Human Rights Campaign and the Foundations of the relevant caucuses including Congressional Black Caucus Foundation, Congressional Hispanic Caucus Institute and Leadership Institute and Asian Pacific American Institute for Congressional Studies.

We engage in regular briefings, meetings and case-by-case sessions to consult with these leaders and organizations in order to ensure that our products, policies, and tools comply with civil and human rights standards.

Google's work also includes our innovative tech policy diversity pipeline initiative—the Google Next Generation Policy Leaders, a program now in its 3rd year of identifying, training and supporting the Nation's emerging social justice leaders, across a range of tech policy areas.

*Question 3.* You note in remarks that Google supports legislation for the NIH to study the developmental effects of technology on children. However, Google holds all the important data on this issue. What data does Google plan to provide to the NIH to support such a study?

Answer. We recognize that Google is one of the many entities that hold important data on this issue. We have supported CAMRA which will make highly necessary NIH funding available to researchers working on issues surrounding children and media. While we are not able to provide user data to researchers directly, we sometimes work with researchers to field studies with users who have explicitly opted in to this purpose. In the case of our services for children this would also require parental consent. We would be happy to consider participating in NIH-funded studies if they involved the proper user consent.

*Question 4.* Who at Google is responsible for protecting civil rights and civil liberties in its A.I. system and what is their role in product development?

Answer. The Google AI Principles specifically state that we will not design or deploy AI technologies whose purpose contravenes widely accepted principles of international law and human rights. There are several teams with relevant experts who hold responsibility for helping Google live up to its AI Principles commitment and protecting civil rights and liberties in design and development of our AI systems, including Responsible Innovation, Trust & Safety, Privacy & Security, Research, Product Inclusion, Human Rights & Social Impact, and Government Affairs & Public Policy.

*Loot Boxes.* Google has a substantial role as a gatekeeper to protect consumers with the Play Store and Android.

*Question 5.* Has Play Store restricted loot boxes in games that can be played children and minors?

Answer. Google provides tools and safeguards to help parents guide their child's online experience and make informed choices with regard to the apps their child can use. Children under the age of thirteen (age varies based on country) must have their account managed by a parent through Family Link. Family Link's parental controls, by default, require parental approval for app downloads, as well as for app and in-app purchases. Family Link also gives parents the ability to filter the apps that their child can browse in the Play store by rating and to block apps that have previously been downloaded on their child's Android and ChromeOS devices.

Our gaming content ratings system helps inform parents and guardians of the type of content displayed in an in-game experience. We are also actively working with third party partners to ensure their ratings reflect loot box experiences in game play. As an additional safeguard and to ensure parent oversight, Play also requires password authentication for all in-app purchases for apps that are in the Designed for Families Program.

*Question 6.* Why does the Play Store not specifically warn when games offer loot boxes?

Answer. We do provide notice to users that games include in-app purchases at the store level, and we do require disclosure of loot box probabilities before a purchase is made (in-game). Our current policy language is as follows: Apps offering mechanisms to receive randomized virtual items from a purchase (*i.e.,* "loot boxes") must clearly disclose the odds of receiving those items in advance of purchase.

*Data Privacy and Manipulative Technologies.*

*Question 7.* Has Google ever conducted research to determine if it could target users based on their emotional state?

*Question 8.* Who is responsible for ensuring that Google's ad targeting practices do not exploit users based on their emotional state and vulnerabilities?

Answer. We answer questions 7 and 8 together. To act responsibly and serve our users well, we regularly conduct research about how our products and services affect our users. For example, we collect data to understand whether our users find what they are looking for when they click on a particular ad in our search results page. We also run studies to learn whether users prefer ads that include additional information or features, such as use of photos. However, we do not allow ad personalization to our users based on sensitive emotional states. We also have clear policy restrictions prohibiting ad personalization using information about potential user vulnerabilities. Specifically, we don't allow ads that exploit the difficulties or struggles of users and we don't allow ads to be personalized based on categories related to personal hardships. Furthermore, we don't allow ads to be personalized based on mental health conditions or disabilities. Our Ads Product and Trust and Safety teams work in tandem to ensure this is enforced. More information about this can be found in our *Personalized Advertising Policy Principles.*

*YouTube's Promotion of Harmful Content.* According to your prepared remarks, YouTube has made changes to reduce the recommendation of content that "comes close to violating our community guidelines or spreads harmful misinformation." According to your account the number of views from recommendations for these videos has dropped by "over 50 percent in the U.S." These are views from YouTube's recommendation system—directed by YouTube itself—from systems that it controls.

