AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

S. HRG. 118–154

# OVERSIGHT OF A.I.: LEGISLATING ON ARTIFICIAL INTELLIGENCE

## HEARING

BEFORE THE

SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW

OF THE

COMMITTEE ON THE JUDICIARY

UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

FIRST SESSION

SEPTEMBER 12, 2023

**Serial No. J–118–30**

Printed for the use of the Committee on the Judiciary

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

U.S. GOVERNMENT PUBLISHING OFFICE
53–879 WASHINGTON : 2025

COMMITTEE ON THE JUDICIARY

RICHARD J. DURBIN, Illinois, *Chair*

DIANNE FEINSTEIN, California
SHELDON WHITEHOUSE, Rhode Island
AMY KLOBUCHAR, Minnesota
CHRISTOPHER A. COONS, Delaware
RICHARD BLUMENTHAL, Connecticut
MAZIE K. HIRONO, Hawaii
CORY A. BOOKER, New Jersey
ALEX PADILLA, California
JON OSSOFF, Georgia
PETER WELCH, Vermont

LINDSEY O. GRAHAM, South Carolina, *Ranking Member*
CHARLES E. GRASSLEY, Iowa
JOHN CORNYN, Texas
MICHAEL S. LEE, Utah
TED CRUZ, Texas
JOSH HAWLEY, Missouri
TOM COTTON, Arkansas
JOHN KENNEDY, Louisiana
THOM TILLIS, North Carolina
MARSHA BLACKBURN, Tennessee

JOSEPH ZOGBY, *Chief Counsel and Staff Director*
KATHERINE NIKAS, *Republican Chief Counsel and Staff Director*

SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW

RICHARD BLUMENTHAL, Connecticut, *Chair*

AMY KLOBUCHAR, Minnesota
CHRISTOPHER A. COONS, Delaware
MAZIE K. HIRONO, Hawaii
ALEX PADILLA, California
JON OSSOFF, Georgia

JOSH HAWLEY, Missouri, *Ranking Member*
JOHN KENNEDY, Louisiana
MARSHA BLACKBURN, Tennessee
MICHAEL S. LEE, Utah
JOHN CORNYN, Texas

DAVID STOOPLER, *Democratic Chief Counsel*
JOHN EHRETT, *Republican Chief Counsel*

# CONTENTS

## OPENING STATEMENTS


| | Page |
|---|---|
| Blumenthal, Hon. Richard | 1 |
| Hawley, Hon. Josh | 3 |


## WITNESSES


| | |
|---|---|
| Dally, William | 4 |
| Prepared statement | 46 |
| Responses to written questions | 54 |
| Hartzog, Woodrow | 7 |
| Prepared statement | 55 |
| Responses to written questions | 70 |
| Smith, Brad | 6 |
| Prepared statement | 105 |
| Responses to written questions | 115 |


## APPENDIX


| | |
|---|---|
| Items submitted for the record | 45 |

# OVERSIGHT OF A.I.: LEGISLATING ON ARTIFICIAL INTELLIGENCE

**TUESDAY, SEPTEMBER 12, 2023**

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

The Subcommittee met, pursuant to notice at 2:35 p.m., in Room 226, Dirksen Senate Office Building, Hon. Richard Blumenthal, Chair of the Subcommittee, presiding.

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

## OPENING STATEMENT OF HON. RICHARD BLUMENTHAL, A U.S. SENATOR FROM THE STATE OF CONNECTICUT

Chair BLUMENTHAL. The hearing of our Subcommittee on Privacy, Technology, and the Law will come to order. I want to welcome our witnesses, all of the audience who are here, and say a particular thanks to Senator Schumer, who has been very supportive and interested in what we're doing here. And also to Chairman Durbin whose support has been invaluable in encouraging us to go forward here.

I have been grateful, especially to my partner in this effort, Senator Hawley, the Ranking Member. He and I, as you know, have produced a framework, basically a blueprint for a path forward to achieve legislation. Our interest is in legislation, and this hearing along with the two previous ones, have to be seen as a means to that end.

[Poster is displayed.]

Chair BLUMENTHAL. We're very result oriented, as I know you are, from your testimony. And I've been enormously encouraged and emboldened by the response so far just in the past few days, and from my conversations with leaders in the industry, like Mr. Smith. There is a deep appetite, indeed a hunger for rules and guardrails, basic safeguards for businesses and consumers, for people in general, from the panoply of potential perils. But there's also a desire to make use of the tremendous potential benefits.

And our effort is to provide for regulation in the best sense of the word. Regulation that permits and encourages innovation, and new businesses, and technology and entrepreneurship. But at the same time provides those guardrails, enforceable safeguards, that can encourage trust and confidence in this growing technology. It's not a

new technology entirely. It's been around for decades, but artificial intelligence is regarded as entering a new era.

And make no mistake, there will be regulation. The only question is, how soon and what? And it should be regulation that encourages the best in American free enterprise, but at the same time provides the kind of protections that we do in other areas of our economic activity.

To my colleagues who say, "There's no need for new rules, we have enough laws protecting the public," yes, we have laws that prohibit unfair and deceptive competition. We have laws that regulate airline safety and drug safety. But nobody would argue that simply because we have those rules, we don't need specific protections for medical device safety or car safety.

Just because we have rules that prohibit discrimination in the workplace doesn't mean that we don't need rules that prohibit discrimination in voting. And we need to make sure that these protections are framed and targeted in a way that apply to the risks involved. Risk-based rules. Managing the risks is what we need to do here.

So our principles are pretty straightforward, I think. We have no pride of authorship. We have circulated this framework to encourage comment. We won't be offended by criticism from any quarter. That's the way we can make this framework better and eventually achieve legislation we hope—I hope, at least, by the end of this year.

And the framework is, basically: establishing a licensing regime for companies that are engaged in high-risk AI development; creating an independent oversight body that has expertise with AI and works with other agencies to administer and enforce the law; protecting our national and economic security to make sure we aren't enabling China, or Russia, and other adversaries to interfere in our democracy or violate human rights; requiring transparency about the limits and use of AI models.

And at this point, includes rules like watermarking, digital disclosure when AI is being used, and data access for researchers, and ensuring that AI companies can be held liable when their products breach privacy, violate civil rights, endanger the public, deepfakes, impersonation, hallucination. We've all heard those terms. We need to prevent those harms.

And Senator Hawley and I, as former attorneys general of our States, have a deep and abiding affection for the potential enforcement powers of those officials—State officials. But the point is, there must be effective enforcement. Private rights of action as well as Federal enforcement are very, very important.

So let me just close by saying—before I turn it over to my colleague—we're going to have more hearings. The way to build a coalition in support of these measures is to disseminate as widely as possible the information that's needed for our colleagues to understand what's at stake here.

We need to listen to the kinds of industry leaders and experts that we have before us today, and we need to act with dispatch more than just deliberate speed. We need to learn from our experience with social media that if we let this horse get out of the barn, it will be even more difficult to contain than social media. And we

are seeking to act on social media, the harms that it portends right now as we speak.

We're literally at the cusp of a new era. I asked Sam Altman, when he sat where you are, what his greatest fear was. I said my nightmare is the massive unemployment that could be created. That is an issue that we don't deal with directly here, but it shows how wide the ramifications may be.

And we do need to deal with potential worker displacement and training. And this new era is one that portends enormous promise, but also perils. We need to deal with both. I'll turn now to Ranking Member, Senator Hawley.

**OPENING STATEMENT OF HON. JOSH HAWLEY, A U.S. SENATOR FROM THE STATE OF MISSOURI**

Senator HAWLEY. Thank you, Mr. Chairman. Thank you for organizing this hearing. This is now, as the Chairman said, the third of these hearings that we've done. I've learned a lot in the previous couple. I think, you know, some of what we're learning about the potentials of AI is exhilarating. Some of it is horrifying.

And I think what I hear the Chairman saying, and what I certainly agree with, is we have a responsibility here now to do our part to make sure that this new technology, which holds a lot of promise but also peril, actually works for the American people.

That it's good for working people. That it's good for families. That we don't make the same mistakes that Congress made with social media where, 30 years ago now, Congress basically outsourced social media to the biggest corporations in the world. And that has been, I would submit to you, nearly an unmitigated disaster.

We've had the biggest, most powerful corporations, not just in America, but on the globe, and in the history of the globe, doing whatever they want with social media. Running experiments basically every day on America's kids, inflicting mental health harms the likes of which we've never seen, messing around in our elections in a way that is deeply, deeply corrosive to our way of life. We cannot make those mistakes again.

So we are here, as Senator Blumenthal said, to try to find answers, and to try to make sure that this technology is something that actually benefits the people of this country. I have no doubt, with all due respect to the corporatists who are in front of us, those heads of these corporations, I have no doubt it's going to benefit your companies. What I want to make sure is, is that actually benefits American people. And I think that's the test that we're engaged in. I look forward to this today. Thank you, Mr. Chairman.

Chair BLUMENTHAL. Thank you. I want to introduce our witnesses. And then, as is our custom, I will swear them in and ask them to submit their testimony. Welcome, to all of you.

William Dally is NVIDIA's chief scientist. He joined NVIDIA in January, 2009, as chief scientist after spending 12 years at Stanford University where he was chairman of the computer science department. He has published over 250 papers. He holds 120 issued patents, and he's the author of four textbooks.

Brad Smith is vice chair and president of Microsoft. As Microsoft's vice chair and president, he is responsible for spearheading the company's work and representing it publicly in a wide

variety of critical issues involving the intersection of technology and society, including artificial intelligence, cybersecurity, privacy, environmental sustainability, human rights, digital safety, immigration, philanthropy, and products and business for non-profit customers. And we appreciate your being here.

Professor Woodrow Hartzog is professor of law, and Class of 1960 Scholar at Boston University School of Law. He's also a non-resident fellow at the Cordell Institute for Policy in Medicine & Law at Washington University, a faculty associate at the Berkman Klein Center for Internet & Society at Harvard University, and an affiliate scholar at the Center for Internet and Society at Stanford Law School.

I could go on about each of you at much greater length with all your credentials, but suffice it to say, very impressive. And if you'll now stand, I'll administer the oath.

[Witnesses are sworn in.]

Chair BLUMENTHAL. Thank you. Why don't we begin with you, Mr. Dally?

**STATEMENT OF WILLIAM DALLY, CHIEF SCIENTIST AND SENIOR VICE PRESIDENT OF RESEARCH, NVIDIA CORPORATION, SANTA CLARA, CALIFORNIA**

Mr. DALLY. Chairman Blumenthal, Ranking Member Hawley, esteemed Judiciary Committee Members, thank you for the privilege to testify today. I'm NVIDIA's chief scientist and head of research, and I'm delighted to discuss our artificial intelligence journey and future.

NVIDIA's at the forefront of accelerated computing and generative AI, technologies that have the potential to transform industries, address global challenges, and profoundly benefit society. Since our founding in 1993, we have been committed to developing technology to empower people and improve the quality of life worldwide.

Today, over 40,000 companies use NVIDIA platforms across media and entertainment, scientific computing, healthcare, financial services, internet services, automotive, and manufacturing to solve the world's most difficult challenges, and bring new products and services to consumers worldwide.

At our founding in 1993, we were a 3D graphic startup. One of dozens of startups competing to create an entirely new market for accelerators to enhance computer graphics for games. In 1999, we invented the graphics processing unit, or GPU, which can perform a massive number of calculations in parallel.

When we launched the GPU for AI in gaming, we recognized that GPUs could theoretically accelerate any application that could benefit from massively parallel processing. And this bet paid off.

Today, researchers worldwide innovate on NVIDIA GPUs. Through our collective efforts, we have made advances in AI that will revolutionize and provide tremendous benefits to society across sectors such as healthcare, medical research, education, business, cybersecurity, climate, and beyond.

However, we also recognize that like any new product or service, AI products and services have risks, and those who make, and use,

or sell AI-enabled products and services are responsible for their conduct.

Fortunately, many uses of AI applications are subject to existing laws and regulations that govern the sectors in which they operate. AI-enabled services in high-risk sectors could be subject to enhanced licensing and certification requirements when necessary while other applications with less risk of harm may need less stringent licensing and/or regulation. With clear, stable, and thoughtful regulation, AI developers will work to benefit society while making products and services as safe as possible.

For our part, NVIDIA is committed to the safe and trustworthy development and deployment of AI. For example, NeMo Guardrails, our open-source software, empowers developers to guide generative AI applications to produce accurate, appropriate, and secure text responses. NVIDIA has implemented model risk management guidance, ensuring a comprehensive assessment and management of risks associated with NVIDIA-developed models.

Today, NVIDIA announced it is endorsing the White House's voluntary commitments on AI. As we deploy AI more broadly, we can and will continue to identify and address risks.

No discussion of AI would be complete without addressing what is often described as frontier AI models. Some have expressed fear that frontier models will evolve into uncontrollable artificial general intelligence which could escape our control and cause harm.

Fortunately, uncontrollable artificial general intelligence is science fiction, not reality. At its core, AI is a software program that is limited by its training, the inputs provided to it, and the nature of its output. In other words, humans will always decide how much decision-making power to cede to AI models.

So long as we are thoughtful and measured, we can ensure safe, trustworthy, and ethical deployment of AI systems without suppressing innovation. We can spur innovation by ensuring that AI tools are widely available to everyone, not concentrated in the hands of a few powerful firms.

I will close with two observations.

First, the AI genie is already out of the bottle. AI algorithms are widely published and available to all. AI software can be transmitted anywhere in the world at the press of a button. And many AI development tools, frameworks, and foundational models are open-sourced.

Second, no nation, and certainly no company, controls a chokepoint to AI development. Leading U.S. computing platforms are competing with companies from around the world. While U.S. companies may currently be the most energy efficient, cost efficient, and easiest to use, they're not the only viable alternatives for developers abroad.

Other nations are developing AI systems, with or without U.S. components, and they will offer those applications in the worldwide market. Safe and trustworthy AI will require multilateral and multi-stakeholder cooperation, or it will not be effective.

The United States is in a remarkable position today, and with your help, we can continue to lead on policy and innovation well into the future.

NVIDIA stands ready to work with you to ensure that the development and deployment of generative AI and accelerated computing serve the best interest of all. Thank you for your opportunity to testify before this Committee.

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

Chair BLUMENTHAL. Thank you, very much. Mr. Smith.

**STATEMENT OF BRAD SMITH, VICE CHAIR AND PRESIDENT, MICROSOFT CORPORATION, REDMOND, WASHINGTON**

Mr. SMITH. Chairman Blumenthal, Ranking Member Hawley, Members of the Subcommittee, my name is Brad Smith. I'm the vice chair and president of Microsoft. And thank you for the opportunity to be here today. And I think more importantly, thank you for the work that you have done to create the framework you've shared.

Chairman Blumenthal, I think you put it very well.

First, we need to learn and act with dispatch. And Ranking Member Hawley, I think you offered real words of wisdom. Let's learn from the experience the whole world had with social media, and let's be clear-eyed about the promise and the peril in equal measure as we look to the future of AI.

I would first say, I think, your framework does that. It doesn't attempt to answer every question by design. But it's a very strong and positive step in the right direction, and puts the U.S. Government on the path to be a global leader in ensuring a balanced approach that will enable innovation to go forward with the right legal guardrails in place.

As we all think about this more, I think it's worth keeping three goals in mind.

First, let's prioritize safety and security, which your framework does. Let's require licenses for advanced AI models and uses in high-risk scenarios. Let's have an agency that is independent, and can exercise real and effective oversight over this category. And then, let's couple that with the right kinds of controls that will ensure safety of the sort that we've already seen, I think, start to emerge in the White House commitments that were launched on July 21st.

Second, let's prioritize, as you do, the protection of our citizens and consumers. Let's prioritize national security—always in a sense, in some ways, the first priority of the Federal Government. But let's think as well as you have about protecting the privacy, the civil rights, and the needs of kids among many other ways of working and ensure we get this right.

Let's take the approach that you are recommending. Namely, focus not only on those companies that develop AI, like Microsoft, as well as companies that deploy AI, like Microsoft. In different categories, we're going to need different levels of obligations.

And as we go forward, let's think about the connection between, say, the role of a central agency that will be on point for certain things, as well as the obligations that, frankly, will be part of the work of many agencies and, indeed, our courts as well.

And let's do one other thing, as well—maybe it's one of the most important things we need to do so that we ensure that the threats

that many people worry about remain part of science fiction and don't become a new reality. Let's keep AI under the control of people. It needs to be safe. And to do that, as we've encouraged, there need to be safety breaks, especially for any AI application or system that controls critical infrastructure.

If a company wants to use AI to, say, control the electrical grid, or all of the self-driving cars on our roads, or the water supply, we need to learn from so many other technologies that do great things but also can go wrong. We need a safety break. Just like we have a circuit breaker in every building and home in this country to stop the flow of electricity if that's needed.

Then I would say let's keep one third goal in mind, as well. This is the one where I would suggest you maybe consider doing a bit more to add to the framework. Let's remember the promise that this offers. Right now if you go to State capitals, you go to other countries, I think there's a lot of energy being put on that.

When I see what Governor Newsom is doing in California, or Governor Burgum in North Dakota, or Governor Youngkin in Virginia, I see them at the forefront of figuring out how to use AI to, say, improve the delivery of healthcare, advance medicine, improve education for our kids, and, maybe most importantly, make government services more accessible and more efficient.

Let's see if we can find a way to not only make government better by using this technology, but cheaper—or use the savings to provide more and better services to our people. That would be a good problem to have the opportunity to consider.

In sum, Professor Hartzog has said this is not a time for half measures. It is not. He is right. Let's go forward as you have recommended. Let's be ambitious and get this right. Thank you.

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

Chair BLUMENTHAL. Thank you. Thank you, very much. And Mr. Hartzog, I've read your testimony, and you are very much against half measures. So we look forward to hearing what the full measures that you recommend are.

Professor HARTZOG. That's correct, Senator.

**STATEMENT OF WOODROW HARTZOG, PROFESSOR OF LAW, BOSTON UNIVERSITY SCHOOL OF LAW, AND FELLOW, CORDELL INSTITUTE FOR POLICY IN MEDICINE & LAW, WASHINGTON UNIVERSITY IN ST. LOUIS, BOSTON, MASSACHUSETTS**

Professor HARTZOG. Chair Blumenthal, Ranking Member Hawley, and Members of the Committee, thank you for inviting me to appear before you today.

My name is Woodrow Hartzog, and I'm a professor of law at Boston University. My comments today are based on a decade of researching law and technology issues, and I'm drawing from research on artificial intelligence policy that I conducted as a fellow with colleagues at the Cordell Institute at Washington University in St. Louis.

Committee Members, up to this point, AI policy has largely been made up of industry-led approaches, like encouraging transparency, mitigating bias, and promoting principles of ethics. I'd like

to make one simple point in my testimony today. These approaches are vital, but they are only half measures. They will not fully protect us.

To bring AI within the rule of law, lawmakers must go beyond these half measures to ensure that AI systems and the actors that deploy them are worthy of our trust. Half measures like audits, assessments, and certifications are necessary for data governance, but industry leverages procedural checks like these to dilute our laws and to managerial box-checking exercises that entrench harmful surveillance-based business models.

A checklist is no match for the staggering fortune available to those who exploit our data, our labor, and our precarity to develop and deploy AI systems. And it's no substitute for meaningful liability when AI systems harm the public.

Today I'd like to focus on three popular half measures, and why lawmakers must do more.

First, transparency is a popular proposed solution for opaque systems, but it does not produce accountability on its own. Even if we truly understand the various parts of AI systems, lawmakers must intervene when these tools are harmful and abusive.

A second laudable, but insufficient approach is when companies work to mitigate bias. AI systems are notoriously biased along lines of race, class, gender, and ability. While mitigating bias in AI systems is critical, self-regulatory efforts to make AI fair are half measures doomed to fail. It's easy to say that AI systems should not be biased. It's very difficult to find consensus on what that means and how to get there.

Additionally, it's a mistake to assume that if a system is fair, then it's safe for all people. Even if we ensure that AI systems work equally well for all communities, all we will have done is create a more effective tool that the powerful can use to dominate, manipulate, and discriminate.

A third AI half measure is committing to ethical principles. Ethics are important, and these principles sound impressive, but they are a poor substitute for laws. It's easy to commit to ethics, but industry doesn't have the incentive to leave money on the table for the good of society.

I have three recommendations for the Committee to move beyond AI half measures.

First, lawmakers must accept that AI systems are not neutral, and regulate how they are designed. People often argue that lawmakers should avoid design rules for technologies because there are no bad AI systems, only bad AI users.

This view of technologies is wrong. There is no such thing as a neutral technology, including AI systems. Facial recognition technologies empower the watcher. Generative AI systems replace labor. Lawmakers should embrace established theories of accountability like product liability's theory of defective design, or consumer protection's theory of providing the means and instrumentalities of unfair and deceptive conduct.

My second recommendation is to focus on substantive laws that limit abuses of power. AI systems are so complex and powerful that it can seem like trying to regulate magic, but the broader risks and

benefits of AI systems are not so new. AI systems bestow power. This power is used to benefit some and harm others.

Lawmakers should borrow from established legal approaches to remedying power imbalances, to require broad non-negotiable duties of loyalty, care, and confidentiality, and implement robust bright-line rules that limit harmful secondary uses and disclosures of personal data in AI systems.

My final recommendation is to encourage lawmakers to resist the idea that AI is inevitable. When lawmakers go straight to putting up guardrails, they fail to ask questions about whether particular AI systems should exist at all. This dooms us to half measures. Strong rules would include prohibitions on unacceptable AI practices like emotion recognition, biometric surveillance in public spaces, predictive policing, and social scoring.

In conclusion, to avoid the mistakes of the past, lawmakers must make the hard calls. Trust and accountability can only exist where the law provides meaningful protections for humans, and AI half measures will certainly not be enough. Thank you, and I welcome your questions.

[The prepared statement of Professor Hartzog appears as a submission for the record.]

Chair BLUMENTHAL. Thank you, Professor Hartzog. And I take very much to heart your imploring us against half measures. I think listening to both Senator Hawley and myself, you have a sense of boldness and initiative.

And we welcome all of the specific ideas, most especially Mr. Smith, your suggestion that we can be more engaged and proactive at the State level or Federal level in making use of AI in the public sector.

But taking the thought that Professor Hartzog has so importantly introduced, AI technology in general is not neutral. How do we safeguard against the downsides of AI, whether it's discrimination or surveillance? Will this licensing regime and oversight entity be sufficient? And what kind of powers do we need to give it?

Mr. SMITH. Well, I would say, first of all, I think that a licensing regime is indispensable in certain high-risk scenarios, but it won't be sufficient to address every issue.

But it's a critical start. Because I think what it really ensures is especially, say, for the frontier models, the most advanced, as well as certain applications that are highest risk, frankly, you do need a license from the Government before you go forward.

And that is real accountability. You can't drive a car until you get a license. You can't make the model or the application available until you pass through that gate.

I do think that it would be a mistake to think that one single agency or one single licensing regime would be the right recipe to address everything, especially when we think about the harms that we need to address.

And that's why I think it's equally critical that every agency in the Government that is responsible for the enforcement of the law and the protection of people's rights, master the capability to assess AI. I don't think we want to move the approval of every new drug from the FDA to this agency. So by definition, the FDA is going to need, for example, to have the capability to assess AI. That would

be just one of several additional specifics that I think one can think about.

Chair BLUMENTHAL. I think that's a really important point because AI is going to be used in making automobiles, making airplanes, making toys for kids. So the FAA, the FDA, the Federal Trade Commission, the Consumer Product Safety Commission, they all have presently existing rules and regulations, but there needs to be an oversight entity that uses some of those rules and adapts them, and adopts new rules so that those harms can be prevented.

And there are a lot of different names we could call that entity. Connecticut now has an Office of Artificial Intelligence. You could use different terms, but I think the idea is that we want to make sure that the harms are prevented through a licensing regime focused on risk.

Mr. Dally, you know, you said that autonomous AI is science fiction. AI beyond human control is science fiction. But science fiction has a way of coming true, and I wonder whether that is a potential fear. Certainly, it is one that's widely shared at the moment.

Whether it's fact-based or not, it is in the reality of human perception. And, as you well know, trust and confidence are very, very important. So I wonder how we counter the perception and prevent the science fiction from becoming reality?

Mr. DALLY. So what I said is that artificial general intelligence that gets out of control as science fiction, not autonomous. We use artificial intelligence, for example, autonomous vehicles all the time.

I think the way we make sure that we have control over AI of all sorts is by, for any really critical application, keeping a human in the loop. You know, AI is a computer program. It takes an input, it produces an output. And if you don't connect up something that can cause harm to that output, it can't cause that harm. And so anytime that there is some grievous harm that could happen, you want a human being between the output of that AI model and the causing of harm.

And so I think as long as we're, you know, careful about how we deploy AI to, you know, keep humans in the critical loops, I think we can assure that the AIs, you know, won't take over and shut down our power grid, or, you know, cause airplanes to fall out of the sky. We can keep control over them.

Chair BLUMENTHAL. Thank you. I have a lot more questions, but we're going to adhere to 5-minute rounds. We have a very busy day, as you know, with votes, as a matter of fact. And I'll turn to Senator Hawley.

Senator HAWLEY. Thank you, Mr. Chairman. Thanks, again, to the witnesses for being here. I want to particularly thank you, Mr. Smith. I know that there's a group of other—your colleagues, your counterparts in industry who are gathering, I think, tomorrow. And that is what it is.

But I appreciate you being willing to be here in public and answer questions in front of the presses here. And this is open to anybody who wants to see it. And I think that's the way that this ought to be done. I appreciate you being willing to do that.

You mentioned protecting kids. I just want to start with that, if I could. I want to ask you a little bit about what Microsoft has done and is doing. Kids use your Bing chatbot. Is that fair to say?