*Question 9.* What specific steps has YouTube taken to end its recommendation system's practice of promoting content that sexualizes children?

Answer. We are deeply committed to protecting children and families online, and we work very hard to ensure that our products, including YouTube, offer safe, age-appropriate content for children. We also enforce a strong set of policies to protect minors on YouTube, including those that prohibit exploiting minors, encouraging dangerous or inappropriate behaviors, and aggregating videos of minors in potentially exploitative ways. In the first quarter of 2019 alone, we removed more than 800,000 videos for violations of our child safety policies, the majority of these before they had ten views.

The vast majority of videos featuring minors on YouTube do not violate our policies and are innocently posted—a family creator providing educational tips, or a parent sharing a proud moment. But when it comes to kids, we take an extra cautious approach towards our enforcement and we're always making improvements to our protections. Earlier this year we made significant changes to our systems so we could limit recommendations of videos featuring minors in risky situations. We made this change recognizing the concern that minors could be at risk of online or offline exploitation if those types of videos were recommended. We have applied these recommendations changes to tens of millions of videos across YouTube.

We also recognize that a great deal of content on the platform that features children is not violative or of interest to bad actors, including child actors in mainstream-style content, kids in family vlogs, and more. Turning off recommendations for all videos with children would cut off the ability for these types of creators to reach audiences and build their businesses. We do not think that type of solution is necessary when we can adjust for the type of videos that are of concern. That said, we are always evaluating our policies and welcome further conversations about efforts to protect children online.

*Question 10.* Why has the number of views for harmful content only dropped by half? Why hasn't the amount of traffic that YouTube itself is driving dropped to zero? You can control this.

Answer. This change to YouTube's recommendations system is a new effort that began in January and that we are still improving and rolling out at scale across the platform. Our systems are getting smarter about what types of videos should get this treatment, and we'll be able to apply it to even more borderline videos moving forward. As we do this, we'll also start raising up more authoritative content in recommendations.

As YouTube develops product features to deal with borderline content or misinformation, we also prioritize protecting freedom of expression and freedom of information. We develop and scale these types of changes carefully to try to avoid sweeping changes that may affect certain content that our systems may have a harder time distinguishing from borderline content and misinformation.

*Question 11.* Since you have quantified the amount of engagement with harmful content, what percentage of viewership does this represent overall for video views on YouTube?

Answer. Borderline content and content that could misinform users in harmful ways accounts for less than 1 percent of consumption on the platform.

*Question 12.* Under what conditions does YouTube believe it is appropriate for its recommendation system to promote content that violates its policies or is considered harmful misinformation?

Answer. We are committed to taking the steps needed to live up to our responsibility to protect the YouTube community from harmful content. When content violates our policies, we remove it. For example, between January to March 2019, we removed nearly 8.3 million videos for violating our Community Guidelines, the majority of which were first flagged by machines and removed before receiving a single view. During this same quarter, we terminated over 2.8 million channels and removed over 225 million comments for violating our Community Guidelines.

In addition to removing videos that violate our policies, we also want to reduce the spread of content that comes right up to the line. We are continuing to build on the pilot program we launched in January to reduce recommendations of borderline content and videos that may misinform users in harmful ways, to apply it at scale. This change relies on a combination of machine learning and real people, and so takes time to scale. We work with human evaluators and experts from all over the United States to help train the machine learning systems that generate recommendations. These evaluators are trained using public guidelines and provide critical input on the quality of a video.

It's important to note that this change only affects recommendations of what videos to watch, not whether a video is available on YouTube. Users can still access all videos that comply with our Community Guidelines.

The openness of YouTube's platform has helped creativity and access to information thrive. We think this change to our recommendations system strikes a balance between maintaining a platform for free speech and living up to our responsibility to users. We will continue to expand it in the U.S. and bring it to other countries.

*Recommendations from Ms. Richardson.* Ms. Richardson provided a set of recommendations in her remark for Congress to act, including:

1.) Require Technology Companies to Waive Trade Secrecy and Other Legal Claims That Hinder Oversight and Accountability Mechanisms

2.) Require Public Disclosure of Technologies That Are Involved in Any Decisions About Consumers by Name and Vendor

3.) Empower Consumer Protection Agencies to Apply "Truth in Advertising Laws" to Algorithmic Technology Providers

4.) Revitalize the Congressional Office of Technology Assessment to Perform Pre-Market Review and Post-Market Monitoring of Technologies

5.) Enhanced Whistleblower Protections for Technology Company Employees That Identify Unethical or Unlawful Uses of AI or Algorithms

6.) Require Any Transparency or Accountability Mechanism To Include A Detailed Account and Reporting of The "Full Stack Supply Chain"

7.) Require Companies to Perform and Publish Algorithmic Impact Assessments Prior to Public Use of Products and Services

During the hearing, I requested for you to respond in writing if possible.

*Question 13.* Please provide feedback to Ms. Richardson's suggestions for Congressional action.

Answer. Transparency has long been a priority at Google to help our users understand how our products work. We must balance this transparency with the need to ensure that bad actors do not game our systems through manipulation, spam, fraud and other forms of abuse. Since Google launched our first Transparency Report in 2010, we've been sharing data that sheds light on how government actions and policies affect privacy, security, and access to information online. For Search, our *How Search Works* site provides extensive information to anyone interested in learning more about how Google Search, our algorithms, and Search features operate. The site includes information on our approach to algorithmic ranking. We offer extensive resources to all webmasters to help them succeed in having their content discovered online. We also publish our 160 page *Search Quality Evaluator Guidelines* which explain in great detail what our search engine is aiming to achieve, and which form a crucial part of the process by which we assess proposed changes to our algorithms.

It's important to note, however, that there are tradeoffs with different levels of transparency, and we aim to balance various sensitivities. For example, disclosing the full code powering our product algorithms would make it easier for malicious actors to manipulate or game our systems, and create vulnerabilities that would represent a risk to our users–while failing to provide meaningful, actionable information to well-meaning users or researchers, notably due to the scale and the pace of evolution of our systems. Extreme model openness can also risk exposing user or proprietary information, causing privacy breaches or threatening the security of our platforms.

Regarding transparency in AI algorithms more broadly, in our own consumer research, we've seen that access to underlying source code is not useful to users. Rather, we have found that algorithmic explanation is more useful. We've identified a few hallmarks of good explanations: it accurately conveys information regarding the system prediction or recommendation; is clear, specific, relatable, and/or actionable; boosts understanding of the overall system; and takes appropriate account of context. In our research we have been demonstrating progress in designing interpretable AI models, model understanding, and data and model cards for more transparent model reporting (see our *Responsible AI Practices* for a full list of technical recommendations and work). And we've outlined more details where government—including Congress—in collaboration with civil society and AI practitioners, has a crucial role to play in AI explainability standards, among other areas, in our paper *Perspectives on Issues in AI Governance.*

It is important to note, there is no one-size-fits-all approach: the kind of explanation that is meaningful will vary by audience, since the factors emphasized and level of complexity that a layperson is interested in or can understand may be very different from that which is appropriate for an auditor or legal investigator. The nature of the use case should also impact the timing and manner in which an explanation can be delivered. Finally there are technical limits as to what is currently feasible for complex AI systems. With enough time and expertise, it is usually possible to get an indication of how complex systems function, but in practice doing so will seldom be economically viable at scale, and unreasonable requirements may inadvertently block the adoption of life-saving AI systems. A sensible compromise is needed that balances the benefits of using complex AI systems against the practical constraints that different standards of explainability would impose.

*Question 14.* What other steps or actions should Congress consider in regulating the use or consumer protection regarding persuasive technologies or artificial intelligence?

Answer. Harnessed appropriately, we believe AI can deliver great benefits for economies and society, and support decision-making which is fairer, safer, and more inclusive and informed. But such promises will not be realized without great care and effort, and Congress has an important role to play in considering how the development and use of AI should be governed. In our paper *Perspectives on Issues in AI Governance,* we outline five areas where government, in collaboration with civil society and AI practitioners, can play a crucial role: explainability standards, approaches to appraising fairness, safety considerations, requirements for human-AI collaboration, and general liability frameworks.

---

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO DR. STEPHEN WOLFRAM

*A.I. Accountability and Civil Rights.* One tech company, Facebook, announced that it is conducting an audit to identify and address discrimination. It has also formed Social Science One, which provides external researchers with data to study the platform's effects of social media on democracy and elections.

*Question 1.* What specific datasets and information would you need to scrutinize Facebook and Google's systems on civil rights and disinformation?