Mr. SMITH. Yes. We have certain age controls, so we don't let a child of any age. But yes, in general, it is possible for children to register if they're of a certain age.

Senator HAWLEY. And the age is?

Mr. SMITH. I'm trying to remember, as I sit here. I'll get you——

Senator HAWLEY. I think it's 13.

Mr. SMITH [continuing]. The answer.

Senator HAWLEY. Does that sound right?

Mr. SMITH. It would——

Senator HAWLEY. Maybe Senator——

Mr. SMITH. I was going to say 12 or 13.

Senator HAWLEY. Okay.

Mr. SMITH. I'll take it, at 13.

Senator HAWLEY. Do you have some sort of age verification? I mean, how do we know what age—I mean, obviously, the kid can put in whatever age he or she wants to. Is there some form of age verification for Bing?

Mr. SMITH. We do have age verification systems that then involve, typically, getting permission from a parent. And we use this across our services, including for gaming. I don't remember off the top of my head exactly how it works, but I'd be happy to get you the details.

Senator HAWLEY. Great. My impression is that Bing Chat doesn't really have an enforceable age verification. I mean, there's no way really to know, but again, you correct me if that's wrong.

Let me ask you this, what happens to all of the information that our hypothetical 13-year-old is putting into the tool as it's having this chat? You know, I mean, they could be chatting about anything and going back and forth on any number of subjects. What happens to that info that the kid puts in?

Mr. SMITH. Well, the most important thing I would say, first, is that it all is done in a manner that protects the privacy of children. This is——

Senator HAWLEY. And how is that?

Mr. SMITH. Well, we follow the rules in COPPA, which is, you know, exists to protect child online privacy. And it forbids using it for tracking. It forbids its use for advertising or for other things. It seeks to put very tight controls around the use and the retention of that information.

The second thing I would just add to that is in addition to protecting privacy, we are hyper-focused on ensuring that, in most cases, people of any age, but especially children, are not able to use something like Bing Chat in ways that would cause harm to themselves or to others.

Senator HAWLEY. And how do you do that?

Mr. SMITH. We basically have a safety architecture that we use across the board. Think about it like this, there's two things around a model.

The first is called a classifier so that if somebody asks, "How can I commit suicide tonight?" "How can I blow up my school tomor-

row?" that hits a classifier that identifies a class of questions, or prompts, or issues.

And then second, there's what we call Meta-Prompts and that we intervene so that the question is not answered. If someone asks how to commit suicide, we typically would provide a response that encourages someone to get mental health assistance and counseling, and tells them how. If somebody wants to know how to build a bomb, it says, no, you cannot use this to do that.

And that fundamental safety architecture is going to evolve, it's going to get better. But in a sense, it's at the heart, if you will, of both what we do and, I think, the best practices in the industry. And I think part of what this is all about, that we're talking about here, is how we take that architectural element and continue to strengthen it.

Senator HAWLEY. Very good. That's helpful. Let me ask you about the information—back to the kids' information for a second. Is it stored in the United States? Is it stored overseas?

Mr. SMITH. If the child is in the United States, the data is stored in the United States. It's true not only for children, it's for adults as well.

Senator HAWLEY. And who has access to that data?

Mr. SMITH. The child has access if—you know, the parents may or may not have access. Typically, we give——

Senator HAWLEY. In what circumstances would the parents have access?

Mr. SMITH. I would have to go get you the specifics on that. Our general principle is this—and this is something we've implemented in the United States even though it's not legally required in the United States. It is legally required, as you may know, in Europe—people, we think, have a right to find out what information we have about them. They have the right to see it. They have the right to ask us to correct it, if it's wrong. They have the right to ask us to delete it, if that's what they want us to do.

Senator HAWLEY. And you do that? If they ask you to delete it, you delete it?

Mr. SMITH. We'd better. Yes. That's our promise. And we do a lot to comply with that.

Senator HAWLEY. Let me just ask about—and I have a lot more questions—I'm going to try to adhere to the timeline, Mr. Chairman. Five minutes, Mr. Chairman.

Chair BLUMENTHAL. We'll have a second round.

Senator HAWLEY. All right. That's great news for us. Not such great news for the witnesses.

[Laughter.]

Senator HAWLEY. Sorry. Right.

Let me just—before I leave this subject, last thing, just about the kids' personal data and where it's stored. I'm asking you this, as I'm sure you can intuit, because we've seen other technology companies in the social media space who have major issues about where data is stored and major access issues.

And I'm thinking of it, it shouldn't be hard to guess, I'm thinking, in particular, of China, where we've seen other social media companies who say, "Oh, well, America's data is stored in America."

But guess what? Lots of people in other countries can access that data.

So is that true for you, Mr. Smith? Is a child's data that they've entered into the Bing Chat that's stored in the United States—you just said, if they're an American citizen—can that be accessed in, let's say, China, by a Microsoft China-based engineer?

Mr. SMITH. I don't believe so. I'd love to go back and just, you know, confirm that. But I don't believe——

Senator HAWLEY. Would you? Would you be able to get that for me for the record?

Mr. SMITH. Sure.

Senator HAWLEY. Thank you. Okay. I'll have lots of more questions later. Thanks, Mr. Chairman.

Chair BLUMENTHAL. Thanks, Senator Hawley. Senator Klobuchar.

Senator KLOBUCHAR. Thank you, very much. Thank you, all of you. I think I will lead with some elections questions, since I'm the Chair of the Rules Committee.

Mr. Smith, in your written testimony you talked about how watermarks could be helpful disclosure of AI material.

As you know, and we have talked about, I have a bill that I lead—that Representative Clark leads in the House—to require disclaimer and some kind of mark on AI-generated ads. I think we have to go further. We'll get to that in a minute, Professor Hartzog.

But could you talk about what you mean by in your written testimony, "the health of democracy and meaningful civic discourse will undoubtedly benefit from initiatives that help protect the public against deception or fraud facilitated by AI-generated content"?

Mr. SMITH. Absolutely. And here I do think things are moving quickly, both in a positive and a worrisome direction, in terms of what we're seeing. On the positive side, I think you're seeing the industry come together. You're seeing a company like Adobe, I think, exercise real leadership, and there's a recipe that I see emerging.

I think it starts with the first principle: People should have the right to know if they're getting a phone call from a computer, from AI, if there's content coming from an AI system rather than a human being.

We then need to make that real with legal rights that back it up. We need to create what's called a provenance system, watermarking for legitimate content so that it can't be altered easily without our detection to create a deepfake.

We need to create an effort that brings the industry and, I think, governments together so we know what to do and there's a consensus when we do spot deepfakes, especially, say, even deepfakes that have altered legitimate content.

Senator KLOBUCHAR. Thank you.

Mr. SMITH. So that would——

Senator KLOBUCHAR. And let's get to that——

Mr. SMITH [continuing]. Be the first one.

Senator KLOBUCHAR [continuing]. Hot off the press.

Senator Hawley and I have introduced our bill today with Senator Collins and—who led the Electoral Reform Act, as we know,

and Senator Coons—to ban the use of deceptive AI-generated content in elections.

So this would work in concert with some watermark system, but when you get into the deception where it is fraudulent—AI-generated content pretending to be the elected official or the candidate when it is not, and we've seen this used against people on both sides of the aisle—which is why it was so important that we be bipartisan in this work. And I want to thank him for his leadership on not only of the framework, but also on the work that we're doing.

And I guess I'll go to you, Senator Hartzog—Mr. Hartzog—I just promoted you.

[Laughter.]

Senator KLOBUCHAR. Maybe, I mean, it's very debatable.

[Laughter.]

Senator KLOBUCHAR. Could you—in your testimony, you advocate for some outright prohibitions, which we're talking about here. Now, we do have, of course, a constitutional exception for satire and humor because we love satire so much—the Senators do. Just kidding.

[Laughter.]

Senator KLOBUCHAR. Could you talk about why you believe there has to be some out-wide ban of misleading AI conduct related to Federal candidates and political ads? Talk about that.

Professor HARTZOG. Sure. Absolutely. Thank you for the question. I—of course, keeping in mind free expression constitutional protections that would apply to any sort of legislation, I do think that bright-line rules and prohibitions around such deceptive ads are critical.

Because we know that procedural walkthroughs, as I said in my testimony, are often—give the veneer of protection without actually protecting us. And so to outright prohibit these practices, I think, is really important.

And I would even go potentially a step further and think about ways in which we can prohibit not just those that we would consider to be deceptive, but practices that we'd consider even abusive that leverage our internal limitations and our desire to believe, or want to believe, things against us. And there's a body of law that sort of runs alongside unfair and deceptive trade practices around abusive trade practices.

Senator KLOBUCHAR. Okay. All right. Mr. Dally, thinking of that—and I've talked to Mr. Smith about this, as well, AI used as a scam—I had someone actually that I know well who has a kid in the Marines who's deployed somewhere where they don't even know where it is. Fake voice calls them, asks for money to be sent somewhere in Texas, I believe.

Could you talk about what companies do—and I appreciate the work you've done to ensure that AI platforms are designed so they can't be used for criminal purposes because that's got to be part of the work that we do.

Mr. DALLY. Yes——

Senator KLOBUCHAR. Because it's not just scams against elected officials.

Mr. DALLY. Yes. Well, I think the best measures against deepfakes, and Mr. Smith mentioned it in his testimony, is the use of provenance and authentication systems where you can have authentic images, authentic voice recordings signed by the device, whether it's a camera or an audio recorder, that has recorded that voice.

And then when it's presented, it can be authenticated as being genuine and not a deepfake. That's sort of the flip side of watermarks, which would require that anything that is synthetically generated be identified as such.

And those two technologies in combination can really help people sort out, along with a certain amount of public education, and make sure people understand, you know, what the technology is, you know, capable of and are on guard for that. But it can help them sort out what is real from what is fake.

Senator KLOBUCHAR. Okay. I'll ask Mr. Smith back where I started here. Some AI platforms use local news content without compensating journalists and papers, including by using their content to train AI algorithms.

The Journalism Competition and Preservation Act, a bill I have with Senator Kennedy, would allow local news organizations to negotiate with online platforms, including generative AI platforms, that use their content without compensation.

Could you talk about the impacts that AI could have on local journalism? You talked about, in your testimony, about the importance of investment and quality journalism. But where we're getting at, we've got to find a way that we make sure that the people who are actually doing the work are compensated in many ways, but also in journalism. Mr. Smith.

Mr. SMITH. I would just say three quick things.

Number one, look, we need to recognize that local journalism is fundamental to the health of the country and the electoral system. And it's ailing. So we need to find ways to preserve and promote it.

Number two, generally, I think we should let local journalists and publications, you know, make decisions about whether they want their content to be available for training, or grounding, and the like. And that's a big topic, and it's worthy of more discussion. And we should certainly let them, in my view, negotiate collectively, because that's the only way local journalism is really going to negotiate effectively.

Senator KLOBUCHAR. I appreciate your words. You want to add one thing—I'm going to get in trouble from Senator Blumenthal here. Go ahead.

Chair BLUMENTHAL. That's alright.

Mr. SMITH. No, but then I will just say, and there are ways that we can use AI to help local journalists. And we're interested in that, too. So let's add that to the list.

Senator KLOBUCHAR. Okay. Very good. And thank you, again, both of you. I talked about Senator Hawley's work, but you, Senator Blumenthal, for your leadership.

Chair BLUMENTHAL. Thank you, very much. Thank you for yours, Senator Klobuchar. Senator Hirono.

Senator HIRONO. Thank you, Mr. Chairman. Mr. Smith, it's good to see you, again. So every time we have one of these hearings, we learn something new. But the conclusion I've drawn is that AI is ubiquitous. Anybody can use AI. It can be used in any endeavor. So when I hear you folks testifying about how we shouldn't be taking half measures, I'm not sure what that means.

What does it mean not taking half measure on something as ubiquitous as AI, where there are other regulatory schemes that can touch upon those endeavors that use AI? So, you know, there's always a question I have of when we address something as complex as how AI is looking, that there are unintended consequences that we should care about. Would you agree? Anybody? Mr. Smith?

Mr. SMITH. I would absolutely agree. I think we have to define what's a full measure and what's a half measure. But I bet we can all agree that half measures are not good enough.

Senator HIRONO. Well, that is the thing. How to recognize, you know, going forward, what is actually going to help us with this powerful tool. So I have a question for you, Mr. Smith. It is a powerful tool that can be used for either good, or it can also be used to spread a lot of disinformation and misinformation.

And that happened during the disaster on Maui. And Maui residents were subject to disinformation, some of it coming from foreign governments, i.e., Russia, looking to sow confusion and distress, including, "Don't sign up for FEMA because they cannot be trusted." And I worry that with AI, such information will only become more rampant with future disasters.

Do you share my concern about misinformation in the disaster context in the role AI could play? And what can we do to prevent these foreign entities from pushing out AI disinformation to people who are very vulnerable?

Mr. SMITH. I absolutely share your concern, and I think there's two things we need to think about doing.

First, let's use the power of AI, as we are, to detect these kinds of activities when they're taking place because they can enable us to go faster—as they did in that instance where Microsoft, among others, used AI and other data technologies to identify what people were doing.

Number two, I just think we need to stand up as a country, and with other governments, and with the public, and sayc, there need to be some clear red lines in the world today regardless of how much else or what else we disagree about.

When you think about what happens, typically, in the wake of an earthquake, or a hurricane, or a tsunami——

Senator HIRONO. Mm-hmm.

Mr. SMITH [continuing]. Or a flood, the world comes together. People are generous, they help provide relief.

And then, let's look at what happened after the fire in Maui. It was the opposite of that. We had some people, not necessarily directed by the Kremlin, but people who regularly spread Russian propaganda, trying to discourage the people of Lahaina from going to the agencies that could help them. That's inexcusable.

And we saw what we believe is Chinese-directed activity trying to persuade the world, in multiple languages, that the fire was caused——

Senator HIRONO. Mm-hmm.

Mr. SMITH [continuing]. By the United States Government itself using a meteorological weapon.

Those are the things that we should all try to bring the international community together and agree they're off limits.

Senator HIRONO. Well, how do we identify that this is even occurring? That there is a, you know, a China- or Russia-directed misinformation going on? How do we—I didn't know this was happening, by the way.

And even in the Energy Committee, on which I sit, we had people testify and I asked, you know, regarding the Maui disaster. I asked one of the testifiers whether he was aware that there had been disinformation put out by a foreign government in that example. And he said yes.

But I don't know that the people of Maui recognized that that was going on.

So how do we, one, even identify that that's going on and then, two, to come forward and say this is happening and to name names—identify that, which country it is that it's spreading this kind of disinformation and misinformation?

Mr. SMITH. I think we have to think about two things.

First, I think we at a company like Microsoft have to lean in, and we are—with data, with infrastructure, with experts, and real-time capability to spot these threats, find the patterns, and reach well-founded conclusions.

And then the second thing, this is the harder thing. This is where it's going to need all of your help. What do we do if we find that a foreign government is deliberately trying to spread false information next year in a Senate or Presidential campaign about a candidate? How do we create the room so that information can be shared and people will consider it?

Senator HIRONO. Yes.

Mr. SMITH. You all, with this—the most important word in your framework is bipartisan. How do we create the bipartisan framework so that when we find this, we create a climate where people can listen? I think we have to look at both of those parts of the problem together.

Senator HIRONO. Well, I hope we can do that. And Mr. Chairman, if you don't mind, one of the concerns about AI from the worker standpoint is that their jobs will be gone.

And Professor Hartzog, you mentioned that the—that generative AI can result in job losses. And for both you and Mr. Smith, what are the kinds of jobs that will be lost to AI?

Professor HARTZOG. That's an excellent question. It's difficult to project that into the future. But I would start with saying that not necessarily something that can be automated effectively, but things that—I think that those that control the purse strings think could be automated effectively. And if it gets to the point where it appears as though it could, I imagine you'll see industry move in that direction.

Senator HIRONO. Mr. Smith, I think you mentioned in your book, which I am listening to, that things like ordering something out of a drive-through, that those jobs could be gone through AI.

Mr. SMITH. Yes. It was 4 years ago, we published our book—with my co-author, behind me—and we said, "What's the first job that we think might be eliminated by AI?" We don't have a crystal ball, but I bet it's taking an order in the drive-through of a fast food restaurant.

You're not really establishing a rapport with a human being. All the person does is listen and type into a computer what you're saying. So when AI can hear as well as a person, it can enter that in. And indeed, I was struck a few months ago, Wendy's, I think it was, announced that they were starting to consider whether they would automate with AI, the drive-through.

I think there's a lesson, though, in that. And it should give us both pause, but I think a little bit of optimism. There's no creativity involved in a drive-through, at least, listening and entering an order. There are so many jobs that do involve creativity.

So the real hope, I think, is to use AI to automate the routine, maybe even the work that's boring, to free people up so they can be more creative, so they can focus more on paying attention to other people and helping them. And if we just apply that recipe more broadly, I think we might put ourselves on a path that's more promising.

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

Chair BLUMENTHAL. Thank you, Senator Hirono. Senator Kennedy.

Senator KENNEDY. Thank you, Mr. Chairman. And thank you for calling this hearing. Mr. Dally—am I saying your name correctly, sir?

Mr. DALLY. That's correct.

Senator KENNEDY. Yes. Mr. Dally, if I am a recipient of content created by generative AI, do you think I should have a right to know that that content was generated by a robot?

Mr. DALLY. Yes, I think you do. I think the details would depend on the context, but in most cases, I think, you know——

Senator KENNEDY. Okay.

Mr. DALLY [continuing]. I or anybody else, if I received something, I'd like to know, is this real or was this generated?

Senator KENNEDY. Mr. Smith?

Mr. SMITH. Generally, yes. What I would say is, if you're listening to an audio, if you're watching a video, if you're seeing an image and it was generated by AI, I think people have a right to know.

The one area where I think there's a nuance is if you're using AI to, say, help you write something. Maybe it's helping you write the first draft. Just as I don't think any of us would say that when our staff helps us write something, we are obliged to give the speech and say, "Now I'm going to read the paragraph that my staff wrote." You make it your own.

And I think the written word is a little more complex. So we need to think that through. But as a broad principle, I agree with that principle.

Senator KENNEDY. Professor?

Professor HARTZOG. There are situations where you probably wouldn't expect to be dealing with the product of generative AI. And in those instances——

Senator KENNEDY. Well, that's the problem.

Professor HARTZOG. Right. But as times change, it's possible that our expectations change. And so——

Senator KENNEDY. But as a principle, do you think that people should have a right to know when they're being fed content from generative AI?

Professor HARTZOG. If they were—if they—well, it's—I tell my students it depends on the context. Generally speaking, if you're vulnerable to generative AI, then the answer is absolutely yes.

Senator KENNEDY. What do you mean, "if you're vulnerable"?

Professor HARTZOG. So there may be——

Senator KENNEDY. I'm just looking for——

Professor HARTZOG. Sure. Sure.

Senator KENNEDY [continuing]. No disrespect——

Professor HARTZOG. No. Not at all.

Senator KENNEDY [continuing]. A straight answer.

Professor HARTZOG. Absolutely.

Senator KENNEDY. I kind of like—I like two things. Breakfast food and straight answers.

Professor HARTZOG. I love them.

[Laughter.]

Senator KENNEDY. And if a robot is feeding me information, and I don't know it's a robot, am I entitled to know it's a robot as a consumer? Pretty straight up.

Professor HARTZOG. Yes. I think the answer's yes. In——

Senator KENNEDY. All right.

Professor HARTZOG [continuing]. A lot of context.

Senator KENNEDY. Let's start back from Mr. Dally. Am I entitled to know who owns that robot and where that content came from? I know it came from a robot, but somebody had to goose the robot to make it give me that content. Am I entitled as a consumer to know who owns the robot?

Mr. DALLY. I think that's a harder question that depends on the particular context. I think, you know, if somebody is feeding me a video and it's been identified as being generated by AI, I now know that it's generated, it's not real. You know, if it's, you know, being used, for example, in a political campaign, then——

Senator KENNEDY. But, but——

Mr. DALLY [continuing]. I would want to know who——

Senator KENNEDY [continuing]. But let me stop you.

Mr. DALLY [continuing]. Informed it. Yes.

Senator KENNEDY. Let's suppose I'm looking at a video and it was generated by a robot. Would it make any difference to you whether that robot was owned by, let's say, President Biden or President Trump? Don't you want to know in evaluating the content who owns the robot and who prompted it to give you this information?

Mr. DALLY. I would probably want to know that. I don't know that I would feel it would be required for me to know that.

Senator KENNEDY. How about you, Mr. Smith?

Mr. SMITH. I'm generally a believer in letting people know not only that it's generated by a computer, but who owns the——

Senator KENNEDY. Yes.

Mr. SMITH [continuing]. Program that's doing it. The only qualification I would offer, and it's something you all should think about and would know better than me, there are certain areas in political speech where one has to decide whether you want people to act with anonymity. The Federalist Papers were first published under a pseudonym. And I think in the world today, I'd rather have everybody know who's speaking——

Senator KENNEDY. Okay.

Mr. SMITH [continuing]. think about it.

Senator KENNEDY. Professor?

Professor HARTZOG. I'm afraid I'm going to disappoint you again, Senator——

Senator KENNEDY. Okay.

Professor HARTZOG [continuing]. With a not-straight answer. But I agree——

Senator KENNEDY. How do you feel about breakfast food?

[Laughter.]

Professor HARTZOG. Right. I am pro-breakfast food.

Senator KENNEDY. Okay.

Professor HARTZOG. So we agree on that. I agree with Mr. Smith. I think that there are circumstances where you'd want to preserve anonymous speech, and there are some where you absolutely would want to know who did it.

Senator KENNEDY. Okay. Well, I'm not—I don't want to go too over. Obviously, this is an important subject, and the extent to which I think—let me rephrase that. The extent of most Senators' knowledge in terms of the nuances of AI, their general impression is that AI has extraordinary potential to make our lives better if it doesn't make our lives worse first. And that's about the extent of it.

And in my judgment, we're not nearly ready to be able to craft a bill that looks like somebody designed it on purpose. I think we're more likely to take baby steps.

And I ask you those questions, predictably, because Senator Schatz and I have a bill. It's very simple. It says if you own a robot that's going to spit out artificial content to consumers, consumers have the right to know that it was generated by a robot and who owns the robot. And I think that's a good place to start.

But again, I want to thank my colleagues here, my Chair and my Ranking Member. They know a lot about this subject, and I want to hear their questions, too. Thank you, all, for coming.

Senator HAWLEY [presiding]. Thank you, Senator Kennedy. On behalf of the Chairman, we're going to start a second round. And I guess I'll go first since I'm the only one sitting here. It's bad news for the—bad news for the witnesses.

Senator KENNEDY. Well, I came just to hear you.

Senator HAWLEY. Oh, I'm sure.

Senator KENNEDY. I mean it.

Senator HAWLEY. Yes, yes.

Let me just—Mr. Smith, let me come back to this. We were talking about kids, and kids' privacy and safety. Thanks for the information you're going to get me. Let me give you an opportunity, though, to maybe make a little news today in the best possible way.

Thirteen, the age limit for Bing's Chat. That's such a young age. I mean, listen, I've got three kids at home; 10, 8, 2 are my kids. I can't—I don't want my kids to be interacting in the chatbots anytime soon at all. But 13 is so incredibly young.

Would you commit today to raising that age? And would you commit to a verifiable age verification procedure such that parents can know, they can have some sense of confidence that their 12-year-old is not just saying to Bing, "Yes, yes, yes, I'm 13. Yes, I'm 15. Sure. Go right on ahead. 'Now let's get into it back and forth with this robot'"—as Senator Kennedy said. Would you commit to those things on behalf of child safety today?

Mr. SMITH. Look, as you can imagine, the teams that work at Microsoft let me go out and speak, but they probably have one principle they want me to remember: Don't go out and make news without talking to them first.

Senator HAWLEY. But you're the boss.

Mr. SMITH. Yes. Let's just say wisdom is important. And most mistakes you make, you make when you make them by yourself. I'm happy to go back and talk more about what the right age should be. That's——

Senator HAWLEY. Don't you think 13 is awfully low, though?

Mr. SMITH. It depends for what, actually.

Senator HAWLEY. To interact with a robot——

Mr. SMITH. No.

Senator HAWLEY [continuing]. Who could be telling you to do any number of things. Don't you think that's awfully young?

Mr. SMITH. Not necessarily. Let me describe——

Senator HAWLEY. Really?

Mr. SMITH. It is the scenario. When I was in Seoul, Korea, a couple of months ago, we met with the deputy prime minister, who's also the minister of education. And they're trying to create for three topics that are very objective: math, coding, and learning English—a digital textbook with an AI tutor. So that if you're doing math and you don't understand a concept, you can ask the AI tutor to help you solve the problem.

And by the way, I think it's useful not only for the kids, I think it's useful for the parents. And I think it's good. Let's just say a 14-year-old, let's say—what's the age of eighth grade algebra? You know, most parents say—I found when my kids were in eighth grade algebra, I tried to help them with their homework, they didn't believe I ever made it through the class. I think we want kids, in a controlled way with safeguards, to use something that way.

Senator HAWLEY. But we're not talking here about tutors. Well, I'm talking about your AI chat, Bing Chat. I mean, famously, earlier this year, your chatbot—you had a technology writer for The New York Times who wrote about this. I'm looking at the article right now. Your chatbot was urging this person to break up his marriage.

Mr. SMITH. I'm not sure——

Senator HAWLEY. Do we want 13-year-olds to be having those conversations?

Mr. SMITH. No, of course not.

Senator HAWLEY. Okay.

Mr. SMITH. Which is why——

Senator HAWLEY. Well, will you commit to raising the age?

Mr. SMITH. I actually don't want Bing Chat to break up anybody's marriage.

Senator HAWLEY. I don't either.

Mr. SMITH. So—but, but, but that——

Senator KENNEDY. There might be some exceptions.

[Laughter.]

Mr. SMITH. Yes. But we're not going to make the decision on the exception. The—no, but it goes to—this is—we have multiple tools. Age is a very red line.

Senator HAWLEY. It is a very red line——

Mr. SMITH. And there's——

Senator HAWLEY [continuing]. That's why I like it.

Mr. SMITH. And my point is, there is a safety architecture that we can apply to bring——

Senator HAWLEY. But your safety architecture didn't stop an adult. Didn't stop the chatbot from having this discussion with an adult in which it said, "You don't really love your wife. You want to—your wife isn't good for you. She doesn't really love you."

Now, can you im—this is an adult. Can you imagine the kind of things that your chatbot would say to a 13-year-old? I mean, I'm serious about this. Do you really think this is a good idea?

Mr. SMITH. Yes. But look, wait a second. Let's put that in context. At a point where the technology had been rolled out for only 20,000 people, a journalist for The New York Times spent 2 hours on the evening of Valentine's Day, ignoring his wife, and interacting with a computer, trying to break the system, which he managed to do. We didn't envision that use. And then——

Senator HAWLEY. Have you——

Mr. SMITH [continuing]. The next day we had fixed it——

Senator HAWLEY [continuing]. Have you—but——

Mr. SMITH [continuing]. That's the key thing.

Senator HAWLEY. Are you telling me that you've envisioned all the questions that 13-year-olds might ask, and that I, as a parent, should be—should be absolutely fine with that? Are you telling me that I should trust you in the same way that The New York Times writer did?

Mr. SMITH. What I am saying is, I think as we go forward, we have an increasing capability to learn from the experience of real people and——

Senator HAWLEY. Yes, see, that's what——

Mr. SMITH [continuing]. What the right——

Senator HAWLEY [continuing]. Worries me. That's exactly what worries me. Is what you're saying is we have to have some failures. I don't want 13-year-olds to be your guinea pig. I don't want 14-year-olds to be your guinea pig. I don't want any kids to be your guinea pig. I don't want you to learn from their failures.

You want to learn from the failures of your scientists, go right ahead. Let's not learn from the failures of America's kids. This is what happened with social media. We had social media who made billions of dollars giving us a mental health crisis in this country. They got rich, the kids got depressed, committed suicide. Why

would we want to run that experiment again with AI? Why not raise the age? You could do it.

Mr. SMITH. We shouldn't want—first of all, we shouldn't want anybody to be a guinea pig. I think that, regardless of age or anything else——

Senator HAWLEY. Good. Well, let's roll kids out right here. Right today, right now.

Mr. SMITH. No. But let's also recognize that technology does require real users. What's different about this technology and, which is so fundamentally different, in my view, from the social media experience, is that we not only have the capacity, but we have the will. And we are applying that will to fix things in hours and days.

Senator HAWLEY. Well—yes, to fix things after there's been—after the fact. I mean, I'm sorry. It just sounds to me like you're boiling down and you're saying, "Trust us. We're going to do well with this." I'm just asking you why we should trust you with our children.

Mr. SMITH. I'm not asking for trust. Although, I hope we will work every day to earn it. That's why you have a licensing obligation.

Senator HAWLEY. There isn't a licensing obligation——

Mr. SMITH. That's why——

Senator HAWLEY [continuing]. Right now.

Mr. SMITH [continuing]. In your framework, in my view——

Senator HAWLEY. Well, sure. But I'm asking you as the president of this company to make a commitment now for child safety and protection to say, "You know what? Microsoft is going to"—you can tell every parent in America now: "Microsoft is going to protect your kids. We will never use your kids as a science experiment ever. Never. And therefore, we're not going to allow your—we're not going to target your kids, and we're not going to allow your kids to be used by our chatbots as a source of information if they're younger than 18."

Mr. SMITH. But I think you're talking about—with all due respect, there's two things that you're talking about. And I think we're——

Senator HAWLEY. I'm just talking about protecting kids. This is very simple. Yes.

Mr. SMITH. Yes. No, but we don't want to use kids as a source of information and monetizing, etc. But I am equally of the view, I don't want to cut off an eighth grader today with the right or ability to use this tool that will help them learn algebra or math in a way that they couldn't a year ago.

Senator HAWLEY. Yes. Well [holds up document], with all due respect, it wasn't algebra or math that your chatbot was recommending or talking about when it was trying to break up some reporter's marriage.

Mr. SMITH. Of course not. But now we're mixing things and——

Senator HAWLEY. No, we're not. We're talking about your chatbot. We're talking about Bing Chat.

Mr. SMITH. Of course we're talking about Bing Chat. And I'm talking about the protection of children, and how we make technology better. And yes, there was that episode back in February on

Valentine's Day. Six months later, if that journalist tries to do the same thing again, it will not happen.

Senator HAWLEY. You want me to be done, Senator Klobuchar? [Laughter.]

Senator KLOBUCHAR. I just don't want to miss my vote.

Mr. SMITH. There's other—there's other witnesses.

Senator KLOBUCHAR. I don't want to miss my vote.

Senator HAWLEY. Senator Klobuchar.

Senator KLOBUCHAR. Oh, you are very kind. Thank you. Some of us haven't voted yet.

So I wanted to turn to you, Mr. Dally. In March, NVIDIA announced a partnership with Getty Images to develop models that generate new images using Getty's image library. Importantly, this partnership provides royalties to content creators. Why was it important to the company to partner with and pay for these—of Getty's image library—developing generative AI models?

Mr. DALLY. Well, NVIDIA, we believe in respecting people's intellectual property rights and the—you know, the rights of the, you know, photographers who produce the images that our model is trained on and are expecting income from those images we didn't want to infringe on.

So, you know, we did not just scrape a bunch of images off the web to train our model. We partnered with Getty, and we trained our model Picasso. And when people use Picasso to generate images, the people who provided the original, you know, content get remunerated. And we see this as a way of going forward in general, where people who are providing the IP that trains these models should benefit from the use of that IP.

Senator KLOBUCHAR. Okay. And today, the White House announced eight more companies that are committing to take steps to move toward safe, secure, and transparent development of AI and videos. NVIDIA is one of those companies—could you talk about the steps that you've taken, and what steps do you plan to take to foster ethical and responsible development of AI?

Mr. DALLY. So we've done a lot already. We have, you know, implemented our NeMo Guardrails. So we can basically put guardrails around our own large language model, NeMo, so that inappropriate prompts to the model don't get a response.

If the model inadvertently were to generate something that might be considered offensive, that is detected and intercepted before it can reach the user of the model. We have a set of guidance that we provide for all of our internally generated models and how, you know, they should be appropriately used.

We provide cards that sort of, you know, say where the model came from, what the data set is trained on, and then we test these models very thoroughly. And the testing depends upon the use.

So for certain models, we test them for bias. Right? Well, we want to make sure that when you refer to a doctor, it doesn't automatically assume it's a him. We test them in certain cases for safety. We have a variant of our NeMo model called BioNeMo that's used in the medical profession. We want to make sure that the advice that it gives is safe.

And there are a number of other measures. I——

Senator KLOBUCHAR. Okay.

Mr. DALLY [continuing]. Could give you a full list, if you wanted.

Senator KLOBUCHAR. Very good. Thank you.

Professor Hartzog, do you think Congress should be more focused on regulating the inputs and design of generative AI, or focus more on outputs and capabilities?

Professor HARTZOG. Oh, can't the answer, Senator, be both?

Senator KLOBUCHAR. Of course, it can.

Professor HARTZOG. Certainly, certainly. I think that the area that has been ignored, I think, up to this point, has been the design and inputs to a lot of these tools. And so to the extent that that area could use some revitalization, I would encourage inputs and outputs, design and uses.

Senator KLOBUCHAR. Okay. And I suggest you look at these election bills because, as we've all been talking about, I think we have to move quickly on those. And the fact that it's bipartisan has been a very positive thing, so.

Professor HARTZOG. Absolutely.

Senator KLOBUCHAR. I wanted to thank Mr. Smith for wearing a purple Vikings tie. I know that that maybe was an AI-generated message that you got to know that this would be a smart move with me after their loss on Sunday. I will remind you, they're playing Thursday night, so.

[Laughter.]

Mr. SMITH. As a native of Wisconsin, I can assure you it was an accident.

[Laughter.]

Senator KLOBUCHAR. Very good. All right. Thank you, all of you. We have a lot of work to do. Thanks.

Chair BLUMENTHAL [presiding]. Senator Blackburn.

Senator BLACKBURN. Thank you, Mr. Chairman.

Mr. Smith, I want to come to you first and talk about China and the Chinese Communist Party. The way they have gone about—and we've seen a lot of it on TikTok. They have these influence campaigns that they are running to influence certain thought processes with the American people.

I know you all just did a report on China. You covered some of the disinformation, some of the campaigns. So talk to me a little bit about how Microsoft, but then the industry as a whole, can combat some of these campaigns.

Mr. SMITH. I think there's a couple of things that we can think more about and do more about.

The first is, we all should want to ensure that our own products, and systems, and services are not used, say, by foreign governments in this manner.

And I think that there's room for the evolution of export controls and next generation export controls to help prevent that. I think there's also room for a concept that's worked since the 1990s in the world of banking and financial services. It's these "Know Your Customer" requirements.

And we've been advocates for those. So that if there is abuse of systems, the company that is offering the service knows who is doing it and is in a better position to stop it from happening.

I think the other side of the coin is using AI and advancing our defensive technologies, which really start with our ability to detect

what is going on. And we've been investing heavily in that space. That is what enabled us to produce the report that we published. It is what enables us to see the patterns in communications around the world.

And we're seeking to be a voice with many others that really calls on governments to, I'll say, lift themselves to a higher standard so that they're not using this kind of technology to interfere in other countries, and especially in other countries' elections

Senator BLACKBURN. In the report that you all did, and you were looking at China, did you look at, what I call the other members of the axis of evil, Russia, Iran, North Korea?

Mr. SMITH. We did. And that specific report that you're referring to was focused on what we call—it was East Asia. Yes, we see especially prolific activities, some from China, some from Iran. And really, the most global actor in this space is Russia.

And we've seen that grow during the war, but we've seen it, you know, really spiral in the recent years going back to the middle of the last decade. We estimate that the Russian government is spending more than $1 billion dollars a year on a global—what we call cyber influence operation.

Part of it targets the United States. I think their fundamental goal is to undermine public confidence in everything that the public cares about in the United States, but it's not unique to the United States. We see it in the South Pacific. We see it across Africa. And I do think it's a problem we need to do more to counter.

Senator BLACKBURN. So summing it up, you would see something like a "Know Your Customer" or a SWIFT system, things that apply to banking, that is there to help weed it out. You think that companies should increase their due diligence to make certain that their systems are appropriate, and then being careful about doing business with countries that may misuse a certain technology?

Mr. SMITH. Generally, yes. I think——

Senator BLACKBURN. Yes.

Mr. SMITH [continuing]. One can look at the specific scenarios and what's more high-risk, but a "Know Your Customer" requirement. We've also said a "Know Your Cloud" requirement, in effect——

Senator BLACKBURN. Okay.

Mr. SMITH [continuing]. So that these systems are deployed in secure data centers.

Senator BLACKBURN. Okay. Mr. Hartzog, let me come to you. I think one of the things as we look at AI detrimental impacts, you know, and we don't always want to look at the doomsday scenarios.

But we are looking at some of the reports on surveillance with the CCP surveilling the Uyghurs, with Iran surveilling women. And I think there are other countries that are doing the same type surveillance. So what can you do to prevent that? How do we prevent that?

Professor HARTZOG. Senator, I've argued in the past that facial recognition technologies and certain sorts of biometric surveillance are fundamentally dangerous, and that there's no world in which that actually should be safe for any of us, and that we should prohibit them outright. In the very least, prohibition of biometric surveillance in public spaces, prohibition of emotion recognition.

This is what I refer to as the strong bright-line measures that draws absolute lines in the sands rather than procedural ones that ultimately, I think, end up entrenching this kind of harmful surveillance.

Senator BLACKBURN. Okay. Mr. Chairman, can I take another 30 seconds? Because Mr. Dally was shaking his head in agreement on some things. I was catching that. Do you want to weigh in before I close my questioning on either of these topics?

Mr. DALLY. I was in general agreement, I guess, when I was shaking my head. I think, you know, we need to be very careful about who we sell our technology to. And, you know, NVIDIA, we, you know, try to, you know, sell to, you know, people who are using this for good commercial purposes and not to, you know, suppress others. And, you know, we will continue to do that because we don't want to see this technology misused to oppress anybody.

Senator BLACKBURN. Got it. Thank you. Thanks.

Chair BLUMENTHAL. Thanks, Senator Blackburn. My colleague Senator Hawley mentioned that we have a forum tomorrow, which I welcome. I think anything to aid in our education and enlightenment—"our," being Senators—is a good thing. And I just want to express the hope that some of the folks who are appearing in that venue will also cooperate and appear before this Subcommittee.

We will certainly be inviting more than a few of them. And I want to express my thanks to all of you for being here, but especially to Mr. Smith, who has to be here tomorrow to talk to my colleagues privately. And our effort is complimentary, not contradictory to what Senator Schumer is doing, as you know.

I'm very focused on election interference because elections are upon us. And I want to thank my colleagues, Senators Klobuchar, and Hawley, Coons, and Collins for taking a first step toward addressing the harms that may result from deepfakes, impersonation, all of the potential perils that we've identified here.

And it seems to me that authenticating the truth or ads that embody true images and voices is one approach. And then banning the deepfakes and impersonations is another approach. And obviously, banning anything in the public realm, in public discourse endangers running afoul of the First Amendment, which is why disclosure is often the remedy that we seek, especially in campaign finance.

So maybe I should ask all of you whether you see that banning certain kinds of election interference—and Mr. Smith, you raised the spectre of foreign interference, and the frauds, and scams that could be perpetrated, as they were in 2016.

And I think it is one of those nightmares that should keep us up at night because we are an open society. We welcome free expression. And AI is a form of expression, whether we regard it as free or not, and whether it's generated and high-risk, or simply touching up some of the background in the TV ad.

Maybe you can, each of you, talk a little bit about what you see the potential remedies there. Mr. Dally?

Mr. DALLY. So I think it is of grave concern, with the election season coming up, that the American public may be misled by deepfakes of various kinds. I think, as you mentioned, that the use of provenance to authenticate a true image or voice at its source,

and then tracking that to its deployment will let us know what a real image is. And if we insist on AI content—AI-generated content being identified as such, then people are at least tipped off that this is generated and not the real thing.

You know, I think that we need to avoid, you know, having some, you know, especially foreign, entity interfere in our elections. But at the same time, you know, AI-generated content is speech, and I think it would be a dangerous precedent to try to ban something. I think it's much better to have disclosure, as you suggested, than to ban something outright.

Chair BLUMENTHAL. Mr. Smith?

Mr. SMITH. Three thoughts.

Number one, 2024 is a critical year for elections, not only in this country. But it's not only for the United States, it's for the United Kingdom, for India, across the European Union. More than two billion people will vote for who is going to represent them. And so this is a global issue for the world's democracies.

Number two, I think you're right to focus in particular on the First Amendment because it's such a critical cornerstone for American political life and the rights that we all enjoy. And yet, I will also be quick to add, I don't think the Russian government qualifies for protection under the First Amendment.

And if they're seeking to interfere in our elections, then I think that the country needs to take a strong stand, and a lot of thought needs to be given as to how to do that effectively.

But then number three—and this I think goes to the heart of your question and why it's such a good one—I think it's going to require some real thought discussion and an ultimate consensus to emerge, let me say, around one specific scenario.

Let's imagine for a moment that there is a video that involves a Presidential candidate that originally was giving a speech. And then let's imagine that someone uses AI to put different words into the mouth of that candidate, and uses AI technology to perfect it to a level that it is difficult for people to recognize as fraudulent.

Then you get to this question, what should we do? And at least as I've been trying, and we've been trying to think this through, I think we have two broad alternatives. One is we take it down, and the other is we relabel it.

If we do the first, then we're acting as censors. And I do think that makes me nervous. I don't think that's really our role to act as censors. And the Government really cannot, I think, under the First Amendment. But relabeling to ensure accuracy, I think that is probably a reasonable path. But really what this highlights is the discussion still to be had, and I think, the urgency for that conversation to take place.

Chair BLUMENTHAL. And I will just say—and then I want to come to you, Professor Hartzog—that I agree emphatically with your point about the Russian government, or the Chinese government, or the Saudi government as potential interferers.

They're not entitled to the protection of our Bill of Rights when they are seeking to destroy those rights, and purposefully trying to take advantage of a free and open society to, in effect, decimate our freedoms.

So I think there is a distinction to be made there in terms of national security. And I think that rubric of national security, which is part of our framework, applies with great force in this area. And that is different from a Presidential candidate putting up an ad that, in effect, puts words in the mouth of another candidate.

And as you may know, we began these hearings with introductory remarks from me that were impersonation taken from my comments on the floor—taking my voice from speeches that I made on the floor of the United States Senate with content generated by ChatGPT that sounded exactly like something I would say in a voice that was indistinguishable from mine.

And, obviously, I disclosed that fact at the hearing. But in real time, as Mark Twain famously said, "A lie travels halfway around the world before the truth gets out of bed." And we need to make sure that there is action in real time if you're going to do the kind of identification that you suggested. Real time, meaning real time in a campaign which is measured in minutes and hours, not in days and months. Professor Hartzog?

Professor HARTZOG. Thank you, Senator. Like you, I'm nervous about just coming out and saying we're going to ban all forms of speech, particularly, when you're talking about something as important as this, political speech.

And like you, I also worry about disclosure alone as a half measure. And earlier in this hearing, it was asked, "What is a half measure?" And I think that goes toward answering your question today. I think the best way to think about half measures is an approach that is necessary, but not sufficient. That risks giving us the illusion that we've done enough, but ultimately—and I think this is the pivotal point—doesn't really disrupt the business model and the financial incentives that have gotten us here in the first place.

And so to help answer your question, one thing that I would recommend is thinking about throwing lots of different tools—which I applaud your bipartisan framework for doing, is bringing lots of different tools to bear on this problem, thinking about the role that surveillance, advertising plays in powering a lot of these harmful technologies and ecosystems that doesn't allow the system—the lie just to be created, but flourish and to be amplified.

And so I would think about rules and safeguards that we could do to help limit those financial incentives, borrowing from standard principles of accountability. Things like, we use disclosures where it's effective, it's not effective. You have to make it safe. And if you can't make it safe, it shouldn't exist.

Chair BLUMENTHAL. Yes. I think I'm going to turn to Senator Hawley for more questions. But I think this is a real conundrum. We need to do something about it. We need more than half measures. We can't delude ourselves by thinking with a false sense of comfort that we've solved the problem if we don't provide effective enforcement.

And to be very blunt, the Federal Elections Commission often has been less than fully effective—a lot less than fully effective in enforcing rules relating to campaigns.

And so, there again, an oversight entity with strong enforcement authority, sufficient resources, and the will to act is going to be

very important if we're going to address this problem in real time. Senator Hawley.

Senator HAWLEY. Mr. Smith, let me just come back to something you said, thinking about now workers. You talked about Wendy's, I think it was, that automating the drive-through and talking about, you know, this was a good thing. I just want to press on that a little bit.

Is it—is it a good thing that workers lose their jobs to AI, whether it's at Wendy's, or whether it's at Walmart, or whether it's at the local hardware store? I mean, is it—you pointed out that—your comment was that there's really no creativity involved in taking orders through the drive-through. But that is a job, oftentimes a first job for younger Americans.

But hey, in this economy where the wages of blue collar workers have been flat for 30, 40 years and running, what worries me is that oftentimes what we hear from the tech sector, to be honest with you, is that jobs that don't have creativity, as tech defines, that don't have value. I'm frankly scared to death that AI will replace lots of jobs that tech types think aren't creative and will leave even more blue collar workers without any place to turn.

So my question to you is, can we expect more of this? And is it really progress for folks to lose those kind of jobs that, you know—I'd suspect that's not the best paying job in the world, but at least it's a job. And do we really want to see more of these jobs lost?

Mr. SMITH. Well, to be clear, at first I didn't say whether it was a good or bad thing. I was asked to predict what jobs would be impacted and I identified that job as one that likely would be.

But let's, I think, step back, because I think your question is critically important. Let's first reflect on the fact that, you know, we've had about 200 years of automation that have impacted jobs, sometimes for the better, sometimes for the worst.

In Wisconsin, where I grew up, or in Missouri, where my father grew up, if you go back 150 years, it took 20 people to harvest an acre of wheat or corn. And now it takes one. So 19 people don't work on that acre anymore. And that's been an ongoing part of technology.

The real question is this, how do we ensure that technology advances so that we help people get better jobs, get the skills they need for those jobs, and hopefully, do it in a way that broadens economic opportunity rather than narrows it.

I think the thing we should be the most concerned by is that since the 1990s, and I think this is the point you're making, if you look at the flow of digital technology—you know, fundamentally we've lived in a world that has widened the economic divide.

Those people with a college or graduate education have seen their incomes rise in real terms. Those people with, say, a high school diploma or less, have seen their income level actually drop compared to where it was in the 1990s.

So what do we do now? Well, I'll at least say what I think our goal should be. Can we use this technology to help advance productivity for a much broader range of people, including people who didn't have the good fortune to go to, say, where you or I went to college or law school?

And can we do it in a way that not only makes them more productive, but actually reaps some of the dividends of that productivity for themselves in a growing income level? I think it's that conversation that we need to have.

Senator HAWLEY. Yes. I agree with you, and I hope that that is—I hope that that's what AI could do. You talked about the farm used to take 20 people to do what one person could do. It used to take thousands of people to produce textiles, or furniture, or other things in this company where now it's zero. So we can tell the tale in different ways.

I'm not sure that seeing working class jobs go overseas or be replaced entirely is a success story. In fact, I'd argue it's not at all. It's not a success story. And I'd argue more broadly that our economic policy the last 30 years has been downright disastrous for working people.

And tech companies, and financial institutions, and, certainly, banks, and Wall Street, they have reaped huge profits, but blue collar workers can barely find a good-paying job.

I don't want AI to be the latest accelerant of that trend. And so I don't really want every service station in America to be manned by some computer such that nobody can get a job anymore, get their foot in the door, or start their climb up the ladder. That that worries me.

Let me ask you about something else here in my expiring time. You mentioned national security.

Mr. SMITH. Mm-hmm.

Senator HAWLEY. Critically important. Of course, there's no national security threat that is more significant for the United States than China. Let me just ask you, is Microsoft too entwined with China?

You have the Microsoft Research Asia that was set up in Beijing back in the late 1990s. You've got centers now in Shanghai and elsewhere. You've got all kinds of cooperation with Chinese state-owned businesses. I'm looking at an article here from Protocol Magazine where one of their contributors said that Microsoft had been the alma mater of Chinese big tech.

Are you concerned about your degree of entwinement with the Chinese government? Do you need to be decoupling in order to make sure that our national security interests aren't fatally compromised?

Mr. SMITH. I think it's something that we need to be and are focused on. To some degree in some technology fields, Microsoft is the alma mater of the technology leaders in every country in the world because of the role that we've played over the last 40 years.

But when it comes to China today, we are and need to have very specific controls on who uses our technology, and for what, and how. That's why we don't, for example, do work on quantum computing, or we don't provide facial recognition services, or focus on synthetic media, or a whole variety of things.

While at the same time, when Starbucks has stores in China, I think it's good that they can run their services in our data center rather than a Chinese company's data center.

Senator HAWLEY. Well, just on facial recognition, I mean, back in 2016, your company released this database, MS-Celeb, 10 mil-

lion faces without the consent of the folks who were in the database. You eventually took it down although it took 3 years. China used that database to train much of its facial recognition software and technology.

I mean, it isn't that a problem? You said that Microsoft might be the alma mater of many companies AI, but China's unique. No? I mean, China is running concentration camps using digital technology like we've never seen before. I mean, isn't that a problem for your company to be in any way involved in that?

Mr. SMITH. We don't want to be involved in that in any way, and I don't believe we are. I think that——

Senator HAWLEY. Are you going to close your centers in China, your Microsoft Research Asia, Beijing, or your center in Shanghai?

Mr. SMITH. I don't think that will accomplish what you are asking us——

Senator HAWLEY. You're running thousands of people through your centers out into the Chinese government and Chinese state-owned enterprises. Isn't that a problem?

Mr. SMITH. First of all, there's a big premise, and I don't embrace the premise that that is in fact what we're doing. We are——

Senator HAWLEY. Well, which part is wrong?

Mr. SMITH. The notion that we're running thousands of people through, and then they're going into the Chinese government——

Senator HAWLEY. Is that not right? I thought you had 10,000 employees in China whom you've recruited from Chinese state-owned agencies, Chinese state-owned businesses. They come work for you, and then they go back to these state-owned entities.

Mr. SMITH. We have employees in China. In fact, we have that number. I don't—to my knowledge, that is not where they're coming from. That is not where they're going. We are not running that kind of revolving door. And it's all about what we do and who we do it with that I think is of paramount importance, and that's what we're focused on.

Senator HAWLEY. You'd condemn what the Chinese government's doing to the Uyghurs, and the Xinjiang Province, and all of that?

Mr. SMITH. We do everything we can to ensure that our technology is not used in any way for that kind of activity in China and around the world, by the way.

Senator HAWLEY. But you condemn it, to be clear?

Mr. SMITH. Yes.

Senator HAWLEY. What are your safeguards that you have in place such that your technology is not further enabling the Chinese government, given the number of people you employ there and the technology you develop there?

Mr. SMITH. Well, you take something like facial recognition, which is at the heart of your question. We have very tight controls that limit the use of facial recognition in China, including controls that, in effect, make it very difficult, if not impossible, to use it for any kind of real time surveillance at all.