Answer. It's difficult to say. First, one would need a clear, computable definition of "civil rights and disinformation". Then one could consider a black-box investigation, based on a very large number (? billions) of inputs and outputs. This would be a difficult project. One could also consider a white-box investigation, involving looking at the codebase, at machine-learning training examples, etc. But, as I explained in my written testimony, under most circumstances, I would expect it to be essentially impossible to derive solid conclusions from this.

*Loot Boxes.* One of the most prolific manipulative practices in the digital economy is "loot boxes." Loot boxes are, in effect, gambling—selling gamers randomly-selected virtual prizes. The games do everything they can to coax people to taking chances on loot boxes. There is increasing scientific evidence that loot boxes share the same addictive qualities as gambling.

*Question 2.* Do you agree with me that loot boxes in video games share the same addictive qualities as gambling, particularly when targeting children?

Answer. This is outside my current areas of expertise.

*Question 3.* Would you support legislation like the Protecting Children from Abusive Games Act, which would prohibit the sale of loot boxes in games catering to children?

Answer. This is outside my current areas of expertise.

*Data Privacy and Manipulative Technologies.* Google and Facebook have an intimate understanding of the private lives of their users. They know about our family relationships, our financial affairs, and our health. This rich profile of our lives is intensively mined to exploit our attention and target us with ever-more manipulative advertising. However, while persuasive technologies take advantage of information about users, their users know little about them.

*Question 4.* Would Google and Facebook, if they wanted to, be able to specifically single out and target people when they are emotionally vulnerable or in desperate situations based on the data they collect?

Answer. I would think so.

*Question 5.* Currently, would it be against the law to do so—for example, were Facebook to target teenagers that it predicts feel like "a failure" with ads?

Answer. I can't offer an informed opinion.

*Question 6.* How can we ensure data privacy laws prevent the use of personal data to manipulate people based on their emotional state and vulnerabilities?

Answer. I am extremely skeptical that this will be possible to achieve through data privacy laws alone, without severely reducing the utility of automatic content selection services, at least until there have been substantial advances with computational contracts, which are still a significant time in the future. I favor a market-based approach, as I discussed in my written testimony. I think this could be implemented now.

*Recommendations from Ms. Richardson.* Ms. Richardson provided a set of recommendations in her remark for Congress to act, including:

1.) Require Technology Companies to Waive Trade Secrecy and Other Legal Claims That Hinder Oversight and Accountability Mechanisms

2.) Require Public Disclosure of Technologies That Are Involved in Any Decisions About Consumers by Name and Vendor

3.) Empower Consumer Protection Agencies to Apply "Truth in Advertising Laws" to Algorithmic Technology Providers

4.) Revitalize the Congressional Office of Technology Assessment to Perform Pre-Market Review and Post-Market Monitoring of Technologies

5.) Enhanced Whistleblower Protections for Technology Company Employees That Identify Unethical or Unlawful Uses of AI or Algorithms

6.) Require Any Transparency or Accountability Mechanism To Include A Detailed Account and Reporting of The "Full Stack Supply Chain"

7.) Require Companies to Perform and Publish Algorithmic Impact Assessments Prior to Public Use of Products and Services

During the hearing, I requested for you to respond in writing if possible.

*Question 7.* Please provide feedback to Ms. Richardson's suggestions for Congressional action.

Answer.

1. If this advocates removing all trade secret protection for technology then it is certainly overly broad.
2. "Involved in decisions about consumers" seems overly broad. Would this include underlying software infrastructure or basic data sources, or only something more specific? Also, "technologies" don't always have names, particularly when they are newly invented.
3. This is outside my current areas of expertise.
4. This is outside my current areas of expertise.
5. This is outside my current areas of expertise.
6. How far down would this go? Software only? Hardware? Networking? For aggregated data sources (*e.g.,* Census Bureau) the details of underlying data are protected by privacy requirements.
7. Without more details of what's proposed, I can't really offer useful input. I would note the phenomenon of computational irreducibility (discussed in my written testimony) which provides fundamental limits on the ability to foresee the consequences of all but unreasonably limited computational processes.

*Question 8.* What other steps or actions should Congress consider in regulating the use or consumer protection regarding persuasive technologies or artificial intelligence?

Answer. I made specific suggestions in my written testimony, and I am encouraged by feedback since the hearing that these suggestions are both practical and valuable.

---

RESPONSE TO WRITTEN QUESTIONS SUBMITTED BY HON. RICHARD BLUMENTHAL TO RASHIDA RICHARDSON

*A.I. Accountability and Civil Rights.* One tech company, Facebook, announced that it is conducting an audit to identify and address discrimination. It has also formed Social Science One, which provides external researchers with data to study the platform's effects of social media on democracy and elections.