And by the way, the thing we should remember, the U.S. is a leader in many AI fields. China's the leader in facial recognition technology and the AI for it. And——

Senator HAWLEY. Well, in part because of the information that you helped them acquire. No?

Mr. SMITH. No. It's because they have the world's most data.

Senator HAWLEY. Well, yes, but you gave them——

Mr. SMITH. No.

Senator HAWLEY [continuing]. Ten million.

Mr. SMITH. Well, it—I—I don't think that's——

Senator HAWLEY. You don't think that had anything to do with it?

Mr. SMITH. I don't think—when you have a country of 1.4 billion people and you decide to have facial recognition used in so many places, it gives that country a mass of data.

Senator HAWLEY. But are you saying that the database that Microsoft released in 2016, MS-Celeb, you're saying that that wasn't used by the Chinese government to train their facial recognition?

Mr. SMITH. I am not familiar with that, and I add it to the list. I'd be happy to provide you with information.

Senator HAWLEY. Okay.

Mr. SMITH. But my goodness, the advance in that facial recognition technology—if you go to another country where they're using facial recognition technology, it's highly unlikely it's American technology. It's highly likely that it's Chinese technology because they are such leaders in that field.

Which I think is fine. I mean, if you want to pick a field where the United States doesn't want to be a technology leader, I'd put facial recognition technology on that list. But let's recognize it's homegrown.

Senator HAWLEY. How much money has Microsoft invested in AI development in China?

Mr. SMITH. I don't know, but I will tell you this—okay. The revenue that we make in China, which accounts for, what, about one out of every six humans on this planet? It's 1.5 percent of our global revenue. It's not the market for us that it is for other industries or even some other tech companies.

Senator HAWLEY. Sounds then like you can afford to decouple?

Mr. SMITH. But is that the right thing to do?

Senator HAWLEY. Yes. And again, a regime that is fundamentally evil, that is inflicting the kind of atrocities on its own citizens that you just alluded to, that is doing to the Uyghurs, what it's doing, that it's running modern day concentration camps—yes, I think it is.

Mr. SMITH. But there's two questions that I think are at least are worthy of thought. Number one, do you want General Motors to sell or manufacture cars—let's just say sell cars in China? Do you want to create jobs for people in Michigan or Missouri so that those cars can be sold in China? If the answer to that is yes, then think about the second question.

How do you want General Motors in China to run its operations, and where would you like it to store its data? Would you like it to be in a secure data center run by an American company, or would you like it to be run by a Chinese company?

Which will better protect General Motors' trades secrets? I'll argue we should be there so that we can protect the data of American companies, European companies, Japanese companies. Even if

you disagree on everything else, that I believe serves this country well.

Senator HAWLEY. You know, I think you're doing a lot more than just protecting data in China. You have major research centers, tens of thousands of employees.

And to your question, do I want General Motors to be building cars in China? No, I don't. I want them to be making cars here in the United States with American workers. And do I want American companies to be aiding in any way the Chinese government and their oppressive tactics? I don't.

Senator Ossoff, would you like me to yield to you now? Are you ready?

Chair BLUMENTHAL. I have been very hesitant to interrupt——

Senator HAWLEY. You've been very, very patient.

Chair BLUMENTHAL [continuing]. The discussion. The conversation here has been very interesting. And I'm going to call on Senator Ossoff, and then I have a couple of follow-up questions.

Senator OSSOFF. Thank you, Mr. Chairman. And thank you, all, for your testimony. Just getting down to the fundamentals, Mr. Smith. If we're going to move forward with a legislative framework or a regulatory framework, we have to define clearly in legislative text precisely what it is that we are regulating.

What is the scope of regulated activities, technologies, and products? So how should we consider that question? And how do we define the scope of technologies, the scope of services, the scope of products that should be subject to a regime of regulation that is focused on artificial intelligence?

Mr. SMITH. I think there's three layers of technology on which we need to focus in defining the scope of legislation and regulation.

First is the area that has been the central focus of 2023 in the executive branch, and here on Capitol Hill. It's these so-called frontier or foundation models that are the most powerful, say, for something like generative AI.

In addition, there are the applications that use AI, or as Senators Blumenthal and Hawley have said, "the deployers of AI." If there is an application that calls on that model in what we consider to be a high-risk scenario, meaning it could make a decision that would have an impact on, say, the privacy rights, the civil liberties, the rights of children or needs of children, then I think we need to think hard, and have law and regulation that is effective to protect Americans.

And then the third layer is the data center infrastructure where these models and where these applications are actually deployed. And we should ensure that those data centers are secure, that there are cybersecurity requirements that the companies, including ours, need to meet.

We should ensure that there are safety systems at one, two, or all three levels if there is an AI system that is going to automate and control, say, something like critical infrastructure, such as the electrical grid.

So those are the areas where we would say start there with some clear thinking, and a lot of effort to learn and apply the details—but focus there.

Senator OSSOFF. As more and more models are trained and developed to higher levels of power and capability, there will be a proliferation—there may be a proliferation of models—perhaps not the frontier models, perhaps not those at the bleeding edge—that use the most compute of all, powerful enough to have serious implications.

So is the question, which models are the most powerful in a moment in time, or is there a threshold of capability or power that should define the scope of regulated technology?

Mr. SMITH. I think you've just posed one of the critical questions that, frankly, a lot of people inside the tech sector, and across the Government, and in academia, are really working to answer. And I think the technology is evolving and the conversation needs to evolve with it.

Let's just posit this: There's something like GPT-4 from OpenAI. Let's just posit it can do 10,000 things really well. It's expensive to create, and it's relatively easy to regulate in the scheme of things because there's 1, or 2, or 10.

But now let's go to where you're going, which I think is right. What does the future bring in terms of proliferation? Imagine that there's an academic at Professor Hartzog's university who says, "I want to create an open source model. It's not going to do 10,000 things well, but it's going to do 4 things well. It won't require as many NVIDIA GPUs. It won't, you know, require as much data."

But let's imagine that that could be used to create the next virus that could spread around the planet. Then you'd say, "Well, we really need to ensure that there's safety, architecture, and controls around that, as well."

And that's the conundrum. That's why this is a hard problem to solve. It's why we're trying to build safety architecture in our data centers so that open-source models can, say, run in them, and still be used in ways that will prohibit that kind of harm from taking place.

But as you think about a licensing regime, this is one of the hard questions. Who needs a license? You don't want it to be so hard that only a small number of big companies can get it. But then you also need to make sure that you're not requiring people to get it when they really, we would say, don't need a license for what they're doing.

And, you know, the beauty of the framework, in my view, is it starts to frame the issue. It starts to define the question——

Senator OSSOFF. Let me ask this question——

Mr. SMITH [continuing]. And how we work on——

Senator OSSOFF [continuing]. Is it a license to train a model to a certain level of capability? Is it a license to sell, or license access to that model? Or is it a license to purchase or deploy that model? Who is the licensed entity?

Mr. SMITH. That's another question that is key and may have different answers in different scenarios, but mostly I would say it should be a license to deploy that.

You know, I think that there may well be obligations to disclose to, say, an independent authority when a training run begins depending on what the goal—when the training run ends, so that an

oversight body can follow it just the way, say, might happen when a company's building a new commercial airplane.

And then there are—what's emerging—the good news is there's emerging a foundation of, call it best practices, for then how the model should be trained, what kind of testing there should be, what harms should be addressed. That's a big topic that needs——

Senator OSSOFF. When you say——

Mr. SMITH [continuing]. Discussion.

Senator OSSOFF. Forgive me, Mr. Smith.

Mr. SMITH. Yes.

Senator OSSOFF. When you say, "a license to deploy"——

Mr. SMITH. Yes.

Senator OSSOFF [continuing]. Do you mean, for example, if a Microsoft Office product wishes to use a GPT model for some user-serving purpose within your suite, you would need a license to deploy GPT in that way? Or do you mean that GPT would require a license to offer to Microsoft?

And putting aside whether or not this is a plausible commercial scenario, the question is what's the structure of the licensing arrangement?

Mr. SMITH. In this case, it's more the latter. Imagine—look, think about it like Boeing. Boeing builds a new plane. Before it can sell it to United Airlines and United Airlines can start to fly it, the FAA is going to certify that it's safe.

Now, imagine where it—call it GPT–12, whatever you want to name it. You know, before that gets released for use, I think you can imagine a licensing regime that would say that it needs to be licensed after it's been, in effect, certified as safe.

And then you have to ask yourself, well, how do you make that work so that we don't have the Government slow everything down? And what I would say is you bring together three things.

First, you need industry standards so that you have a common foundation and well-understood way as to how training should take place.

Second, you need national regulation.

And third, if we're going to have a global economy, at least in the countries where we want these things to work, you probably need a level of international coordination.

And I'd say, look at the world of civil aviation. That's fundamentally how it has worked since the 1940s. Let's try to learn from it and see how we might apply something like that, or other models, here.

Senator OSSOFF. Mr. Dally, how would you respond to the question, in a field where the technical capabilities are accelerating at a rapid rate, future rate unknown, where and according to what standard, or metric, or definition of power do we draw the line for what requires a license for deployment and what can be freely deployed without oversight by the Government?

Mr. DALLY. I think it's a tough question because I think you have to balance two important considerations. The first is, you know, the risks presented by a model of whatever power, and on the other side is the fact that, you know, we would like to ensure that the U.S. stays ahead in this field.

And to do that, we want to make sure that, you know, individual academics and entrepreneurs with a good idea can, you know, move forward, and innovate, and deploy models without huge barriers.

Senator OSSOFF. So it's the capability of the model. It's the risk presented by its deployment without oversight. Is that the—because the thing is, we're going to have to write legislation, and the legislation is going to have to, in words, define the scope of regulated products. And so we're going to have to bound that which is subject to a licensing arrangement or wherever we land and that which is not. So——

Mr. DALLY. I think it is——

Senator OSSOFF [continuing]. What I want is the very—and so how do you—I mean——

Mr. DALLY. It is dependent on the application. Because if you have a model which is, you know, basically determining a medical procedure, there's a high-risk with that, you know, depending on the patient outcome.

If you have another model which is, you know, controlling the temperature in your building, if it gets a little bit wrong you may be, you know, consume a little bit too much power, or maybe, you know, you're not as comfortable as you would be. But it's not a life-threatening situation.

So I think you need to regulate the things that are, have high consequences if the model goes awry.

Senator OSSOFF. And I'm on the Chairman's borrowed time. So just tap the gavel when you want me to stop.

Chair BLUMENTHAL. You had to wait.

Senator OSSOFF. That's true.

Chair BLUMENTHAL. So we'll give you a couple of——

Senator OSSOFF. Okay. Good.

Chair BLUMENTHAL [continuing]. Extra minutes.

Senator OSSOFF. Okay. Professor—and I'd be curious to hear from others as concisely with respect for the Chairman's follow-ups—how does any of this work without international law?

I mean, isn't it correct that a model, potentially a very powerful and dangerous model, for example, whose purpose is to unlock CBRN or mass destructive virological capabilities to a relatively unsophisticated actor, once trained, it's relatively lightweight to transport? And without, A, an international legal system, and B, a level of surveillance, that seems inconceivable into the flow of data across the internet—how can that be controlled and policed?

Professor HARTZOG. It's a great question, Senator. And with respect to being efficient, in my answer, I'll simply say that there are going to be limits even assuming that we do need international cooperation, which I would agree with you.

I mean, we've already started thinking about ways in which, for example, within the EU, which has already deployed some significant AI regulation, we might design frameworks that are compatible with that that requires some sort of interaction.

But ultimately what I worry about is actually deploying a level of surveillance that we've never before seen in an attempt to perfectly capture the entire chain of AI. And that's simply not possible. And so——

Senator OSSOFF. And I share that concern about privacy, which is in part why I raised the point. I mean, how can we know what folks are loading a lightweight model, once trained, onto perhaps a device that's not even online anymore?

Professor HARTZOG. Right. Yes, there are limits, I think, to what——

Senator OSSOFF. Right.

Professor HARTZOG [continuing]. We'll ever be able to know——

Senator OSSOFF. Either of you want to take a stab before I get gaveled out here?

Mr. SMITH. I would just say you're right. There's going to need—to be a need for international coordination. I think it's more likely to come from like-minded governments than perhaps global governance, at least in the initial years. I do think there's a lot we can learn.

We were talking with Senator Blackburn about the SWIFT system for financial transactions.

And, you know, somehow we've managed globally and especially in the United States for 30 years to have "Know Your Customer" requirements obligations for banks. Money has moved around the world. Nothing is perfect. I mean, that's why we have laws. But it's worked to do a lot of good to protect against, say, terrorist or criminal uses of money that would cause concern.

Mr. DALLY. Well, I think you're right in that these models are very portable. You could put the parameters of most models, even the very large ones, on a large USB drive and, you know, carry it with you somewhere. You could also train them in a data center anywhere in the world.

So, you know, I think it's really the use of the model and the deployment that you can effectively regulate. It's going to be hard to regulate the creation of it because if people can't create them here, they'll create them somewhere else.

And I think we have to be very careful if we want the U.S. to stay ahead—that we want the best people creating these models here in the U.S. and not to go somewhere else where, you know, the regulatory climate has driven them.

Senator OSSOFF. Thank you. Thank you, Mr. Chairman.

Chair BLUMENTHAL. Thank you, Senator Ossoff. I hope you are okay with a few more questions. We've been at it for a while. You've been very patient.

Mr. SMITH. Do we have a choice?

[Laughter.]

Chair BLUMENTHAL. No.

[Laughter.]

Chair BLUMENTHAL. But thank you very much. It's been very useful. I want to follow up on a number of the questions I've asked.

First of all, on the international issue, there are examples and models for international cooperation. Mr. Smith, you mentioned civil aviation. The 737 MAX—I think I have it right—when it crashed, it was a plane that had to be redone in many respects.

And companies' airlines around the world looked to the United States for that redesign and then approval. Civil aviation, atomic energy—not always completely effective, but it has worked in many respects.

And so I think there are international models here where, frankly, the United States is a leader by example and best practices are adopted by other countries when we support them. And frankly, in this instance, the EU has been ahead of us in many respects regarding social media. And we are following their leadership by example.

I want to come to this issue of having centers—whether they're in China or, for that matter, elsewhere in the world—requiring safeguards so that we are not allowing our technology to be misused in China against the Uyghurs, and preventing that technology from being stolen, or people we train there from serving bad purposes.

Are you satisfied, Mr. Smith, that it is possible, in fact, that you are doing it in China, that is preventing the evils that could resolve from doing business there in that way?

Mr. SMITH. I would say two things. First, I feel good about our track record, and our vigilance, and the constant need for us to be vigilant about what services we offer to whom and how they're used. It's really those three things. And I would take from that what I think is probably the conversation we'll need to have as a country about export controls more broadly.

There's three fundamental areas of technology where the United States is today, I would argue, the global leader.

First, the GPU chips from a company like NVIDIA.

Second, the cloud infrastructure from a company like, say, Microsoft.

And the third, is the foundation model from a firm such as OpenAI, and, of course, Google, and AWS, and other companies are global leaders as well.

And I think if we want to feel that we're good in creating jobs in the United States by inventing and manufacturing here—as you said, Senator Hawley, which I completely endorse—and good, the technology is being used properly, we probably need an export control regime that weaves those three things together.

For example, there might be a country in the world—let's just set aside China for a moment. Leave that out. Let's just say there's another country where you, all, in the executive branch would say, "We have some qualms, but we want U.S. technology to be present and we want U.S. technology to be used properly," the way that would make you feel good.

You might say then we'll let NVIDIA export chips to that country to be used in, say, a data center of a company that we trust that is licensed, even here, for that use, with the model being used in a secure way in that data center with a "Know Your Customer" requirement, and with guardrails that put certain kinds of use off limits.

That may well be where Government policy needs to go, and how the tech sector needs to support the Government and work with the Government to make it a reality.

Chair BLUMENTHAL. I think that that answer is very insightful and raises other questions. I would kind of analogize this situation to nuclear proliferation. We cooperate over safety in some respects with other countries, some of them adversaries, but we still do ev-

erything in our power to prevent American companies from helping China or Russia in their nuclear programs.

Part of that non-proliferation effort is through export controls. We impose sanctions. We have limits and rules around selling and sharing certain chokepoint technologies related to nuclear enrichment, as well as biological warfare, surveillance, and other national security risks. And our framework, in fact, envisions sanctions and safeguards precisely in those areas for exactly the reasons we've been discussing here.

Last October, the Biden administration used existing legal authorities as a first step in blocking the sale of some high-performance chips and equipment to make those chips to China. And our framework calls for export controls, and sanctions, and legal restrictions.

So I guess a question that we will be discussing—we're not going to resolve it today, regrettably. But we would appreciate your input going forward, and I'm inviting any of the listening audience here in the room, or elsewhere, to participate in this conversation on this issue and others—how should we draw a line on the hardware and technology that American companies are allowed to provide, anyone else in the world, any other adversaries or friends? Because as you've observed, Mr. Dally, and I think all of us accept, it's easily proliferated.

Mr. DALLY. Yes. If I could comment on this?

Chair BLUMENTHAL. Sure.

Mr. DALLY. You drew an analogy to, you know, nuclear regulation and mentioned the word chokepoint. And I think the difference here is that there really isn't a chokepoint. And I think there's a careful balance to be made between, you know, limiting, you know, where, you know, our chips go and what they're used for.

And, you know, disadvantaging American companies and the whole food chain that feeds them because—you know, we're not the only people who make chips that can do AI. I wish we were, but we're not.

There are companies around the world that can do it. There are other American companies. There are companies in Asia. There are companies in Europe. And if people can't get the chips they need to do AI from us, they will get them somewhere else.

And what will happen then is—you know, it turns out that chips aren't really the things that make them useful. It's the software. And if all of a sudden the standard chips for people to do AI become something from—you know, pick a country, Singapore, you know, all of a sudden all the software engineers will start writing all the software for those chips.

They'll become the dominant chips. And, you know, the leadership of that technology area will have shifted from the U.S. to Singapore or whatever other country becomes dominant.

So we have to be very careful to balance, you know, the national security considerations and the abusive technology considerations against preserving the U.S. lead in this technology area.

Chair BLUMENTHAL. Mr. Smith.

Mr. SMITH. Yes. It's a really important point. And what you have is the argument-counter argument. Let me for a moment channel what Senator Hawley often voices that I think is also important.

Sometimes you can approach this and say, "Look, if we don't provide this to somebody, somebody else will. So let's not worry about it." I get it.

But at the end of the day, you know, whether you're a company or a country, I think you do have to have clarity about how you want your technology to be used.

And, you know, I fully recognize that there may be a day in the future after I retire from Microsoft when I look back, and I don't want to say, "Oh, we did something bad. Because if we didn't, somebody else would have."

I want to say, "No, we had clear values, and we had principles, and we had in place guardrails and protections. And we turned down sales so that somebody couldn't use our technology to abuse other people's rights. And if we lost some business, that's the best reason in the world to lose some business."

And what's true of a company is true as a country. And so I'm not trying to say that your view shouldn't be considered. It should. That's why this issue is complicated. How to strike that balance?

Chair BLUMENTHAL. Professor Hartzog, do you have any comment?

Professor HARTZOG. I think that was well said. And I would only add that it's also worth considering in this discussion about how we sort of safeguard these incredibly dangerous technologies and the risk that could happen if they, for example, proliferated. If it's so dangerous, then we need to revisit the existential question again.

And I just bring it back to thinking not only about how we put guardrails on, but how we lead by example, which I think you brought up which is really important. And we don't win the race to violate human rights. Right? And that's not one that we want to be running.

Chair BLUMENTHAL. And it isn't simply Chinese companies importing chips from the United States and building their own data centers. Most AI companies rent capabilities from cloud providers. We need to make sure that the cloud providers are not used to circumvent our export controls or sanctions.

Mr. Smith, you raised the "Know Your Customer" rules. Knowing your customers would require cloud AI, cloud providers whose models are deployed to know what companies are using those models.

If you're leasing out a supercomputer, you need to make sure that your customer isn't the People's Liberation Army. That it isn't being used to subjugate Uyghurs. That it isn't used to do facial recognitions on dissidents or opponents in Iran, for example.

But I do think that you've made a critical point, which is, there is a moral imperative here. And I think there—there is a lesson in the history of this great country, the greatest in the history of the world, that when we lose our moral compass, we lose our way.

And when we simply do economic or political interests, sometimes it's very shortsighted, and we wander into a geopolitical swamp and quicksand. So I think these kinds of issues are very important to keep in mind when we lead by example.

I want to just make a final point, and then if Senator Hawley has questions, we're going to let him ask. But on this issue of worker displacement, I mentioned at the very outset, I think we are on the cusp of a new industrial revolution.

We've seen this movie before, as they say. And it didn't turn out that well in the industrial revolution where workers were displaced en masse. Those textile factories and the mills in this country and all around the world went out of business, essentially, or replaced the workers with automation and mechanics.

And I would respond by saying, we need to train those workers. We need to provide education. You've alluded to it, and it needn't be a 4-year college.

You know, in my State of Connecticut—Electric Boat, Pratt & Whitney, Sikorsky—defense contractors are going to need thousands of welders, electricians, tradespeople of all kinds who will have not just jobs, they'll have careers that require skills that, frankly, I wouldn't begin to know how to do. And I haven't the aptitude to do. And that's no false modesty.

So I think there are tremendous opportunities here, not just in the creative spheres that you have mentioned where, you know, we may think higher human talents may come into play, but in all kinds of jobs that are being created daily already in this country.

And as I go around the State of Connecticut, the most common comment I hear from businesses: "We can't find enough people to do the jobs we have right now. We can't find people to fill the openings that we have." And that is, in my view, maybe the biggest challenge for the American economy today.

Mr. SMITH. I think that is such an important point. And it's really worth putting everything we think about for jobs because I wholeheartedly endorse, Senator Hawley, what you were saying before, about let's—we need—we want people to have jobs, we want them to earn a good living, etc.

First, let's consider the demographic context in which jobs are created. The world has just entered a shift of the kind that it literally hasn't seen since the 1400s, namely, populations that are leveling off or in much of the world now declining.

One of the things we look at is every country and measure over 5-year periods, is the working age population increasing or decreasing? And by how much?

From 2020 to 2025, the working age population in this country, people aged 20 to 64, is only going to grow by one million people. The last time it grew by that smaller number, you know who was President of the United States? John Adams. That's how far back you have to go.

And if you look at a country like Italy, take that group of people over the next 20 years, it's going to decline by 41 percent. And what's true of Italy is true almost to the same degree in Germany. It's already happening in Japan, in Korea.

So we live in a world where for many countries, we suddenly encounter what you actually find, I suspect, when you go to Hartford, or St. Louis, or Kansas City. People can't find enough police officers, enough nurses, enough teachers, and that is a problem we need to desperately focus on solving. So how do we do that? I do think AI is something that can help.

And even in something like a call center. One of the things that's fascinating to me, we have more than 3,000 customers around the world running proofs of concept. One fascinating one is a bank in the Netherlands.

They said, you go into a call center today, the desks of the workers look like a trading floor in Wall Street. They have six different terminals. Somebody calls, they're desperately trying to find the answer to a question.

You know, with something like GPT–4, with our services, six terminals can become one. Somebody who's working there can ask a question, the answer comes up. And what they're finding is that the person who's answering the phone, talking to a customer, can now spend more time concentrating on the customer and what they need.

And I appreciate all the challenges. There's so much uncertainty. We desperately need to focus on skilling, but I really do hope that this is an era where we can use this to, frankly, help people fill jobs, get training, and focus more—let's just put it this way. I'm excited about artificial intelligence. I'm even more excited about human intelligence.

And if we can use artificial intelligence to help people exercise more human intelligence and earn more money doing so, that would be something that would be way more exciting to pursue than everything that we've had to grapple with the last decade around, say, social media and the like.

Chair BLUMENTHAL. Well our framework very much focuses on treatment of workers, about providing more training. It may not be something that this entity will do, but it is definitely something that it has to address.

And it's not only displacement, but also working conditions and opportunities within the workplace for promotion, to prevent discrimination, to protect civil rights. We haven't talked about it in detail, but we deal with it in our framework in terms of transparency around decision-making.

And, you know, China may try to steal our technology, it can't steal our people. And China has its own population challenges with the need for more people, skilled labor. But I say about Connecticut, you know, we don't have gold mines, or oil wells. What we have is a really able workforce, and that's going to be the key to, I think, America's economy in the future. And AI can help promote development of that workforce. Senator Hawley, anything from you?

Senator HAWLEY. Nothing further.

Chair BLUMENTHAL. You all have been really patient and so has our staff. I want to thank our staff for this hearing, but most important, we're going to continue these hearings. It is so helpful to us. I can go down our framework and tie the proposals to specific comments made by Sam Altman or others who have testified before, and we will enrich and expand our framework with the insights that you have given us.

So I want to thank all of our witnesses, and, again, look forward to continue our bipartisan approach here. You made that point, Mr. Smith. We have to be bipartisan and adopt full measures, not half measures.

Thank you, all. This hearing is adjourned.

[Whereupon, at 4:52 p.m., the hearing was adjourned.]

[Additional material submitted for the record follows.]

# A P P E N D I X

**Submitted by Chair Blumenthal:**

California Privacy Protection Agency (CPPA), letter ..... 119
Center for AI and Digital Policy (CAIDP), letter and attachment ..... 123

**Testimony of Bill Dally, Chief Scientist and Head of Research, NVIDIA**
Before the
U.S. Senate Judiciary Committee
Subcommittee on Privacy, Technology, and the Law

Hearing on "Oversight of A.I.: Legislating on Artificial Intelligence"
September 12, 2023

Chairman Blumenthal, Ranking Member Hawley, and esteemed Judiciary Committee members, thank you for the privilege to testify today. I am NVIDIA's Chief Scientist and Head of Research, and I am delighted to provide our perspective on our artificial intelligence (AI) journey and future.

**Our History**

NVIDIA is at the forefront of accelerated computing and generative AI, technologies that have the potential to transform industries, address global challenges, and profoundly benefit society.

Since our founding in 1993, we have been committed to developing technology to empower people and improve the quality of life worldwide.

- 1993: Founded to bring 3D graphics to the gaming and multimedia markets.
- 1999: Invented the GPU, the graphics processing unit, which would later reshape the computing industry.
- 2006: Introduced the CUDA API, providing a GPU programming capability to researchers for applications in every field, including science, research, and enterprise.
- 2012: Spurred AI research and supported the breakthrough AlexNet neural network.
- 2018: Reinvented computer graphics with NVIDIA RTX, the first GPU capable of real-time ray tracing. From 2012 to 2022, pioneered architecture innovations resulting in a 1000x increase in AI performance, enabling modern generative AI.

Today, over 40,000 companies use NVIDIA platforms – across media and entertainment, scientific computing, healthcare, financial services, internet services, automotive, and manufacturing – to solve the world's most difficult challenges and bring new products and services to consumers worldwide.

We provide over three hundred libraries of accelerated code built on CUDA, our parallel computing developer platform and application programming interface. These libraries help developers take advantage of GPU-accelerated platforms. Today, our GPUs accelerate thousands of applications in every field, benefitting enterprises, consumers, and academics worldwide. Our ecosystem has attracted a global community of over 4 million developers, making NVIDIA a globally adopted computing platform.

**Our Long AI Journey**

AI has captured headlines in recent months, but AI did not arrive overnight.

Modern AI algorithms, based on deep neural networks, which are algorithms that use interconnected nodes in a layered structure, have been around since the 1950s. At that

time, neural networks were largely academic pursuits; industry and policymakers paid them little attention. In the 80s, AI was dominated by logic programming and expert systems. A large industrial effort was made to commercialize expert systems, but they had little long-term impact.

At our founding in 1993, we were a 3D graphics startup, one of dozens of startups competing to create an entirely new market for accelerators to enhance graphics for computer games. At first, nobody seemed to see a need for another processor, much less one made by an untested startup. While others fell by the wayside, we forged ahead, inventing the GPU in 1999.

In contrast to CPUs, which are serial processors, our GPUs perform a massive number of calculations in parallel. When we launched the GPU for high-end gaming graphics, we recognized that GPUs could theoretically accelerate any application that could benefit from massively parallel processing.

But we needed to build an ecosystem and would have to convince developers to write applications for our products, one developer at a time. We firmly believed in our vision and set to work on it. In 2006, we unveiled the CUDA programming model, opening the parallel processing capabilities of GPUs to science and research. Even with CUDA, success did not come overnight. One by one, we worked with researchers and developers to teach the benefits of GPUs and CUDA.

We kept our nose to the grindstone, investing everything we could in research and development and building the ecosystem, even in lean years. We continued to refine CUDA and advance GPU technology, convinced that if we could innovate at the speed of light, the applications and the benefits to society would follow.

Over time, developers and researchers saw that accelerated computing could save time, energy, and money in various fields. As researchers learned to use CUDA, they realized that GPUs could accelerate the neural network algorithms that had previously resided only in textbooks. Neural networks require massively parallel processing, the raison d'etre of the GPU.

Thirteen years after NVIDIA invented the GPU, in 2012, researchers trained the AlexNet computer vision model with 14 million images on a pair of NVIDIA GPUs. The AlexNet model won the ImageNet challenge— an annual contest for image recognition algorithm accuracy — by a wide margin, igniting a flurry of research, largely performed on our GPU-powered platforms.

AlexNet didn't result in an immediate wave of commercial AI applications, but it left no doubt that we were on the right track. We redoubled our investment in our GPU-accelerated platforms and systems for the world's leading AI researchers. And each year, researchers clamored for more.

Five years after AlexNet, researchers using NVIDIA platforms invented the transformer model. After another five years, OpenAI scaled this model and its data set to release the GPT-3 large language model running on thousands of NVIDIA GPUs. This achievement led to the creation of ChatGPT, delighting and amazing users worldwide. Today, researchers across the globe innovate on NVIDIA GPUs.

**AI's Potential and Responsible Design**

As we look toward the future, we are thrilled to see generative AI's enormous potential, thanks to the remarkable developers who had faith in NVIDIA platforms. We are grateful to them and proud that our collective efforts have led to advances in AI that will revolutionize healthcare, medical research, education, business, and beyond.

For example, healthcare collaborations with academic institutions have led to large language models that can accurately predict a patient's risk of 30-day readmission, among other clinical outcomes. Imagine a world where AI-driven chatbots assist healthcare providers, helping them ask the right follow-up questions and ensuring that every possible issue is considered and evaluated and that the treatment is the best possible, meets all patient needs, and is explained to the patient in a way they can understand. AI will help radiologists correctly and rapidly interpret scans, dramatically reducing false results and saving thousands of lives. Medical AI will not eliminate the need for, or value of, human judgment, expertise, and care. Still, it will augment and raise the baseline abilities of human practitioners and provide a safety net to protect patients and practitioners alike. AI will never replace your human doctor or nurse, nor should it. But it will dramatically help them provide more personalized, accurate, efficient, and accessible healthcare.

In education, generative AI will expand access and uplift underserved communities worldwide. Teachers at all levels will be able to consult generative AI as a teaching assistant on nearly every topic, empowering educators to reach all students. Generative AI will be a valuable assistant for teachers in grading tests, creating new problem sets, and tutoring students who need more attention. As with healthcare, AI will never eliminate the need for students to connect with teachers, but it will enhance the experience for educators, students, and parents worldwide.

Generative AI can empower new tools to help businesses operate more efficiently and ensure compliance with financial, employment, commercial, and other laws. Just as an AI application can assist a doctor or an educator, an AI application can serve as a valuable assistant to ensure that the books and records of a company are in good order and reported accurately, to identify and mitigate fraud, and to ensure equal access and fair treatment in employment. In other words, we can use AI as a real-time auditor to help us identify and remedy fraud and unwanted bias *before* it causes harm. Again, AI will not eliminate the need for human professionals in business, financial, legal, and employment fields—rather it will empower them to do their jobs better than previously imaginable.

In our own field of computing, generative AI will make programmers and designers more productive, helping to reduce a workforce shortage that is challenging our country's ability to compete in this critical industry. Furthermore, AI-assisted design tools will be developed to provide real-time assistance to software developers—highlighting potential vulnerabilities before they are deployed, and in near real-time, suggesting secure alternatives.

AI will also be critical to the U.S.'s effort to cut greenhouse gases and combat climate change. AI will make every industry more efficient, and that efficiency will drive a reduction in greenhouse gas emissions. For example, AI will enhance and add intelligence to factories, transportation, logistics, HVAC systems, power generation, power grids, construction, and agriculture.

AI-driven innovations and management will create other environmental benefits, including enhancing water quality and reducing pollution. For example, AI-enabled water quality monitoring will swiftly detect and drive mitigation solutions to hazardous conditions and spills in our waterways and reservoirs. AI-enhanced forest monitoring will detect forest pest infestations, allowing for prompt treatment and mitigation. AI-powered air quality monitoring will detect and help mitigate toxic leaks. AI will enable targeted application of herbicides, dramatically reducing the chemical burden in our fields.

And, of course, AI applied to common computational use cases will dramatically reduce global greenhouse gas emissions while simultaneously adding significantly to the gross domestic product of nations that adopt it. Although significant energy is required to train models, once trained, running AI models saves energy. For example, AI-based simulations for fluid dynamics, weather, and climate are up to 100,000 times more efficient than conventional numerical simulations, allowing the same computation to be run at a tiny fraction of the energy use.

We recognize that, like any new product or service, AI products and services have risks. As it should be, the uses of AI systems are subject to our nation's laws and regulations, and everyone who makes, uses, or sells AI-enabled products and services is responsible for their conduct.

For example, like any other medical tool, the use or deployment of a medical AI chatbot must comply with the laws and regulations overseen by the FDA. Likewise, employment decisions must comply with all laws regarding bias and discrimination, whether assisted by an AI application or not. In financial services, lending decisions are subject to strict regulation—no exceptions for AI tools exist. Every endeavor that poses a serious risk of harm— from driving, flying, medicine, banking, engineering, and even legal services— is subject to licensing requirements that protect the public and promote quality, trustworthy products and services. AI-enabled services in high-risk sectors should be subject to

licensing requirements as well. Other applications, with less risk of harm, may need less stringent licensing and/or regulation.

Benefitting from clear, stable, and thoughtful regulation, AI developers will work to benefit society while making our products and services as safe as possible. As we deploy AI more broadly, we can and will continue to identify and address risks— whether data privacy issues, undesired algorithmic biases, or potentials for misuse.
Efforts like NeMo Guardrails, NVIDIA's newly released open-source software, showcase our dedication to responsible AI advancement. NeMo Guardrails empowers developers to guide generative AI applications in producing accurate, appropriate, and secure text responses. Additionally, NVIDIA has established an internal model risk management guidance, ensuring a comprehensive assessment and management of risks associated with AI models.

**Navigating the AI Frontier**

No discussion of AI would be complete without addressing what are often described as "frontier AI models."

When we refer to frontier AI models, we are describing the next-generation, gigantic-scale models that vastly exceed the capabilities of anything we have today. Frontier AI models are not models like current implementations of ChatGPT, Bard, Ernie, or Llama, which have been exhaustively explored by developers and millions of users, and which have well-known and understood capabilities and limitations.

Depending on their size and training strategies the developers use, frontier AI models may possess unexpected, difficult-to-detect new capabilities. As a result, we must be especially thoughtful in our approach to developing and deploying frontier AI models, ensuring that we do not unleash models before they are safe, accurate, reliable, and doing exactly what we want them to do.

Some pundits have expressed fear that frontier AI models will evolve into uncontrollable "artificial general intelligence" products that will escape our control and cause harm. Fortunately, uncontrollable artificial general intelligence is science fiction, not reality. While we may not be able to predict everything that a given AI model *will* do, the way models are built and connected limits what they *can* do.

At its core, AI is a software program, not a nuclear reactor. AI is limited by its training, the inputs provided to it, and the nature of its output. When we create an AI, we first decide what tasks we want our AI to perform, and we train it to perform those tasks.

For example, suppose we train an AI to answer questions – type in a query, and the AI will provide an answer. That AI is not trained to or connected in a way that enables it to drive cars, fly airplanes, or control the power grid. And even if the AI could answer questions

about how to do those tasks, the AI would have no ability to commandeer a car or an aircraft. The AI resides exactly where we put it, can do only what we train it to do, and can only affect what its outputs are connected to.

As a result, we humans will always decide how much decision-making power to cede to AI models. The AI models will never seize power by themselves. They will have only the power we give them. We can keep a human in the loop for any activity we wish.

We have no doubt that industry and government can work collaboratively to ensure that AI models are developed and deployed ethically and responsibly. The Administration's voluntary commitments, endorsed by NVIDIA and many of the companies engaged in frontier AI model development, are a solid start.

**Promoting Innovation and Democratizing AI**

While a few companies tend to dominate the AI headlines, we should recognize that the next great AI applications may come from anywhere in the world.

Today, thousands of startups worldwide are working diligently on foundational models and AI techniques, striving to create new tools and software that will benefit society in a myriad of ways. They work on GPU-powered systems deployed in workstations and the cloud. That work should be celebrated and encouraged and is well within the reach of today's regulatory and legal framework. These startups can and should be regulated as any other business.

So long as we are thoughtful and measured, we can ensure the safe, trustworthy, and ethical deployment of AI systems without suppressing innovation by researchers, academics, and enterprises working on new applications today. And we can spur innovation by ensuring that AI tools are widely available to everyone, not concentrated in the hands of a few powerful firms.

We encourage policymakers to work with industry to address concerns regarding frontier AI models while spurring startups, researchers, and enterprises to innovate to benefit society as rapidly as possible – which leads to a final observation.

Safe and trustworthy AI will require multilateral cooperation on norms and regulations, with the participation of every major power, or it will not be effective. This observation follows from two foundational truths.

First, the AI genie is already out of the bottle. The technology industry is international in scope, with many competitors worldwide. AI algorithms are widely published and available to all. AI software can be transmitted anywhere in the world at the press of a button, and many AI development tools, frameworks, and foundational models are open-sourced,

spurring companies in every country to compete to create new AI-enabled tools and services.

Second, no nation, and certainly no company, controls a "chokepoint" to AI development. Leading U.S. computing platforms are competing with companies from around the world. While U.S. companies may currently be the most energy-efficient, cost-efficient, and/or easiest to use, they are not the only viable alternatives for developers abroad. Other nations are developing AI systems, with or without U.S. components, and they will offer those applications in the worldwide market. To ensure safe and trustworthy AI, we must engage with governments and stakeholders around the world on norms and regulations to ensure responsible and trustworthy AI.

The United States is in a remarkable position today, and with your help, we can lead well into the future.

Thank you for the opportunity to testify before this Committee. NVIDIA stands ready to work with you to ensure that the development and deployment of generative AI and accelerated computing serve the best interests of all.

**Senator Josh Hawley**
**Questions for the Record**

**William Dally**
**Chief Scientist, NVIDIA**

1. **If Taiwan were invaded tomorrow, what would be the effect on NVIDIA's supply chains for semiconductors?**

   If NVIDIA's suppliers in Taiwan were disrupted or halted, NVIDIA would need to find alternative sources for a wide range of components and services.

2. **In your view, what steps should the United States take to ensure robust access to semiconductors?**

   The United States should ensure continued and secure access to semiconductor manufacturing around the world, including Taiwan, Korea, and the United States.

3. **What legislation or policies do you recommend to ensure that the United States can remain the leader in AI technology?**

   The United States should promote policies that ensure that every industry can benefit from AI, and that markets worldwide are open for U.S. firms to compete and lead.

# PREPARED TESTIMONY AND STATEMENT FOR THE RECORD
## OF

**WOODROW HARTZOG**
PROFESSOR OF LAW, BOSTON UNIVERSITY SCHOOL OF LAW
FELLOW, CORDELL INSTITUTE FOR POLICY IN MEDICINE & LAW, WASHINGTON UNIVERSITY IN ST. LOUIS

**on behalf of himself and based on research with scholars from the Cordell Institute for Policy in Medicine & Law, Washington University in St. Lous:**

NEIL RICHARDS
KOCH DISTINGUISHED PROFESSOR IN LAW AT WASHINGTON UNIVERSITY SCHOOL OF LAW AND CO-DIRECTOR OF THE CORDELL INSTITUTE

RYAN DURRIE
ASSOCIATE DIRECTOR FOR POLICY, CORDELL INSTITUTE

**September 12, 2023**

HEARING ON

## "OVERSIGHT OF A.I.: LEGISLATING ON ARTIFICIAL INTELLIGENCE"

BEFORE THE

U.S. SENATE COMMITTEE ON THE JUDICIARY
SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW

## I. INTRODUCTION

Chair Blumenthal, Ranking Member Hawley, and Members of the Committee, thank you for inviting me to appear before you and provide testimony on this important issue. My name is Woodrow Hartzog and I am a Professor of Law at Boston University School of Law and a Non-Resident Fellow at The Cordell Institute for Policy in Medicine & Law at Washington University in St. Louis. I have written extensively on privacy and law and technology issues, including how the unregulated amassing of personal data has fueled high-risk uses of artificial intelligence or "AI". My comments today will address what I've learned from this research. I make these comments in my personal, academic capacity. My remarks are based on research I have conducted in conjunction with scholars from the Cordell Institute for Policy in Medicine & Law at Washington University in St. Louis.[1]

I applaud the Committee for taking the initiative and advancing the conversation about legislating artificial intelligence. People and institutions are using AI systems to reshape how we work, play, socialize, and civically engage. These systems have proven useful, but they also impose an enormous cost on our wellbeing, our environment, and our democracy. So, we must make sure they are worth it for reasons beyond making things cheaper.

In my testimony today, I'll discuss how the bulk of industry-led AI policy approaches over the past several years, such as encouraging transparency, mitigating bias, promoting ethical principles, and giving people choice, are vital, but *they are not enough*, whether individually or collectively. In the end, they will not fully protect society, at the risk of giving the impression that our rules are sufficient and that lawmakers have done enough. In that sense these measures are best thought of as "AI half-measures."

I'd like to make one simple point in my testimony today: ***To bring AI within the rule of law, lawmakers must go beyond half measures to ensure that AI systems and the actors that deploy them are worthy of our trust***.[2] To do that, lawmakers must do three things.

---

[1] The Cordell Institute works to be an internationally recognized voice and resource in ethical health and human information policy, promoting the compelling societal good of advancing research and discovery into human wellness and disease while protecting the privacy and autonomy of patients, consumers, and citizens. We also thank Agnish Chakraburtty for his assistance with these comments.

[2] My Cordell Institute colleague Professor Neil Richards and I have argued that trust and relational vulnerability are the critical lenses through which to view issues of privacy, data protection, and civil rights in the digital age. We believe that issues of trust and vulnerability are equally critical for the design and implementation of AI systems that affect our minds, our health, our life opportunities, our finances, our labor, our markets, our environment, and even our democracy itself. *See, e.g.*, Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961; Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431(2016); Neil Richards & Woodrow Hartzog, *Privacy's Trust Gap: A Review*, 126 YALE L.J. 1180 (2017); Woodrow Hartzog & Neil Richards, *Trusting Big Data Research*, 66 DEPAUL L. REV. 579 (2017); Neil Richards & Woodrow Hartzog, *A Relational Turn for Data Protection?*, 6 EUR. DATA PROT. L. REV. 492 (2020); Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. 985 (2021); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022).

*First, they must accept that AI systems are not neutral.* This includes moving swiftly in holding developers of AI systems accountable for their design choices.

*Second, they must focus on substantive interventions that limit abuses of power.* Such approaches include imposing duties of loyalty, care, and confidentiality and *ex ante* approaches requiring a sound basis for processing data and deploying technologies.

*Third, they must resist the narrative that AI systems are inevitable* by creating strong bright-line rules for the development and deployment of AI systems. For the most dangerous designs and deployments, lawmakers should impose outright prohibitions.

## II. AI HALF-MEASURES ARE IMPORTANT, BUT INADEQUATE

In work with scholars at the Cordell Institute, we have argued that establishing trustworthy and accountable AI systems requires avoiding or at least going beyond "AI half-measures."[3] We have conceived of these half-measures as necessary but insufficient approaches that have a limited efficacy and in isolation may give lawmakers a false sense of security that they've done enough. They also risk tricking people into thinking new AI systems give little cause for worry.

Half measures often take the form of post-deployment controls, audits, assessments, certifications, and similar procedural compliance requirements.[4] These tools are necessary to begin the task of data governance, but industry has routinely leveraged procedural checks such as these to dilute data and consumer protection law into a managerial box-checking exercise that largely serves to entrench harmful surveillance-based business models.[5] A checklist is no match for the staggering fortune available to those who exploit our data, labor, and precarity to develop and deploy AI systems. And it's no substitute from meaningful liability for when AI systems harm the public.

The AI landscape is at a crossroads and now is the time to act. The harms of AI are real, significant, and becoming both entrenched and normalized by the day.[6] If we do not

---

[3] Neil M. Richards, Woodrow Hartzog & Jordan Francis, *Comments of the Cordell Institute on AI Accountability* (June 12, 2023), https://www.regulations.gov/comment/NTIA-2023-0005-1291 (Comment ID: NTIA-2023-0005-1291); *see also* Neil M. Richards, Woodrow Hartzog & Jordan Francis, *Comments of the Cordell Institute on the Prevalence of Commercial Surveillance and Data Security Practices that Harm Consumers* (Nov. 21, 2022), https://www.regulations.gov/comment/FTC-2022-0053-1071 (Comment ID: FTC-2022-0053-1071).

[4] *See* 88 Fed. Reg. 22,433, 22,435, *AI Accountability Policy Request for Comment* (Apr. 13, 2023) ("Governments around the world, and within the United States, are beginning to require accountability mechanisms including audits and assessments of AI systems").

[5] *See generally*, ARI EZRA WALDMAN, INDUSTRY UNBOUND: THE INSIDE STORY OF PRIVACY, DATA, AND CORPORATE POWER (2021).

[6] *See generally*, Grant Ferguson et al., *Generating Harms: Generative AI's Impact & Path's Forward*, ELECTRONIC PRIVACY INFORMATION CENTER (May 2023), https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf; Woodrow Hartzog, Evan Selinger & Johanna Gunawan, *Privacy Nicks: How the Law Normalizes Surveillance*, 101 WASH. U. LAW REV. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4384541.

impose rules to limit abuses of power, we risk eroding our civil liberties, our civil rights, and our democracy itself. The widespread adoption of AI systems implicates many of the rights and values we hold most dear: due process, freedom of expression, anti-discrimination, voting, free and fair elections, privacy, identity formation, the integrity of meaningful and intimate social relationships, and opportunities for safe, healthy, and fulfilling work.[7]

If lawmakers hope to foster trust and accountability in AI systems, they must avoid the seductive appeal of relying solely on "AI half-measures"—regulatory tools and mechanisms like transparency requirements, bias audits, ethical principles, and other procedural requirements that are necessary but not sufficient for true accountability. When implemented as standalone protections rather than as components of broader governance strategies, AI half-measures provide merely a veneer of accountability while failing to prevent or remedy the more serious harms that flow from deployment of untrustworthy AI systems. In so doing, a commitment solely to AI half-measures reveals itself as pernicious—offering the illusion of protection while enabling the festering of harms and other social costs. This might make AI half-measures appealing from an industry perspective but it definitely makes them dangerous for society. In my testimony today, I will identify four such AI half-measures which are necessary but not sufficient for true accountability: transparency, bias mitigation, ethics, and individualistic control.

### A. TRANSPARENCY DOES NOT AUTOMATICALLY MAKE THINGS RIGHT WHEN THINGS GO WRONG

At best, AI transparency can only be a first step, and not an end in itself. Certainly, AI systems themselves are astonishingly opaque at a scale "almost beyond human imagining."[8] As decision making concerning ordinary people increasingly relies on AI systems, we have a right to know the basis upon which those decisions are being made. People subject to AI system decision-making deserve "technological due process" that provides meaningful notice and transparency.[9] But transparency isn't always helpful to people without power. Even if you are aware a system is in use you might not be able to

---

[7] *See generally*, JULIE COHEN, CONFIGURING THE NETWORKED SELF (2012); DANIELLE CITRON, THE FIGHT FOR PRIVACY: PROTECTING DIGNITY, IDENTITY, AND LOVE IN THE DIGITAL AGE (2022); DANIEL SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE (2006); DANIEL SOLOVE, THE FUTURE OF REPUTATION: GOSSIP, RUMOR, AND PRIVACY ON THE INTERNET (2008); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); BRIAN CHRISTIAN, THE ALIGNMENT PROBLEM: MACHINE LEARNING AND HUMAN VALUES (2020); KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2016); FRANK PASQUALE, NEW LAWS OF ROBOTICS: DEFENDING HUMAN EXPERTISE IN THE AGE OF AI (2020); SAFIA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE (2023).

[8] *See generally*, Kate Crawford & Vladan Joler, *Anatomy of an AI System*, MOMA (2018) https://anatomyof.ai; *accord* KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021).

[9] *See generally*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008).

do much about it. It is critical, therefore, that lawmakers also have some level of insight into the design and implementation of these systems.

But looking into a system does not necessarily lead to knowing about it.[10] Nor does it produce accountability on its own. Transparency must work in tandem with due process and additional, substantive legal mechanisms that ensure people are not stripped of their rights.

Transparency is a pleasant-sounding word, but in practice it can mean very different things to different actors in the AI conversation. The importance of transparency in AI accountability, however, depends wholly upon what we mean by transparency and what it gets us—and particularly to what extent transparency furthers our ability to prevent and remedy harmful deployments of AI systems. To protect ourselves from the individual and social harms stemming from untrustworthy AI systems, we must do more than merely look into these systems; we must understand their various parts and change them when they threaten harm and betrayal.

**B. BIAS MITIGATION PROPOSALS ALONE ARE NOT ENOUGH.**

AI systems are notoriously and perhaps inevitably biased. A host of scholars have spent decades identifying the ways in which AI systems are biased against marginalized and underrepresented communities, most notably along the familiar lines of race, class, gender, and ability.[11] To mitigate bias, lawmakers, regulators, and those in industry call

---

[10] *See* Mike Ananny & Kate Crawford, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, 20 NEW MEDIA & SOCIETY 973, 977-982 (2018). Scholars Mike Annay and Kate Crawford have identified ten limitations of the transparency ideal: 1.It does not always follow that the ability to see inside a system results in the power to govern it. If there are not systems in place to process, digest, and use the information revealed to create change, or if the decision-makers are not vulnerable to public exposure, then transparency does not result in meaningful change. 2.Full transparency can be harmful, especially to vulnerable individuals. 3.Transparency can intentionally occlude by making so much information available that it conceals more damaging information. 4.Transparency can create false choices between complete secrecy and total openness if there is not a "nuanced understanding[] of the kind of accountability that visibility is designed to create." 5.Transparency burdens individuals by forcing them to "seek out information about a system, to interpret that information, and determine its significance." Similarly, the transparency ideal also presumes that different systems can easily be compared, allowing individuals to assess and choose between alternative options. 6.There is a dearth of empirical evidence that transparency engenders trust, either trust in organizations and systems or by the organizations and systems making disclosures. 7.Transparency is reliant on professionals who may have their own aims, such as "protecting the exclusivity of their expertise" or are subject to capture. 8.Transparency efforts can prevent deeper understanding of complex systems by focusing on merely seeing into those systems rather than interacting with them more deeply. 9.Technical limitations—resulting from the scale and speed of a system's design—can make a system inscrutable, even to its creators. This problem is especially challenging in the context of machine learning AI systems such as deep learning. 10.Temporal limitations (i.e., whether transparency should mean "future relevance, anticipated revelation, ongoing disclosure, or post hoc visibility") alter the efficacy of transparency obligations because visibility at different moments in an AI system's lifetime may "require or produce different kinds of system accountability."