*Question 1.* What specific datasets and information would you need to scrutinize Facebook and Google's systems on civil rights and disinformation?

Answer. There are a variety of civil rights implications regarding the variety of AI applications so the specific datasets and information needed to assess civil rights liability depends on the specifics law and application. In my testimony, I referenced legal challenges as well as subsequent research regarding Facebook's ad-targeting and delivery system. The cases referenced violations of several civil rights statutes but for brevity, I will focus on Title VII. Title VII prohibits discrimination in the employment context including in advertising. The previously referenced lawsuits claimed that Facebook's ad-targeting enable employers and employment agencies to discrimination based on sex, age, and race, but subsequent research of Facebook's ad-delivery mechanisms found that there were also aspects of the ad-delivery system design (which Facebook has exclusive control over) also contributed to biased outcomes. To assess how Facebook contributes to discriminatory outcomes in ad-targeting, one would want copies of the forms advertisers use to select advertising targets (i.e.—criteria as well as overall design aspects that may influence advertisers choice), information about advertisers targets, and demographic data on ad delivery. To assess how Facebook contributes to discriminatory outcomes in ad-delivery, one would want information about the selection choices of advertisers, actual delivery outcomes, demographic characteristics about the populations that was selected by advertiser and who received the ad, aggregated data about the content of similar ads and delivery outcomes, and information on behaviors of users that received ads. The aforementioned list of information or datasets is not exhaustive but illustrative of the types of information to request to perform a comparative assessment about impact and who is liable. Title VII provides for the use of statistical evidence to show disparate impact

It is also worth noting that based on my current research on the civil rights implications of AI and other emerging technologies, there are several deficiencies with existing Civil Rights and anti-discrimination statutes and caselaw (judicial interpretations) in adequately addressing the range of concerns and consequences that arise from our current understandings of AI use, particularly on Internet platforms. For instance, relying on the Title VII jurisprudence, there is uncertainty about whether an employer may be liable for changing how it uses Facebook's ad targeting and delivery platform or other practices, once made aware of discriminatory outcomes. This is an example of the type of problems inherent in current laws and caselaw.

*Loot Boxes.* One of the most prolific manipulative practices in the digital economy is "loot boxes." Loot boxes are, in effect, gambling—selling gamers randomly-selected virtual prizes. The games do everything they can to coax people to taking chances on loot boxes. There is increasing scientific evidence that loot boxes share the same addictive qualities as gambling.

*Question 2.* Do you agree with me that loot boxes in video games share the same addictive qualities as gambling, particularly when targeting children?

Answer. I do not have expertise in this area to comment.

*Question 3.* Would you support legislation like the Protecting Children from Abusive Games Act, which would prohibit the sale of loot boxes in games catering to children?

Answer. I support legislation that will create more protections for children engaging with persuasive technologies.

*Data Privacy and Manipulative Technologies.* Google and Facebook have an intimate understanding of the private lives of their users. They know about our family relationships, our financial affairs, and our health. This rich profile of our lives is intensively mined to exploit our attention and target us with ever-more manipulative advertising. However, while persuasive technologies take advantage of information about users, their users know little about them.

*Question 4.* Would Google and Facebook, if they wanted to, be able to specifically single out and target people when they are emotionally vulnerable or in desperate situations based on the data they collect?

Answer. Facebook actively monitors depressive states of users and there is evidence to suggest that other companies have capabilities to target advertising

*Question 5.* Currently, would it be against the law to do so—for example, were Facebook to target teenagers that it predicts feel like "a failure" with ads?

Answer. Facebook allows advertisers to target micropopulations which can include certain subjective characterizations of users.

*Question 6.* How can we ensure data privacy laws prevent the use of personal data to manipulate people based on their emotional state and vulnerabilities?

Answer. Data privacy laws need to be enhance to ensure better protections of people using persuasive technologies but there also needs to be government investment in public education to help inform people about big data practices so they can make more informed choices as well as actual investment in public institutions (*e.g.,* schools and libraries) to both improve access to technology but also digital literacy.

*Recommendations on Congressional Action.* Thank you for your recommendations on steps Congress can take.

*Question 7.* What other steps or actions should Congress consider in regulating the use or consumer protection regarding persuasive technologies or artificial intelligence?

Answer. The recommendations in my written comments are the best recommendations I have to date.

○