[11] *See, e.g.,* IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE (2023); MEREDITH BROUSSARD, MORE THAN A GLITCH: CONFRONTING RACE, GENDER, AND

for more care in the development of AI systems.[12] I agree. AI systems will remain dangerously and possibly fatally flawed so long as they reflect harmful societal discriminatory practices. Bias mitigation is undoubtedly worthy of attention, resources, and regulation. However, while indispensable, by itself, de-biasing AI systems risks being a half measure in two ways.

First, what constitutes "fairness" in the context of AI systems is highly contested. It is easy to say that AI systems should not be biased; it is very difficult to find consensus on what that means and how to approach that goal. After all, "the very choice of a mathematical definition of 'fairness' is a political one."[13] Lawmakers must make that difficult decision and give the concept of fairness a clear and firm moral anchor. Leaving the definition of bias to the collective wisdom of self-interested individuals will lead to splintered and self-serving definitions that are likely to enable more bad practices than they prevent. Not to put too fine a point on it, but ***AI self-regulation will be doomed to fail***. Lawmakers should treat disparate impacts of AI systems as an issue of civil rights while also enforcing existing privacy and consumer protection laws.[14]

Second, when lawmakers and industry focus on bias-correction, they seem to assume away important threshold questions about whether AI has virtuous goals and uses, whether in a particular context or more generally. It is tempting to think a less biased AI

---

ABILITY BIAS IN TECH (2023); SAFIA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2021); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS (2015); KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021); Batya Friedman and Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFORMATION SYSTEMS 330 (July 1, 1996); Aylin Caliskan, Joanna J Bryson & Arvind Narayanan, *Semantics derived automatically from language corpora contain human-like biases*, SCIENCE (April 14, 2017) https://pubmed.ncbi.nlm.nih.gov/28408601/; Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Discrimination through Optimization: How Facebook's Ad Delivery can Lead to Biased Outcomes*, 3 PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION 1 (Nov. 7, 2019); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. (2018); Ngozi Okedigbe, *Discredited Data*, 107 CORNELL L. REV. 2007 (2022).

[12] *For example*, the Federal Trade Commission has emphasized that AI tools used for consumer lending be "empirically derived, demonstrably and statistically sound." Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 420 (2022) (*citing* Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N (Apr. 8, 2020), https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms (*quoting* 12 C.F.R. § 1002.2 (2018) (Regulation B))).

[13] Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 420 (2022) (*citing* Arvind Narayanan, *Tutorial: 21 Fairness Definitions and Their Politics*, YOUTUBE (Mar. 1, 2018), https://www.youtube.com/watch?v=jIXIuYdnyyk).

[14] *See generally* Solon Barocas and Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016); Andrew Selbst and Solon Barocas, *Unfair Artificial Intelligence: How FTC Intervention Can Overcome the Limitations of Discrimination Law*, 171 U. PENN. L. REV. 1023 (2023).

system is less dangerous. But accurate AI systems are, if anything, more dangerous.[15] Less-biased AI is more attractive to the powerful, who can abuse it. When industry uses bias mitigation to gloss over the other imbalanced dynamics of AI, they misrepresent the lessons from these important scholars teaching us that bias is the symptom of a larger problem about how power is amassed and wielded against marginalized communities.[16] To put this point simply, even if we ensure that AI works equally well for all communities, such an achievement will still create AI systems that can be used to dominate, damage, misinform, manipulate, and discriminate.

### C. ETHICAL PRINCIPLES ARE NOT ENOUGH WITHOUT ENFORCEMENT

Ethics should certainly be front of mind for companies that design or deploy AI systems. But ethical principles for AI become AI half-measures where these commitments to ethics are commitments in name only, or when they happen after the fact once AI tools have been built.[17] Ethical principles operate as a half measure if there is no way to hold companies accountable for failure to align with their espoused commitments, either because their statements are too vague or self-serving in substance or because ethics boards hold no decision-making or accountability power.[18] When industry does no more than adopt ethical principles, we get easy-to-make promises by companies to avoid the practices that aren't in their business model, but silence regarding the dubious tools that can make them money. We also get companies continuing to develop dubious AI systems that violate their own ethics principles. [19]

Professor Ryan Calo has noted that we are flooded with ethical principles from industry, the government, and even civil society, but this flood has not been accompanied by

---

[15] *See* Evan Selinger and Woodrow Hartzog, *What Happens When Employees Can Read Your Facial Expressions?*, NEW YORK TIMES (Oct. 17, 2019), https://www.nytimes.com/2019/10/17/opinion/facial-recognition-ban.html.

[16] *See generally,* Anita L. Allen, *Dismantling the "Black Opticon": Privacy, Race Equity, and Online Data-Protection Reform*, 132 YALE L.J. FORUM 907 (2022).

[17] In fact many companies have already given up on their own commitments. Gerrit De Vynck & Will Oremus, *As AI booms, tech firms are laying off their ethicists*, THE WASHINGTON POST (March 30, 3023), https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/.

[18] James Vincent, *The Problem with AI Ethics*, VERGE (Apr. 3, 2019, 10:47 AM) https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech ("Academic Ben Wagner says tech's enthusiasm for ethics paraphernalia is just 'ethics washing,' a strategy to avoid government regulation. When researchers uncover new ways for technology to harm marginalized groups or infringe on civil liberties, tech companies can point to their boards and charters and say, 'Look, we're doing something.' It deflects criticism, and because the boards lack any power, it means the companies don't change.").

[19] *Id.* (first citing George Joseph, *Inside the Video Surveillance Program IBM Built for Philippine Strongman Rodrigo Duterte*, INTERCEPT (Mar. 20, 2019, 9:35 AM), https://theintercept.com/2019/03/20/rodrigo-duterte-ibm-surveillance; then citing Shannon Liao, *Google Employees Aren't Convinced That Dragonfly Is Dead*, VERGE (Mar. 4, 2019, 1:30 PM), https://www.theverge.com/2019/3/4/18250285/google-dragonfly-censored-search-engine-code-dead-employees-doubt) (including alleged examples of unethical projects by companies that had publicly committed to AI ethics).

substantive legal rules.[20] Ethical principles are a poor substitute for laws and can even delay eventual rules because espousing principles and pointing to ethics committees can give the illusion of progress. It's easy to publicly commit to ethics, but industry doesn't have the incentive to leave money on the table for the good of society.[21] Ethics are important, but unless they occur throughout the development of AI systems rather than just at the end, and unless they are accompanied by substantive, external, legal constraints and sanctions, ethics are a half-measure.

**D. FOCUSING ON INDIVIDUALS IS JUST PART OF THE PICTURE**

One of the most popular approaches to legislating AI (as with privacy and big data before it) has been to seek to give individuals more control or ownership over their own data and creations. We can see this in new bills that require consent for data practices, give people rights over their data, and seek to give people intellectual property rights in their artistic creations and their names and likenesses. But when legislation becomes too focused on the individual at the expense of society, it can become a half-measure for two reasons. First, strategies to give people more control can backfire, because it's easy to overwhelm people with choices and delude them about what's really going on. Second, focusing on the collective wisdom of trillions of self-motivated, nudgeable, and possibly misinformed decisions isn't always what's best for society.

Thirty years of experience with modern privacy law makes one point clear – when it comes to complex technological systems, control over your informational destiny is not only an impossible illusion but it is one that can operate to make consumers do what the technology designers want them to do. Such commitments to control can have the opposite effect than what they are advertised as doing.[22]

Lawmakers who prioritize individual control, consent, and ownership in a vacuum risk missing how power and information are unequally distributed and deployed. Quite simply, the transparency and control contemplated by these frameworks is impossible in mediated environments. People can only click on the options provided to them and

---

[20] Ryan Calo, *Artificial Intelligence and the Carousel of Soft Law*, IEEE (Sept. 16, 2021), https://ieeexplore.ieee.org/document/9539878 ("[G]enerally speaking, there has been little change to the law and legal institutions in light of the supposedly transformative technology of our time. What we have instead are principles. We are awash in them. The industry has a running supply: Microsoft, Google, Facebook, IBM, and other companies have each released principles, despite joining with civil society to form an organization—the Partnership on AI—which has its own "tenets." The White House has AI principles. The Department of Defense has AI principles. So does the U.S. Chamber of Commerce. The UN, the WTO, and the OECD all have published AI principles. Even organizations with a track record of advocating for concrete legal reform have released or endorsed AI principles.").

[21] *Id.* ("The impulse of so many organizations across nearly every sector of society to promulgate principles in response to the ascendance of AI is understandable. Unlike law, which requires consensus and rigid process, an organization can develop and publish principles unilaterally…and while common principles can lay a foundation for societal change, they are no substitute for law and official policy … No invisible hand guides market participants to charity. The Internet is not Eden. Uber and Airbnb are not sharing with anyone. And AI is not a magical genie-in-training … The role of the law is to understand, channel, and address that change—with rules, not aspirations.").

[22] *See* NEIL RICHARDS, WHY PRIVACY MATTERS 90-100 (2022).

companies have incentive to design their products to nudge and manipulate people into doing what the designers want them to, whether through "dark patterns," hidden options, or "are you sure?" popups. Rules that prioritize individual control over data create incentives for companies to hide the risks of AI systems through manipulative design, vague abstractions, and complex or soothing words as they force us to accept those risks by designing systems where we never stop clicking the "I agree" button. This is, of course, assuming we had a choice about whether to use these systems in the first place. Lawmakers should not seek to give consumers dubious rights with which to try to protect themselves; lawmakers should seek to actually protect consumers regardless of what they choose.

Similarly, decisions about who owns the output of AI trained on or using other's data, works of art, or even likenesses are fundamental decisions for lawmakers. However, these decisions in and of themselves will not mitigate the effects of untrustworthy and unaccountable AI. Apportioning ownership of certain output of an AI that has already ingested and integrated a consumer's information or intellectual property may be able to redress some harms, but only to the extent that the infraction is identifiable and remediable by such an ownership transfer. Ownership is not the answer to the problem of accountability, trust, and equitable access to the powers of AI.

## III. HOW TO MOVE BEYOND HALF-MEASURES

As I have explained so far, a focus merely on concepts of (1) transparency and procedure, (2) bias mitigation, (3) ethical principles, and (4) individual control is insufficient when it comes to the design and implementation of accountable AI systems. This is especially true where dangerous, disruptive systems are being released on the world by for-profit companies with scant regard to the potential larger societal effects produced by these systems.

I recommend that Congress should take three steps to go beyond AI half-measures. First, lawmakers must accept that AI is not neutral and take the design of AI systems seriously. Second, Congress should focus on substantive interventions such as loyalty and other duties that limit abuses of power. Finally, Congress should resist the inevitability narrative by imposing bright-line rules and outright prohibitions where appropriate.

### A. ACCEPT THAT AI IS NOT NEUTRAL

As a first step in determining how AI should be regulated, I urge lawmakers to ***resist the idea that AI systems are simply neutral conduits.*** A common misconception about technologies is that they are value-neutral. People often argue that AI systems can be used for pro-social or anti-social ends, but the technology itself isn't inherently good or bad. In other words, "there are no bad AI systems, only bad AI system users." This mistaken view leads to the common refrain that we should regulate uses of technology, not the technology itself.

This view of technologies is wrong. There is nothing value-neutral about any information technology, including AI systems. Values are deeply embedded into the design of technology human beings build. Every technology sends signals to people and makes a certain task easier or harder. Facial recognition technologies make people easier to surveil, which gives power to the watcher. Social media reduces the cost of speech and gaining attention, empowering those with the incentive to scale misinformation and disinformation efforts. Generative AI systems reduce the cost of countless tasks, reducing the value of certain labor and rewarding those who avoided the cost of the labor, even where that labor is a skill-developing Sophomore English essay. Lawmakers must take the design of AI systems seriously, looking to established theories of accountability like defective design and providing the means and instrumentalities of unfair and deceptive conduct.[23]

**B. FOCUS ON SUBSTANTIVE INTERVENTIONS THAT LIMIT ABUSES OF POWER**

The prospect of omnibus AI legislation is daunting not only because it affects so many different areas of our life, but also because AI systems are so complex and powerful it can seem like trying to regulate magic.[24] But the broader risks and benefits of AI systems are not so new. AI systems bestow power. This power is used in all sorts of ways to benefit some and harm others. Some communities disproportionally benefit from that power, and other communities are marginalized and exploited by it. But I can offer some good news to the committee here: American law is no stranger to these power dynamics. As long as lawmakers keep inequalities and abuses of power and vulnerabilities to power at the center of their regulatory approach, they will be on the right track.

To that end, I recommend that since so many risks of AI systems come from within *relationships* where people are on the bad end of an information asymmetry, lawmakers should implement broad, non-negotiable duties of loyalty, care, and confidentiality as part of any broad attempt to hold those who build and deploy AI systems accountable. These duties would provide a substantive prohibition on self-dealing and harm to limit AI systems in ways that half measures would not. ***And duties of loyalty, care, and confidentiality are tried-and-true tools that American law has used to mitigate power imbalances in relationships for literally hundreds of years.***

A duty of loyalty to people who are made vulnerable to actors using AI systems would be a revolution in regulating technology as the most direct way to target business models

---

[23] *See generally* WOODROW HARTZOG, PRIVACY'S BLUEPRINT, THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES (2018); Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 MAR. L. REV. 785 (2015).

[24] *See* Efraín Foglia, Ferran Esteve, Lucía Lijtmaer, Luis Paadín, Óscar Marín, Miró Ramon & Mas Baucells, *Any sufficiently advanced technology is indistinguishable from magic*, CCCB LAB (Nov. 8, 2018), https://lab.cccb.org/en/arthur-c-clarke-any-sufficiently-advanced-technology-is-indistinguishable-from-magic/ (quoting Arthur Clarke's famous saying "Any sufficiently advanced technology is indistinguishable from magic."); Neil Richards, *Big data isn't magic*, ZOCALO PUBLIC SQUARE (Sept. 24, 2014), https://www.zocalopublicsquare.org/2014/09/24/will-we-have-any-privacy-after-the-big-data-revolution/ideas/up-for-discussion/#Neil+Richards.

that betray people's trust.[25] These duties can support our collective well-being by incorporating civil rights principles and a focus on the social good. Building a framework around loyalty could also be a powerful rallying cry to center our new AI legislative frameworks around an anti-betrayal rule. It's hard to get inspired by waves of bureaucratic jargon about impact assessments and the specifics of neural networks, but every person on earth unfortunately knows what it feels like to be betrayed. Duties like these are proven and effective.

Emboldened consumer protection rules, such as prohibitions on unfair, deceptive, and abusive acts or practices, can also help protect our trust. Lawmakers could implement substantive prohibitions like robust data minimization rules, limits on particular uses of data, and prohibitions on abusive design practices like dark patterns and predatory algorithms.[26] I join the groups Accountable Tech, AI Now, and the Electronic Privacy Information Center in their call for bright-line rules.[27] As part of the implementation of duties of loyalty, care and confidentiality, lawmakers should prohibit unacceptable AI practices like emotion recognition, unconstrained facial recognition, predictive policing, remote biometric identification in social spaces, social scoring, and fully automated hiring and firing. They should prohibit most secondary uses and third-party disclosure of personal data and while also requiring protections against third party access, including data-scraping. Woven together as a comprehensive regulatory fabric, these duties, rules, and commitments can invigorate and strengthen procedural tools such as audits and certifications, to the benefit of people both individually and as a group.

These duties could be crafted not only to protect our privacy, but also implemented in frameworks designed to protect our health and wellbeing, environment, and workplace. A

---

[25] *See generally* Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961; Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016); Neil Richards & Woodrow Hartzog, *Privacy's Trust Gap: A Review*, 126 YALE L.J. 1180 (2017); Woodrow Hartzog & Neil Richards, *Trusting Big Data Research*, 66 DEPAUL L. REV. 579 (2017); Neil Richards & Woodrow Hartzog, *A Relational Turn for Data Protection?*, 6 EUR. DATA PROT. L. REV. 492 (2020); Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. 985 (2021); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022).

[26] Computer science and policy scholarship demonstrates dark patterns' pervasiveness in consumers' digital lives. *See generally* Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt & Austin L. Toombs, *The Dark (Patterns) Side of UX Design*, EXTENDED ABSTRACTS OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (Apr. 20, 2018), https://pure.psu.edu/en/publications/the-dark-patterns-side-of-ux-design; Arunesh Mathur, Gunes Acar, Michael J Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty & Arvind Narayanan, *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION 1-32 (Nov. 7, 2019), https://dl.acm.org/doi/10.1145/3359183; Johanna Gunawan, David Choffnes, Woodrow Hartzog & Christo Wilson, *A Comparative Study of Dark Patterns Across Mobile and Web Modalities*, 5 PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION 1-29 (Oct. 18, 2021), https://dl.acm.org/doi/10.1145/3479521; Jamie Luguri and Lior Jacob Strahilevitz. 2021. Shining a Light on Dark Patterns. Journal of Legal Analysis 13, 1 (March 2021), 43–109; Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno & Franziska Roesner. *Exploring Deceptive Design Patterns in Voice Interfaces*, EUROUSEC '22: PROCEEDINGS OF THE 2022 EUROPEAN SYMPOSIUM ON USABLE SECURITY (Sept. 29, 2022), https://dl.acm.org/doi/10.1145/3549015.3554213.

[27] *Zero Trust AI Governance*, AI NOW INSTITUTE (Aug. 2023), https://ainowinstitute.org/wp-content/uploads/2023/08/Zero-Trust-AI-Governance.pdf.

new federal agency could coordinate these protective efforts across various legal frameworks. Licensing is one way to structurally ensure that an institution must respect human values and prove themselves worthy of our trust as a prerequisite to market entry. This *ex ante* approach could be a powerful correction to business models that encourage harmful behavior. It would require a sound basis for processing data and deploying technologies. A justification-first approach combined with substantive duties and strong liability rules would also flip the presumption that burdens society with the risk of dangerous systems by requiring companies to justify their systems by proving they will not harm us.[28]

Other structural rules would also help hold developers and deployers accountable while simultaneously benefitting all technology law and regulatory efforts. Lawmakers must protect researchers and whistleblowers. They should refine the Computer Fraud and Abuse Act and create an avenue for researchers to discover abuses of and within AI systems while preserving the trust of people exposed to those systems. They should expand public policy exceptions to NDAs for whistleblowers to report those abuses as in California and Washington's Silenced No More Acts. Governments should invest in their own expertise by ensuring a broad range of methodological experts are staffed in every office that deals with AI systems. Give the Federal Trade Commission, an agency already deeply invested in protecting the public from the abuses of AI systems, more staff and more funding and modify the prohibition on unfair and deceptive trade practices to include a prohibition on abusive trade practices (as at least one state has already done) and remove the cost-benefit requirement in Section 5 that fails to properly elevate human values and wellbeing.

### C. RESIST THE INEVITABILITY NARRATIVE

This brings me to my final point, which is that I encourage lawmakers to resist the inevitability narrative of AI. ***Technologies like AI systems are not inevitable*** – they are intentionally designed and built by people, and people (including many of the people in this room) can prohibit them, they can regulate them, and they can shape their evolution into socially-beneficial tools as well. Lawmakers will make little progress until they accept that the toothpaste is never out of the tube when it comes to questioning and curtailing the design and deployment of AI systems for the betterment of society.[29]

The inevitability narrative often gets woven in with the ideology of "innovation" and cashed out as the necessity of "progress." But by itself, progress is an empty word. Progress for whom? Of what? At what cost? We cannot avoid making choices about what kind of future we want with AI. Similarly, a quasi-religious invocation of citing "tech progress" is not an answer either, much less a panacea. Like "progress," "innovation" is a

---

[28] *See generally* Gianclaudio Malgieri & Frank Pasquale, *From Transparency to Justification: Toward Ex Ante Accountability for AI*, 712 BROOK. L. SCH. LEGAL STUD. (2022).

[29] Amba Kak & Sarah Myers West, *AI Now 2023 Landscape: Confronting Tech Power*, AI NOW INSTITUTE (April 11, 2023), https://ainowinstitute.org/2023-landscape ("[T]here is nothing about artificial intelligence that is inevitable. Only once we stop seeing AI as synonymous with progress can we establish popular control over the trajectory of these technologies and meaningfully confront their serious social, economic, and political impacts.").

buzz word that masks a thousand sins. My co-author Neil Richards has criticized tech company's repeated invocation of "innovation" as a canard.[30] The concept of innovation is selectively vague, meaning it can be whatever a tech company wants it to be, and to hear them tell it, innovation is always good and never bad.[31] It also has the strength of convenience—when advertising the latest product launch innovation seems like a supernatural force, but the moment regulation is proposed, innovation becomes easily "stifled," as fragile as a house of cards, toppled by the slightest regulation.

Perhaps because of this consistent mislabeling of innovation and progress, it is easy to simply assume the rightful existence of AI systems and go straight to building guardrails so they can flourish. Sometimes this works well. For example, AI systems have the potential to help scientists solve vexing problems.[32] But it's dangerous to always assume the virtues of astonishingly powerful AI systems. Professor Evan Selinger and I have argued that some AI systems like face surveillance technologies are too dangerous to ever be safely deployed.[33] I recognize there is room for debate on this topic, but my point here is that when lawmakers go straight to putting up guardrails, they fail to ask the existential question about whether particular AI systems should exist at all, and under what circumstances it should ever be developed or deployed. Even when tech companies initially resist the most dangerous tools like facial recognition, it seems unlikely that all participants in an unregulated industry can hold out forever when there is so much money on the table.[34]

Ignoring the existential question about AI systems dooms us to a framework of half measures. We've already seen evidence of half-hearted approaches to limiting the rampant abuse of facial recognition technologies. Lawmakers and industry first welcomed these systems by demanding transparency, implementing ethical principles, mitigating bias, and requiring consent. Meanwhile companies big and small scraped every bit of biometric and personal data they could get on the open web under the dubious claim it was all "publicly available" and the flimsy pretext that collecting jaw-dropping amounts of data on every person who uses the Internet was necessary for training the system. AI half-measures were no match for facial recognition systems, some of which were powered by the most astonishing and dangerous collection of data grabs I've ever heard of. These efforts have created untold implications for our privacy, our

---

[30] *See* NEIL RICHARDS, WHY PRIVACY MATTERS 177-183 (2021) ("[H]ere's an experiment: take any sentence from a technology company about "innovation," and replace the word "innovation" with "magic" to see if the meaning of the sentence changes at all. In my own experience playing this game many times over the past decade, it almost never changes the meaning.").

[31] *Id.* ("The rhetorical construction of "innovation" by the tech sector slices off everything bad and leaves only the gleaming stainless steel of a technological utopia, one that is all Thomas More and no George Orwell.").

[32] Robert F. Service, *'The game has changed.' AI triumphs at protein folding*, SCIENCE (Dec. 4, 2020), https://www.science.org/doi/10.1126/science.370.6521.1144.

[33] *See* Evan Selinger & Woodrow Hartzog, *What Happens When Employers Can Read Your Facial Expressions?*, N.Y. TIMES (Oct. 17, 2019), https://www.nytimes.com/2019/10/17/opinion/facial-recognition-ban.html; Evan Selinger & Woodrow Hartzog, *The Inconsentability of Facial Surveillance*, 66 LOY. L. REV. 101 (2019).

[34] *See* Kashmir Hill, *The Technology Facebook and Google Didn't Dare Release*, THE NEW YORK TIMES (Sept. 9, 2023), https://www.nytimes.com/2023/09/09/technology/google-facebook-facial-recognition.html.

environment, and our democracy. Bright-line rules that prohibit certain kinds of design and deployments of AI systems and certain kinds of information collection and use can make sure that the development and deployment of AI is justified, and not a power grab that imposes massive external costs on society.

## III. CONCLUSION

The shortcomings of AI half-measures should not be read to mean that transparency, bias mitigation, ethics, and personal control are not worthwhile measures to pursue. Again, this is not to say that procedural protections are wholly ineffective; rather, it is to say that procedural protections must be backed by substantive rules that promote human flourishing.

To properly rise to the challenge of AI technologies, we need a bold, comprehensive, substantive strategy of regulation rather than resignation. Lawmakers must resist the urge to simply stack procedural half-measures on top of one another and declare the AI problem solved. Transparency, bias mitigation, ethical principles, and individual control are merely band aids for the digital lacerations inflicted by unregulated and unconstrained AI systems. Lawmakers must wrestle with the difficult questions of determining what human values and goals constitute fair, trustworthy, and accountable AI. The procedural approach has been tried in the privacy context for the last twenty-five years, and it has been a spectacular failure.[35] Now that the stakes are even higher, it is time to try a different strategy that focuses on power and substance.

In my work with Professor Neil Richards concerning trust and loyalty in the context of privacy law, I have written about the need for "a substantive embrace of a broad array of human values over privacy law's reflexive deference to individual choice, consent, and control."[36] As AI systems are increasingly integrated into our lives, whether consumer-facing or operating in the background, we must flip the presumption that burdens society with the risk of AI systems and instead place the obligation on those who develop these tools, deploy them, and profit from them to serve society.

This does not mean that lawmakers should reflexively ban AI technologies *en masse*. Rather, lawmakers should restrict the implementation and design of AI systems where it is inappropriate and use substantive regulations like duties to encourage the design and implementation of AI systems in ways which embrace human values and promote human flourishing where it is appropriate.

Although artificial intelligence has existed in one form or another for decades, the present moment is notable for the degree to which AI has captured the public imagination. ChatGPT and other public-facing AI systems dominate the headlines, and the latest wave

---

[35] *See e.g.*, Neil M. Richards, Woodrow Hartzog & Jordan Francis, *Comments of the Cordell Institute on the Prevalence of Commercial Surveillance and Data Security Practices that Harm Consumers*, (Nov. 21, 2022) https://www.regulations.gov/comment/FTC-2022-0053-1071, (Comment ID: FTC-2022-0053-1071).
[36] Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C.L. REV. 1687 (2020).

of tech entrepreneurs are painting bold and imaginative visions of a future built upon AI. But we must be cautious, and we have good reason to be skeptical. AI systems also pose huge risks to almost every important aspect of our lives. They are already being used to amass power, harm vulnerable people, and erode social and political institutions. As AI systems kickstart a new phase of information systems revolution, we must avoid the mistakes of the past and proactively approach the difficult issues raised by these technologies. Trust and accountability can only exist where the law provides meaningful protections for humans. And AI half-measures will certainly not be enough.



**The Cordell Institute for Policy in Medicine & Law**
Washington University in St. Louis
Anheuser-Busch Hall, Room 541B
St. Louis, Missouri 63130

**October 13, 2023**

*Submitted via email to Record@judiciary-dem.senate.gov*

**Re: Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law hearing entitled "Oversight of A.I.: Legislating on Artificial Intelligence" – Senator Josh Hawley Questions for the Record**

**Dear Senator Hawley, Committee Members, and Staff,**

Thank you for the opportunity to testify at the Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law hearing entitled "Oversight of A.I.: Legislating on Artificial Intelligence" on Tuesday, September 12, 2023. Below is a written response to the question for the record from Senator Hawley. I file these comments on behalf of myself and in collaboration with my colleagues at the Cordell Institute at Washington University, Neil Richards and Ryan Durrie.

**Senator Hawley's Question for the Record:**

1. *What legislation or policies do you recommend to ensure that companies developing large language models do not exploit users' data?*

Large language models (LLMs) by their very nature are hungry for data,[1] leaky at holding this data,[2] sneaky in how they gather and use data,[3] and exclusory in their results and consequences for consumers.[4] Data exploitation and manipulation, whether in the context of creating and deploying large language models (LLMs) or otherwise, comes from an imbalance of power and information in relationships.[5] Tech companies can control what their customers see and they collect information from across the web

[1] Robert Hart, *Clearview AI Fined $9.4 Million In U.K. For Illegal Facial Recognition Database*, FORBES (May 23, 2022, 6:55am), https://www.forbes.com/sites/roberthart/2022/05/23/clearview-ai-fined-94-million-in-uk-for-illegal-facial-recognition-database/?sh=c9ef95019636; Alex Reisner, *These 183,000 Books Are Fueling The Biggest Fight In Publishing And Tech*, THE ATLANTIC (Sept. 25, 2023), https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/.
[2] Ben Derico, *ChatGPT bug leaked users' conversation histories*, BBC (Mar. 2023), https://www.bbc.com/news/technology-65047304; James Vincent, *Apple restricts employees from using ChatGPT over fear of data leaks*, THE VERGE (May 19, 2023, 3:29 AM), https://www.theverge.com/2023/5/19/23729619/apple-bans-chatgpt-openai-fears-data-leak.
[3] Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 MAR. L. REV. 785 (2015).
[4] *See* Neil Richards & Woodrow Hartzog, *Against Engagement* (conference draft , Privacy Law Scholars Conference 2022) (on file with authors).
[5] *See, e.g.* Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022).

as though all accessible human information is theirs for the taking. Compared to individual consumers, tech companies have practically unlimited resources and strong financial incentives to influence people's behavior for their own benefit and profit.

Lawmakers might be tempted to turn to standard data protection rules like forcing transparency and requiring consent from people before processing their data, but this standard approach would be doomed to fail. These "fair information processing" rules haven't yet held tech companies accountable and they aren't responsive to the full scope of risks posed by LLMs. To combat the fact that LLMs are hungry, leaky, sneaky, and exclusory, lawmakers should focus on consumer relationships with the companies that gather their data or design and implement LLMs based upon this data. This relationships-focused approach would be to how the law requires confidentiality, care, and most importantly, loyalty from physicians, lawyers, and other professionals with a power imbalance related to their patients, clients, and customers.[6]

Duties of loyalty protect against self- dealing, while related duties of care placed on relationships protect against dangerous behavior and the risks of harm. In other areas of the law, the extent of these duties is proportional to the vulnerability of the trusting parties.[7] The more exposed people are to LLMs, the more loyalty, care, and confidentiality lawmakers should demand from those deploying the tools.

Many of the problems of surveillance capitalism come down to the problem of self-dealing, where an organization exploits an advantage over a trusting party to its own benefit.[8] The lack of meaningful abilities to protect consumers under American privacy law has enabled such corporate opportunism and manipulation of consumers using human information, and this failure will only be exacerbated by the increased speed and efficiency of large language models to analyze and use this data. This problem is particularly serious in the context of LLMs and other technologies that promise to understand consumers so that they can better satisfy their needs and wants. Insufficiently constrained by privacy law and driven to maximize quarterly profits by corporate law, companies can deploy a potent cocktail of techniques derived from cognitive and behavioral science to "nudge" or otherwise influence the choices consumers make.[9] And history shows us that the companies that gather consumer data have not acted as benevolently as many had hoped.[10]

Misuse and self-enrichment through data gained in these power asymmetries ultimately costs consumers their time, money, attention, mental well-being, reputation, and significant life opportunities.[11] These costs include everything from their attention being broken via intrusive "notifications," to manipulation subtly shaping the way that consumers shop and vote, to the harms of engagement-driven social media.[12] "Personalization" of the companies' contacts and engagement strategies through the use of this personal

---

[6] *See* Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. Rev. 1687, 1693 (2020); Hartzog & Richards, *supra* note 5, at 1697.

[7] *See, e.g.*, Hartzog & Richards, *supra* note 5; Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. F. 11 (2020) at 13–14.

[8] *See, e.g.*, JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

[9] *See* NEIL RICHARDS, WHY PRIVACY MATTERS 39–50 (2022).

[10] *Id., See generally, Richard H. Thaler & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008) (outlining how companies can promote pro-human outcomes if given the incentives necessary to induce those outcomes)*.

[11] *See* Neil Richards & Woodrow Hartzog, *Against Engagement* (draft manuscript) (on file with authors).

[12] *See generally*, JOHANN HARI, STOLEN FOCUS: WHY YOU CAN'T PAY ATTENTION – AND HOW TO THINK DEEPLY AGAIN (2022); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022)?

data only magnifies these harms.[13] Such "personalization" can be finely calibrated to manipulate consumers into increasing engagement, regardless of any effect on consumers' mental wellbeing.[14] With every click and post, we are further exposed to the appetite, carelessness, and influence of those developing and deploying LLMs.

LLMs will never work for the benefit of all unless society can reliably trust those designing and deploying them. Right now, the trust people are giving these companies is a blind trust that is regularly betrayed. What is needed are rules that make companies deploying LLMs trust-*worthy*. This is where duties of loyalty, care, and confidentiality come in. The core feature of a duty of loyalty is that it creates a substantive duty prohibiting self-dealing at the expense of a trusting party.[15] Relational duties, such as a duty of loyalty, offer distinct advantages for lawmakers looking to address privacy across multiple disparate actors and methods of data consumption.

First, relational duties are sensitive to power disparities within information relationships. Second, relational duties help to mitigate the issues with overwhelming corporate disclosures and requests for consent. Relational duties allow allow lawmakers to move beyond ineffective consent frameworks while preserving meaningful choices for people. These duties allow trusting parties to enter information relationships without accepting the risks of whatever harmful data practices and consequences lurk in the fine print, the business model, or the technology.[16] They can also allow a broader range of potential choices because under a duty of loyalty, people are protected regardless of what they choose.[17] Relationships open the possibility of more robust enforcement rules because they are voluntarily entered into and because they are more consistent with free expression principles. This is why relational rules have long been recognized in American law.[18]

A substantive duty of data loyalty could revolutionize American privacy law. As we have argued in previous articles, comments, and testimony,[19] we believe that creating a broad duty of data loyalty offers three important advantages that other approaches do not. First, a duty of loyalty is substantially more able than a traditional data protection approach to address the novel problems created by the explosion of "big data" processing and analytics.[20] These include algorithmic discrimination, manipulation, oppression, and shaming that are caused by the ubiquity of modern technology platforms. Second, loyalty helps solve privacy law's harm problem in a way that is consistent with the direction of current Supreme Court

---

[13] This is precisely what happened in the Cambridge Analytica scandal, in which Facebook data was used to create finely calibrated psychological profiles of voters identifiedby their real names, suggesting which kinds of arguments would be most effective at getting them to act in the ways that the paying political advertisers wanted them to. *See* RICHARDS, *supra* note 9, at 25–26.

[14] These are the allegations Facebook whistleblower Frances Haugen presented under oath before lawmakers in the United States and around the world in 2021. *See, e.g.*, Billy Perrigo, *Inside Frances Haugen's Decision to Take on Facebook*, TIME (Nov. 22, 2021) https://time.com/6121931/frances-haugen-facebook-whistleblower-profile/ [https://per ma.cc/L8QN-6GD5].

[15] *See generally*, U.S. Sen. Comm. on Health Educ. Lab. & Pensions, Comment Letter on Improving Americans' Health Data Privacy (Sept. 28, 2023), (on file with The Cordell Institute for Policy in Medicine & Law); Hartzog & Richards, *supra* note 5.

[16] *Id.*

[17] For an extended critique of consent-based models for data processing, *see* Hartzog & Richards, *supra* note 5; *see also* Richards & Hartzog, *supra* note 11, and HARI, *supra* note 12.

[18] HARI *Supra* note 12.

[19] *See e.g., Neil Richards & Woodrow Hartzog, A Duty of Loyalty for Privacy Law, 99 WASH. U. L. REV. 961; U.S. Sen. Comm. on Health Educ. Lab. & Pensions, Comment Letter on Improving Americans' Health Data Privacy (Sept. 28, 2023), (on file with The Cordell Institute for Policy in Medicine & Law); Fed. Trade Comm'n, Comments of the Cordell Institute on the Prevalence of Commercial Surveillance and Data Security Practices that Harm Consumers (Nov. 22, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4284020; U.S. Senate Comm. on the Judiciary, Subcomm. on Priv. Tech. and the L., Testimony of Woodrow Hartzog on "Legislating of Artificial Intelligence" (Sept. 12, 2023), https://www.judiciary.senate.gov/imo/media/doc/2023-09-12_pm_-_testimony_-_hartzog.pdf; Nat'l Telecomm. and Info. Admin., Comments of the Cordell Institute on AI Accountability (June 12, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4477426.*

[20] *See generally* Woodrow Hartzog & Neil Richards, *Trusting Big Data Research*, 66 DEPAUL L. REV. 579 (2017)

doctrine. The exploitation of a relationship against a trusting party's interests, such as in a case of conflict of interest, can be a legally-cognizable concrete harm even if no other tangible harm manifests.[21] This is significant because American plaintiffs in privacy and data breach lawsuits have struggled to articulate harm that courts will recognize, particularly as the federal courts have tightened the rules for what constitutes a recognizable harm.[22] By contrast, because our common law duties of loyalty are literally older than the United States itself, they offer a tried and tested mechanism to resolve the power imbalances in relationships like those between doctors and patients and platforms and consumers.

A third benefit of a loyalty-based approach to privacy law is that loyalty duties have a long and established development in our law, most famously in the law of fiduciaries. A duty of data loyalty could draw heavily from this tradition and its proven ability to protect against the power imbalances in relationships in a fair, principled, and meaningful way. (We note that the professional ethics of both lawyers and doctors already require that they be loyal to their clients and patients; perhaps those of data scientists should as well.)[23] Loyalty, care, and confidentiality are not just foundational concepts in American law, they are also deeply intuitive. Lawmakers should not underestimate loyalty's rhetorical potential. A rallying cry requiring companies to "act in our best interests" could motivate American privacy reform in the way that "the right to be let alone" did at the turn of the twentieth century. Technocratic terms like "data minimization" and "legitimate interests of the data controller" do little for public imagination or comprehension. By contrast, loyalty is clear, it is easy to understand, and it is potentially robust enough to counterbalance spurious industry claims about the importance of "innovation" or the idea that commercial data processing implicates significant First Amendment issues. GDPR-style ideas like requiring companies to undergo data protection impact assessments can feel wonky and feeble, but every person in America likely knows how it feels to be betrayed.

There is also a roadmap for lawmakers looking to impose duties of loyalty on the powerful. Fiduciary law scholars have identified a tried and tested two-step process that lawmakers use to implement loyalty obligations in such a fair and just way.[24] Lawmakers first articulate a primary, general duty of loyalty—one that can be relatively permissive but acts as a residual backstop against betrayal. Second, courts and lawmakers go about the task of creating and refining what have been referred to as "subsidiary" duties that are more specific and sensitive to context. These subsidiary duties target the most opportunistic contexts for self-dealing and typically result in a mix of overlapping open-ended rules, maxims, more specific standards, and context-specific rules.

Thus, we propose that a duty of data loyalty should be implemented on two levels through what we have called the "loyalty two-step."[25] The first level is a broad and general "catch all" prohibition on substantial conflicts with the trusting party's best interests. This would prevent the most egregious forms of disloyalty across the board, and it would also serve to orient the company's incentives generally against betrayal

---

[21] *See* TAMAR FRANKEL, FIDUCIARY LAW 107–08 (2011) ("The duty of loyalty supports the main purpose of fiduciary law: to prohibit fiduciaries from misappropriating or misusing entrusted property or power. Thus, the duty of loyalty is manifested by important preventative rules. Such rules prohibit actions even though they are not necessarily injurious to entrustors."); see also *Spokeo v. Robbins* 136 S.Ct. 1540 (2016).
[22] See, e.g., *TransUnion LLC v. Ramirez,* 141 S.Ct. 2190 (2021); *Spokeo, Inc. v. Robins,* 136 S.Ct. 1540 (2016).
[23] Richards & Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961, 968 (2021).
[24] *See e.g.* Robert H. Sitkoff, *Other Fiduciary Duties: Implementing Loyalty and Care* in THE OXFORD HANDBOOK OF FIDUCIARY LAW 419 (Evan J. Criddle et al., 2019). ("The duties of loyalty and care, which we might call the primary fiduciary duties, are typically structured as broad, open-ended standards that speak generally. By contrast, the other fiduciary duties, which we might call the subsidiary or implementing fiduciary duties, are typically structured as rules or at least as more specific standards that speak with greater specificity.").
[25] *See* Hartzog & Richards, *supra* note 5.

rather than micromanaging specific instances. It would also supply a backstop against novel or innovative forms of betrayal that allows the law to evolve for new circumstances. The second level subsidiary duty of loyalty rules should be more specific and, where necessary, restrictive. This would involve the articulation of specific and substantive rules targeting particular contexts and actions that provide clearer rules than the general duty and would leave less wiggle room to ensure accountability. This clarity will keep the frameworks from becoming watered down. In the health care context, for example, bright-line rules should be more restrictive where companies are using personal health data for marketing or persuasion, or where they are collecting location data, but more permissive where personal health data is being used for biomedical research in the public interest. Through this layered approach, a duty of data loyalty could provide both general applicability as well as sensitivity to specific contexts.[26]

There is already bipartisan support for a duty of loyalty, including the proposed American Data Privacy and Protection Act (ADPPA).[27] However, the best starting point for statutory language is the proposed bipartisan Digital Consumer Protection Commission Act of 2023 (DCPCA) for online platform regulation.[28] The relevant language appears in Section 2411:

> "**SEC. 2411. DUTY OF LOYALTY.**
>
> A covered entity may not process personal data or design information technologies in a way that substantially conflicts with the best interests of a person with respect to—
>
> (1) the experience of the person when using a platform owned or controlled by the covered entity;
> or
> (2) the personal data of the person."[29]

An advantage of using this language as a starting point is that it builds on any bipartisan support that the DCDPA already has. If this Committee is further interested in how a duty of loyalty might be developed in legislation, we have attached our article *Legislating Data Loyalty*, appearing in the *Notre Dame Law Review Reflection*. And of course, we stand ready to help in any further ways that the Committee might find helpful as it tackles this complex and important set of policy issues.

In conclusion, we believe that a duty of loyalty provides the strongest, and most comprehensive protections against the misuse of consumer data by the creators and deployers of LLMs. This duty, implemented through the loyalty two-step described above could address the potentially hungry, leaky, sneaky, and exclusory effects of the unfettered proliferation of LLMs.

---

[26] *Id.*
[27] *See* Woodrow Hartzog & Neil Richards, *We're So Close to Getting Data Loyalty Right*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (June 14, 2022), https://iapp.org/news/a/were-so-close-to-getting-data-loyalty-right/.
[28] Digital Consumer Protection Commission Act of 2023, S. 2597, 118th Cong. (2023).
[29] *Id.* at § 2411.

Respectfully submitted,

Woodrow Hartzog, Professor of Law at Boston University and Cordell Institute Fellow [30]
Neil Richards, Koch Distinguished Professor of Law at Washington University and Cordell Institute Faculty Director[31]
Ryan Durrie, Cordell Institute Associate Director[32]

[30] Professor of Law, Boston University; Faculty Associate, Berkman Klein Center for Internet & Society at Harvard University; Affiliate Scholar, Stanford Law School Center for Internet & Society.
[31] Koch Distinguished Professor in Law and Director, Cordell Institute, Washington University in St. Louis; Affiliated Fellow, Yale Information Society Project; Faculty Associate, Berkman Klein Center for Internet & Society at Harvard University; Affiliate Scholar, Stanford Law School Center for Internet & Society.
[32] Associate Director for Policy, Cordell Institute, Washington University in St. Louis. We would also like to thank our Cordell Institute Research Fellow Agnish Chakraburtty for his assistance with this response.

76

# LEGISLATING DATA LOYALTY

*Woodrow Hartzog* & Neil Richards***

*Lawmakers looking to embolden privacy law have begun to consider imposing duties of loyalty on organizations trusted with people's data and online experiences. The idea behind loyalty is simple: organizations should not process data or design technologies that conflict with the best interests of trusting parties. But the logistics and implementation of data loyalty need to be developed if the concept is going to be capable of moving privacy law beyond its "notice and consent" roots to confront people's vulnerabilities in their relationship with powerful data collectors.*

*In this short Essay, we propose a model for legislating data loyalty. Our model takes advantage of loyalty's strengths—it is well-established in our law, it is flexible, and it can accommodate conflicting values. Our Essay also explains how data loyalty can embolden our existing data privacy rules, address emergent dangers, solve privacy's problems around consent and harm, and establish an antibetrayal ethos as America's privacy identity.*

*We propose that lawmakers use a two-step process to (1) articulate a primary, general duty of loyalty, then (2) articulate "subsidiary" duties that are more specific and sensitive to context. Subsidiary duties regarding collection, personalization, gatekeeping, persuasion, and mediation would target the most opportunistic contexts for self-dealing and result in flexible open-ended duties combined with highly specific rules. In this way, a duty of data loyalty is not just appealing in theory—it can be effectively implemented in practice just like the other duties of loyalty our law has recognized for hundreds of years. Loyalty is thus not only flexible, but it is capable of breathing life into America's historically tepid privacy frameworks.*

## INTRODUCTION

American privacy law is in a rut. It has no privacy identity. Its traditional rules mandating transparency and consent are outdated,

© 2022 Woodrow Hartzog & Neil Richards. Individuals and nonprofit institutions may reproduce and distribute copies of this Essay in any format at or below cost, for educational purposes, so long as each copy identifies the authors, provides a citation to the *Notre Dame Law Review Reflection*, and includes this provision in the copyright notice.

* Professor of Law and Computer Science, Northeastern University.

** Koch Distinguished Professor in Law and Director, Cordell Institute, Washington University in St. Louis. Portions and ideas in this Essay are adapted from and developed in much greater detail in the authors' other articles: *A Relational Turn for Data Protection?*, 6 EUR. DATA PROT. L. REV. 492 (2020); *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961 (2021); and *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. (forthcoming 2022).

The authors would like to thank Giuliana Green, Sara Hubaishi, Alexis Johnson, and Nina Sprenger for their research assistance. This research was supported by an award from the Notre Dame-IBM Tech Ethics Lab and by Notre Dame-IBM Tech Ethics Lab and by NSF award 1956393/1955227/1956435/2103439: "SaTC: Frontiers: Collaborative: Protecting Personal Data Flow on the Internet" as part of the ProperData project.

porous, and poorly enforced. It is a far cry from the "adequacy" necessary for a profitable and sustainable data trade with the European Union (EU) and Britain. It has, in short, proven no match for the likes of the modern tech giants and a world awash in data and devices. What's worse, while privacy reform appears to be on the agenda, many of the existing proposals—particularly those touted as "business-friendly"—are so weak as to risk codifying a privacy rights status quo that virtually everyone agrees is unacceptable.[1] In searching for a meaningful new approach to regulating data privacy, lawmakers have begun to seriously explore the idea that tech companies should be bound by a duty of loyalty to those who trust them with their data and online experiences.[2]

Scholars have proposed versions of a duty of loyalty for the past twenty years, but not all lawmakers are convinced.[3] Some may be

---

1 This is an argument we have been making for several years. *See e.g.*, Woodrow Hartzog & Neil Richards, Opinion, *There's a Lot to Like About the Senate Privacy Bill, if It's Not Watered Down*, THE HILL (Dec. 6, 2019), https://thehill.com/opinion/technology/472892-theres-a-lot-to-like-about-the-senate-privacy-bill-if-its-not-watered [https://perma.cc/W87Y-ZGPG]; Woodrow Hartzog & Neil Richards, Opinion, *It's Time to Try Something Different on Internet Privacy*, WASH. POST (Dec. 20, 2018), https://www.washingtonpost.com/opinions/its-time-to-try-something-different-on-internet-privacy/2018/12/20/bc1d71c0-0315-11e9-9122-82e98f91ee6f_story.html [https://perma.cc/W63X-UHCP].

2 *See, e.g.*, Data Care Act of 2019, S. 2961, 116th Cong. § 3(b)(2) (2019) (Duty of Loyalty—An online service provider may not use individual identifying data, or data derived from individual identifying data, in any way that—(A) will benefit the online service provider to the detriment of an end user; and (B) (i) will result in reasonably foreseeable and material physical or financial harm to an end user; or (ii) would be unexpected and highly offensive to a reasonable end user); Consumer Online Privacy Rights Act, S. 2968, 116th Cong. § 101 (2019) (Duty of Loyalty. (a) In General.—A covered entity shall not—(1) engage in a deceptive data practice or a harmful data practice; or (2) process or transfer covered data in a manner that violates any provision of this Act); New York Privacy Act, S. 5642, 2019–2020 Reg. Sess. (N.Y. 2019) ("Every legal entity, or any affiliate of such entity, and every controller and data broker, which collects, sells or licenses personal information of consumers, shall exercise the duty of care, loyalty and confidentiality expected of a fiduciary with respect to securing the personal data of a consumer against a privacy risk; and shall act in the best interests of the consumer, without regard to the interests of the entity, controller or data broker, in a manner expected by a reasonable consumer under the circumstances."); *Commission Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)*, COM (2020) 767 final (Nov. 25, 2020); Data Protection Act 2018, c. 123 (UK); An Act to Provide Facial Recognition Accountability and Comprehensive Enforcement, H.R. 117, 192d Gen. Ct., §2(a) (Mass. 2021) ("A covered entity shall be prohibited from taking any actions with respect to processing facial recognition data or designing facial recognition technologies that conflict with an end user's best interests.").

3 *See, e.g.*, Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. F. 11 (2020) [hereinafter Balkin, *The Fiduciary Model*]; Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2016) [hereinafter Balkin, *Information Fiduciaries*]; Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961 (2021) [hereinafter Richards & Hartzog, *Duty of Loyalty*]; Lauren Henry Scholz, *Fiduciary Boilerplate: Locating Fiduciary Relationships in Information Age Consumer Transactions*,

concerned that it is too vague, or that it would be bad for business. Others wonder what data loyalty would get us that we couldn't get from EU or California-style statutes. Others are uncertain about how a duty of loyalty would work and what specific legislation for data loyalty should look like.

In this short Essay, we propose a model for legislating data loyalty. Our model takes advantage of loyalty's strengths—it is well-established in our law, it is flexible, and it can accommodate conflicting values. Our Essay also explains how data loyalty can and should fit within the existing fabric of information privacy law, building on our research exploring how better privacy rules can protect and build trust in relationships between consumers and companies. It lays out the *what* and the *why* of data loyalty for legislators seeking a robust alternative to the failed "notice-and-choice" regime in the United States.

Our argument is simple—a duty of data loyalty is not just appealing in theory—it can be effectively implemented in practice just like the other duties of loyalty our law has recognized for hundreds of years. Loyalty is not only flexible, but it is capable of breathing life into America's historically tepid privacy efforts. It is a meaningful alternative to ineffective regimes that rely too much upon illusory notions of consent and restrictive notions of harm, while being flexible enough to confront new privacy challenges and accommodating mutually beneficial data practices. A properly implemented duty of loyalty could thus represent an answer to many of the problems of information privacy, creating real value for consumers, businesses, and our society as a whole.

---

46 J. CORP. L. 143 (2020); *see also* ARI EZRA WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE (2018); Claudia E. Haupt, *Platforms as Trustees: Information Fiduciaries and the Value of Analogy*, 134 HARV. L. REV. F. 34 (2020); Lilian Edwards, *The Problem with Privacy: A Modest Proposal*, 18 INT'L REV. L. COMPUTS. & TECH. 309 (2004); Christopher W. Savage, *Managing the Ambient Trust Commons: The Economics of Online Consumer Information Privacy*, 22 STAN. TECH. L. REV. 95 (2019); Jonathan Zittrain, *Engineering an Election*, 127 HARV. L. REV. F. 335, 340 (2014); Lindsey Barrett, *Confiding in Con Men: U.S. Privacy Law, the GDPR, and Information Fiduciaries*, 42 SEATTLE U. L. REV. 1057 (2019); Ariel Dobkin, *Information Fiduciaries in Practice: Data Privacy and User Expectations*, 33 BERKELEY TECH. L.J. 1, 1 (2018); Cameron F. Kerry, *Why Protecting Privacy Is a Losing Game Today—and How to Change the Game*, BROOKINGS (July 12, 2018), https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/ [https://perma.cc/L8DQ-SK79]; Ian R. Kerr, *The Legal Relationship Between Online Service Providers and Users*, 35 CANADIAN BUS. L.J. 419 (2001); Richard S. Whitt, *Old School Goes Online: Exploring Fiduciary Obligations of Loyalty and Care in the Digital Platforms Era*, 36 SANTA CLARA HIGH TECH. L.J. 75 (2020); Kiel Brennan-Marquez, *Fourth Amendment Fiduciaries*, 84 FORDHAM L. REV. 611, 612 (2015); DANIEL J. SOLOVE, THE DIGITAL PERSON (2004); *but see* Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497 (2019).

## I. WHAT IS DATA LOYALTY?

Data loyalty is the simple idea that the organizations we trust should not process our data or design their tools in ways that conflict with our best interests. It borrows from notions of loyalty in fiduciary law, but it is distinct from them. The model we propose here would be crafted by legislators to the specific vulnerabilities and incentives in the relationships between consumers and the data-extractive companies they deal with every day.

Scholars have proposed duties of loyalty in a variety of forms—including loyalty duties for data collectors, "information fiduciaries," or fiduciary boilerplate—in part because loyalty represents a substantive check on the ability of companies to use human data to nudge, influence, coerce, and amass vast profits from the exploitation of human information.[4] It cannot be avoided by trickery, hidden fine print, or manipulative interfaces known as "dark patterns." At its core, it protects the expectations consumers bring to relationships with companies, and it builds trust in those relationships that allows them to flourish to the benefit of both parties.

In other work we have articulated a duty for loyalty for privacy law as the duty of data collectors to act in the best interests of those whose data they collect.[5] A duty of loyalty for privacy law is neither perfect nor a tool for all tasks. But loyalty has one great virtue: it places the focus for information age problems on the relationships that define our social lives rather than on the data which is the byproduct of those relationships. Loyalty shifts the law's attention from the procedural rules of privacy law that are too easy to manipulate ("Did you hide a vague sentence in the privacy policy?" "Did the consumer fail to hit the tiny opt-out button?") to the substantive question of what practices go too far. It is flexible and adaptable across contexts, cultures, and times. Loyalty can thus be a powerful response to what Shoshana

---

4 *See, e.g.*, Balkin, *The Fiduciary Model*, *supra* note 3; Balkin, *Information Fiduciaries*, *supra* note 3; Richards & Hartzog, *Duty of Loyalty*, *supra* note 3; Scholz, *supra* note 3; *see also* WALDMAN, *supra* note 3; Haupt, *supra* note 3; Edwards, *supra* note 3; Savage, *supra* note 3; Zittrain, *supra* note 3, at 340; Barrett, *supra* note 3; Dobkin, *supra* note 3, at 1; Kerry, *supra* note 3; Kerr, *supra* note 3; Whitt, *supra* note 3; Brennan-Marquez, *supra* note 3, at 612; SOLOVE, *supra* note 3.

5 *See, e.g.*, Richards & Hartzog, *Duty of Loyalty*, *supra* note 3; Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. (forthcoming 2022); Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687 (2020) [hereinafter Hartzog & Richards, *Privacy's Constitutional Moment*]; Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016) [hereinafter Richards & Hartzog, *Taking Trust Seriously*]; Neil Richards & Woodrow Hartzog, *The Pathologies of Digital Consent*, 96 WASH. U. L. REV. 1461 (2019) [hereinafter Richards & Hartzog, *Pathologies of Digital Consent*]; Neil Richards & Woodrow Hartzog, *Privacy's Trust Gap: A Review*, 126 YALE L.J. 1180 (2017) (book review).

Zuboff calls "surveillance capitalism," the claiming of "human experience as free raw material for hidden commercial practices of extraction, prediction, and sales . . . As significant a threat to human nature in the twenty-first century as industrial capitalism was to the natural world in the nineteenth and twentieth."[6]

Data loyalty has three key features—it is a (1) *relational duty* (2) that *prohibits self-dealing* (3) at the *expense of a trusting party*. Let's break these three features apart.

### *A. A Relational Duty*

Lawmakers who decide they want to regulate privacy can begin their task by focusing on at least three different things. First, they could focus on the data itself, like what can be collected and whether datasets are deidentified. This is the approach that most federal and European privacy laws have taken to date with laws like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Second, there are structural concerns, like requiring companies to appoint a data privacy officer or focusing on monopoly power. This is the approach familiar to antitrust and corporate law. There's also a third option—lawmakers could focus on our relationships, like requiring confidentiality from physicians, lawyers, and other professionals.[7]

In addition to being one of the oldest contexts for privacy to flourish, relationships have a few distinct advantages for lawmakers looking to fight the excesses and abuses of data-hungry organizations. First, relational duties are acutely sensitive to the *power disparities* within information relationships. Tech companies control what we see, what we can click on, and what sorts of information they want to extract from their customers. They have incredible resources that help them predict and nudge our behavior and have the financial incentive to keep us ever more exposed. Duties of loyalty protect against self-dealing, while related duties of care placed on relationships protect against dangerous behavior and the risks of harm. The greater the power imbalance and the more vulnerable people are through exposure, so should the duty to which the trusted party is held be greater.[8]

Second, relational duties are a way out of *privacy's consent trap*. For years lawmakers, regulators, and companies have been obsessing over whether the consent people gave was a truly meaningful, informed, and revocable choice. People click "I Agree" buttons and slightly

---

6 SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER, at vii (2019).

7 *See* Hartzog & Richards, *Privacy's Constitutional Moment*, *supra* note 5, at 1697.

8 *See, e.g.*, Balkin, *The Fiduciary Model*, *supra* note 3, at 13–14.

wince without reading the terms because it is impossible to do so, even when what they click states that they read and understand the terms. Consent is broken, but lawmakers have stuck to notice and consent regimes anyway, even though it is common knowledge that digital consent is rarely meaningful. Relational duties allow for a decoupling of choice and consent. These duties allow trusting parties to enter into information relationships without accepting the risks of whatever harmful data practices and consequences lurk in the fine print, the business model, or the technology. They can also allow trusting parties to select from a range of choices without fear of betrayal because they would be protected no matter what they chose.[9]

Finally, relationships open the possibility of more *robust enforcement rules* because they are voluntarily entered into and hold a unique place in the law as a result. The concept of contractual privity could also be used to extend relational duties beyond the initial trusting party and entrustee. Under a "chain-link" approach to relational privacy rules, lawmakers could directly—or using mandated terms in data-sharing contracts—link the disclosure of personal information to obligations of loyalty to protect information as it is disclosed downstream.[10] To create the chain of protection, contracts would be used to link each new recipient of information to a previous recipient who wished to disclose the information. At the same time, relational duties raise even fewer free expression issues than other forms of data regulation because they regulate relationships rather than information flows. In relationships, parties assume these duties by soliciting trust and voluntarily entering into these relationships. Moreover, protections for power-imbalanced relationships have a deep tradition in U.S. law in harmony with free expression frameworks. This is, for example, why lawyers do not have a First Amendment right to disclose client confidences, no matter how "newsworthy" they might be.[11]

For these reasons, shifting the focus of privacy law from data to relationships offers significant advantages for effective policy.

---

9 For an extended critique of consent-based models for data processing, see Richards & Hartzog, *Pathologies of Digital Consent*, *supra* note 5.

10 *See* Woodrow Hartzog, *Chain-Link Confidentiality*, 46 GA. L. REV. 657, 658–61 (2012).

11 *Cf.* Balkin, *Information Fiduciaries*, *supra* note 3; Eugene Volokh, *Freedom of Speech and Information Privacy: The Troubling Implications of a Right to Stop People from Speaking About You*, 52 STAN. L. REV. 1049, 1057–58 (2000) (explaining that enforcement of contracts to maintain confidentiality create no First Amendment problems); Andrew J. McClurg, *Kiss and Tell: Protecting Intimate Relationship Privacy Through Implied Contracts of Confidentiality*, 74 U. CIN. L. REV. 887 (2006). *But see* Khan & Pozen, *supra* note 3.

### *B. That Prohibits Self-Dealing*

Many of the problems of surveillance capitalism come down to the problem of self-dealing, where an organization exploits an advantage over a trusting party to its own benefit.[12] The failures of American privacy law have enabled such corporate opportunism and manipulation of consumers using human information. This problem is particularly serious in the context of "personalized" technologies that promise to know us so that they can better satisfy our needs and wants. Insufficiently constrained by privacy law and driven to maximize quarterly profits by corporate law, companies can deploy a potent cocktail of techniques derived from cognitive and behavioral science to "nudge" or otherwise influence the choices we make.[13] These highly capitalized tech companies have not acted like the benevolent choice architects some had hoped for.[14] Technologies—and choice architecture—advertised as serving consumers have instead become weaponized, serving commodified consumers up to the companies and their commercial and political advertiser clients.[15]

Loyalty rules directly prohibit conflicted self-dealing. In so doing, they can change the incentives and business models of entire industries. Many critics believe that U.S. data privacy law has failed to change the corrosive business models that endanger, manipulate, mislead, misinform, and polarize people every day. The law, these critics suggest, merely prunes the edges of wrongdoing rather than getting to the core of the problem.[16] A duty of data loyalty would directly address this problem by taking self-dealing off the table as a general matter. More specific, subsidiary data loyalty rules for targeted advertising, web scraping, manipulative interfaces, and optimized human engagement metrics could revolutionize entire industries with clearer rules of the road. They could make certain abusive business models obsolete overnight. This would be a sharp contrast to the piecemeal and procedural approach of current U.S. data privacy law, which presupposes that all possible extraction models can be valid if they follow the right procedures and give people some semblance of control over their information. Data loyalty rules instead look directly to corporate profit motives and ask if they conflict with a trusting

12 *See* JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019); ZUBOFF, *supra* note 6.

13 *See* NEIL RICHARDS, WHY PRIVACY MATTERS 39–50 (2022).

14 *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 11–13 (2008).

15 *See* RICHARDS, *supra* note 13, at 46–49.

16 *See, e.g.*, Facebook, Inc., F.T.C. No. 1823109 (July 24, 2019) (Chopra, Comm'r, dissenting); Facebook, Inc., F.T.C. No. 1823109 (July 24, 2019) (Kelly, Comm'r, dissenting).

party's best interests. They require profit models to be based on the provision of valuable services rather than exploitation and extraction.

### *C. At the Expense of a Trusting Party*

Loyalty rules safeguard trusting parties from betrayal, looking to whether a trusting party has been disadvantaged by an organization's self-dealing. When organizations enrich themselves with trusting parties' data, people consistently end up paying with their time, attention, mental well-being, reputation, and significant life opportunities.[17] These costs include everything from notifications interrupting our attention to advance the interests of the platform, to manipulative advertising that causes people to buy (or vote) differently in ways that serve advertisers, to the well-documented emotional injuries wrought by engagement-driven social media. Crucially, these costs, impositions, and manipulations are made substantially more damaging by "personalization" enabled by self-dealing in personal data. Thus, it's not just a random notification or one serving your interests like a reminder to attend a meeting, but one teasing you out of the blue that someone you know may have done something cool. It's not just an ad or a political message, but one calculated to your precisely known psychology and vulnerabilities.[18] And it's not just social media telling you what your friends are doing, it's being done in a way that is calibrated to push your buttons to keep you scrolling (or doom-scrolling) with a reckless indifference to your mental health.[19]

The scope of protection that loyalty rules safeguard includes, but is broader than, recognized privacy harms like identity theft, emotional harms, breaches of confidence, and dangerous exposure.[20] It also includes more subtle individual and collective costs to our identity, our ability to create relationships, our collectively held truths, and the obscurity that protects our ability to share and move about freely. As

---

17 *See* Neil Richards & Woodrow Hartzog, *Against Engagement* (draft manuscript) (on file with authors).

18 This is precisely what happened in the Cambridge Analytica scandal, in which Facebook data was used to create finely calibrated psychological profiles of voters identified by their real names, suggesting which kinds of arguments would be most effective at getting them to act in the ways that the paying political advertisers wanted them to. *See* RICHARDS, *supra* note 13, at 25–26.

19 These are the allegations Facebook whistleblower Frances Haugen presented under oath before lawmakers in the United States and around the world in 2021. *See, e.g.*, Billy Perrigo, *Inside Frances Haugen's Decision to Take on Facebook*, TIME (Nov. 22, 2021) https://time.com/6121931/frances-haugen-facebook-whistleblower-profile/ [https://perma.cc/L8QN-6GD5].

20 *See, e.g.*, Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793 (2022); M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131 (2011).

such, it protects against the full range of betrayals that powerful parties in an information relationship can engage in.

* * *

Loyalty duties are thus quite straightforward when understood as relational duties that prevent self-dealing at the expense of a trusting party. They accord with basic notions of fairness and decency—if you have power over someone who trusts you, you shouldn't betray them or manipulate them to serve your own interests. It is undoubtedly for these reasons that our law has placed duties of loyalty on relationships with power imbalances for centuries in a wide variety of contexts.

## II. WHY DATA LOYALTY?

One common question that proposals for a duty of data loyalty often face is, "What does a duty of loyalty get you that other approaches to regulation do not?" This is an excellent question that asks why a duty of loyalty might be the right regulatory tool rather than some other approach. We believe that duties of data loyalty offer four important advantages that other approaches do not.

First, loyalty represents a central policy commitment that could be the missing ingredient to embolden existing U.S. privacy frameworks. Second, it is substantially more capable than a traditional data protection approach when it comes to modern privacy problems like algorithmic discrimination, manipulation, oppression, and shaming that are caused by unceasing digital contact and the astonishing scale and power of modern technology platforms. Third, loyalty helps solve privacy law's harm problem in a way that is consistent with the direction of current Supreme Court doctrine. Finally, data loyalty has a straightforward and strong rhetorical appeal; it offers a clear explanation for better privacy rules, it could help define America's privacy identity, and it could be used to gather broad popular support for stronger privacy rules.

### *A. To Embolden Existing Data Privacy Frameworks*

Law professor Ryan Calo is fond of saying that technology law's biggest problem is that we lack the political will to enforce the rules we already have.[21] We believe that this problem persists in privacy law as

21 *See, e.g.*, Ryan Calo, *Artificial Intelligence and the Carousel of Soft Law*, 2 IEEE TRANSACTIONS ON TECH. & SOC'Y 171, 171 (2021) ("But ultimately what is missing is not knowledge about the content of ethics as much as political will."); *Enlisting Big Data in the Fight Against Coronavirus: Hearing Before the S. Comm. on Com., Sci., & Transp.*, 116th Cong. (2020) (statement of Ryan Calo, Law Professor, University of Washington) ("It is also

well. Many privacy regulators lack the same political will and support from lawmakers and the executive branch to enforce existing data rules in a robust way. Many privacy rules are also vague, leaving their interpretation (and enforcement based upon that interpretation) up in the air. For example, what constitutes an "unfair trade practice," a "reasonable expectation of privacy," or the collection of "more data than is necessary" is a perennial topic of debate.

One of the reasons why U.S. data privacy frameworks tend to wilt is that they lack a clear touchstone to guide interpretation that would lead to effective enforcement. The collection of U.S. privacy statutes, enforcement actions, and common law remedies adhere to basic commitments like "do not lie," "do not harm," and "follow the Fair Information Practices (FIPS)."[22] But such edicts tend not to interrogate the wrongful motives of data processors and do little to force companies into any practice beyond bare compliance.

A duty of loyalty could change that. Lawmakers should use loyalty duties to embolden and revitalize *existing* approaches to regulating data privacy, such as robust implementation of data minimization requirements, rules against unfair and deceptive trade practices, and expansion of products liability theories of accountability. Data loyalty can empower lawmakers to use tools that have already been developed, by expanding the contexts in which rules should be followed, who must follow them, and the level of adherence necessary for compliance.

Take as an example data minimization, the idea that organizations should only collect, maintain, and use data that is necessary to fulfill a designated and legitimate purpose. Data minimization rules are a fundamental commitment of data protection and data security laws. They are scattered throughout U.S. law, including the California Consumer Privacy Act,[23] the Wiretap Act,[24] and are implicitly a part of

---

important to note that a lack of political will is sometimes the greater hurdle than a lack of information.").

22 *See, e.g.*, WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 15 (2018); Hartzog & Richards, *Privacy's Constitutional Moment*, *supra* note 5, at 1704 & n.66; Richards & Hartzog, *Duty of Loyalty*, *supra* note 3, at 42.

23 CAL. CIV. CODE § 1798.100(c) (2018) ("A business' collection, use, retention, and sharing of a consumer's personal information shall be reasonably necessary and proportionate to achieve the purposes for which the personal information was collected or processed, or for another disclosed purpose that is compatible with the context in which the personal information was collected, and not processed in a manner that is incompatible with those purposes.").

24 18 U.S.C. § 2518(5) (2018) ("No order entered under this section may authorize or approve the interception of any wire, oral, or electronic communication for any period longer than is necessary to achieve the objective of the authorization, nor in any event longer than thirty days. . . . Every order and extension thereof shall contain a provision that the authorization to intercept shall be executed as soon as practicable, shall be conducted in such a way as to minimize the interception of communications not otherwise subject to

the data security requirements of Section 5 of the FTC Act.[25] Data minimization, if robustly interpreted and enforced as a way for companies to remain loyal to trusting parties, could be a remarkably effective tool for regulators since it targets both collection and use of data and is meant to counter abusive purpose creep by companies.[26] If data loyalty became a guiding obligation for data minimization rules, it would give regulators and judges interpreting potential violations an additional layer of interrogation. Data loyalty would compel an examination of a company's motives *and* the potential adverse consequences to consumers in determining if more data than necessary was collected or if the use of data deviated too far from its original purpose. Such foundational support would prevent an arid and strictly textual analysis by explicitly forcing regulators and judges to look at the big picture of exploitative motives of organizations and the trusting parties' wellbeing.

Another example would be laws based on the Fair Information Practices, the most common standard for privacy laws worldwide. Under current U.S. privacy law, perhaps the most important question for regulators and compliance professionals is whether consumers have been given "notice and choice." In principle, this is a good thing, emphasizing consent to data practices and evoking the gold standard of "knowing and voluntary" consent familiar to lawyers and medical researchers. But in practice, under current American law, "notice and choice" all too often means just that consumers have merely vague "notice" of data practices that are buried in the fine print and illusory "choice" with respect to these practices such as a take-it-or-leave-it choice about whether to use the service.

In practice, such rules not only place few constraints on companies, but they also represent a kind of cookbook to create and justify even deeply disloyal data practices by checking the boxes of fictional notice and illusory consent. This is likely why companies like Amazon have been engaged in aggressive lobbying in many state capitols to get weak notice-and-choice (and only weak notice-and-

interception under this chapter, and must terminate upon attainment of the authorized objective, or in any event in thirty days.").

25 *See* FTC, COMMISSION STATEMENT MARKING THE FTC'S 50TH DATA SECURITY SETTLEMENT (2014) ("The Commission has also provided educational materials to industry and the public about reasonable data security practices. These materials explain that, while there is no single solution, such a program follows certain basic principles. . . . [Among them,] companies should limit the information they collect and retain based on their legitimate business needs so that needless storage of data does not create unnecessary risks of unauthorized access to the data.").

26 *See, e.g.*, DANIEL SOLOVE & WOODROW HARTZOG, BREACHED! WHY DATA SECURITY LAW FAILS AND HOW TO IMPROVE IT (forthcoming 2022).

choice) laws on the books.[27] But here, too, a duty of loyalty could help. If data loyalty became a guiding obligation for data processing, notice-and-choice requirements would become more than a checkbox compliance exercise at best and a cookbook for manipulation at worst. Instead, "notice" would become an obligation of honesty, ensuring that consumers actually understood what was happening with their data before they agreed to it, and preventing companies from all sorts of self-interested practices where meaningful understanding was not present. "Choice" would mean knowing and voluntary agreement to particular data practices among reasonable alternatives that do not conflict with a trusting party's best interests, rather than a "choice" about whether to live in the modern world or not.

In these ways, by reorienting the question for companies from "What can we get away with" to "Are we being loyal to our human customers," a duty of data loyalty could breathe new life into existing regimes that are moribund at best and exploitation-enabling at their worst.

### *B. To Address Emergent Dangers*

A second benefit of data loyalty is that it can safeguard consumers against novel and emerging digital risks. Data loyalty duties can *go beyond* the standard data processing concerns and traditional privacy harms. In crafting such rules, lawmakers should look to the ways in which the affordances of modern technologies endanger people by bestowing power in trusted entities. Data loyalty duties should scrutinize how those organizations have incentives to use the power human information gives them in self-interested ways that conflict with a trusting party's best interests. Duties crafted in this way would meaningfully respond to concerns about manipulative user interfaces (sometimes called "dark patterns"), the wrongful extraction of human labor by dominant platforms, algorithmic discrimination, and protection against third parties and other users while using a service. Duties of data loyalty can thus go beyond often hard-to-quantify injuries of individual pieces of data and address the structural power imbalances and inequalities that characterize the relationships between individual harried consumers and the richest corporations in the history of the world.

---

27 *See* Jeffrey Dastin, Chris Kirkham & Aditya Kalra, *Amazon Wages Secret War on Americans' Privacy, Documents Show*, REUTERS (Nov. 19, 2021), https://www.reuters.com/investigates/special-report/amazon-privacy-lobbying/ [https://perma.cc/LE8N-PBCM].

### *C. To Solve Privacy's Harm Problem*

Third, and related to the problem of emergent dangers, a duty of data loyalty would help lawmakers solve one of privacy law's most difficult problems: the problem of cognizable harm. Many privacy rules require some kind of economic, physical, emotional, or other kind of concrete and traditionally recognized harm to be legally cognizable. However, loyalty rules look to the trusted party's inequitable conduct of wrongfully exploiting an advantage gained by an information relationship. The exploitation of the relationship against a trusting party's interests can itself be the wrong, such as in a case of conflict of interest, even if no other tangible harm manifests.[28]

In privacy cases, this is significant because American plaintiffs in privacy and data breach lawsuits have struggled to articulate diffuse but real informational injuries, and this situation has been made worse in recent years as courts have tightened the rules for what counts as a legally cognizable injury under Article III standing doctrine.[29] Critically, loyalty duties do not have this problem—not just because the legal injury in loyalty cases is the disloyalty itself, but because this injury is one that has been already recognized by courts as legally sufficient within standing doctrine.[30]

The focus of loyalty is on the integrity of a relationship and removing an incentive and ability to wrongfully profit by taking advantage of a power disparity. Because loyalty duties are rooted in betrayal rather than harm or injury, they have significant consumer

---

28 *See* TAMAR FRANKEL, FIDUCIARY LAW 107–08 (2011) ("The duty of loyalty supports the main purpose of fiduciary law: to prohibit fiduciaries from misappropriating or misusing entrusted property or power. Thus, the duty of loyalty is manifested by important preventative rules. Such rules prohibit actions even though they are not necessarily injurious to entrustors.").

29 *See, e.g.*, TransUnion LLC v. Ramirez, 141 S. Ct. 2190, 2214 (2021); Spokeo, Inc. v. Robins, 578 U.S. 330, 342 (2016).

30 To get a bit technical for a moment, in *TransUnion*/*Spokeo* terms, then, a breach of a legally imposed duty of loyalty would be a "concrete" and "traditionally recognized" intangible harm. To satisfy this requirement, *Spokeo* requires courts "to consider whether an alleged intangible harm has a close relationship to a harm that has traditionally been regarded as providing a basis for a lawsuit in English or American courts." *Spokeo*, 578 U.S. at 341. *Ramirez* uses a slightly different formulation—asking whether an intangible injury bears "a close relationship to harms traditionally recognized as providing a basis for lawsuits in American courts." *Ramirez*, 141 S. Ct. at 2204. But because a breach of a duty of loyalty has been recognized as a basis for lawsuits for centuries, duties of loyalty simply do not raise concreteness problems. *See also* Citron & Solove, *supra* note 20. By contrast, although duties of care in general would be concrete, statutory causes of action rooted in novel theories of harm (including procedural data protection requirements) would seem to have to run through the *Spokeo* test, with an uncertain likelihood of success.

protection advantages over existing privacy rules that demand proof of injury.[31]

### *D. To Define America's Privacy Identity*

Finally, a duty of data loyalty could offer a defining value for America's privacy law identity, rather than forcing it to adopt a watered-down and sometimes ill-fitting version of the European GDPR approach. While American privacy law is weak, permissive, and seemingly rudderless, in Europe, privacy law is on firmer ground. Privacy and data protection are both considered fundamental human rights in the EU.[32] The GDPR is the manifestation of these rights, a commitment to the idea that people should be able to determine their informational fates for themselves. Bilyana Petkova has argued that data protection is "the main tenet of constitutional identity" in the EU.[33] This is why European data protection law often seems so strikingly powerful to American observers compared to domestic consumer privacy rights. As much as anything, then, for Europeans the GDPR is a state of mind. And it is why a U.S. version of the GDPR would inevitably be both a weak and inadequate version of the real GDPR, something we have elsewhere called a "GDPR-lite."[34]

A duty of loyalty could fill this definitional role for U.S. privacy law. It could supply a political lodestar for privacy reform that defines America's privacy identity on its own terms rather than those of the EU. Lawmakers should not underestimate loyalty's rhetorical potential. A rallying cry requiring companies to "act in our best interests" could motivate American privacy reform in the way that "the right to be let alone" did at the turn of the twentieth century. Technocratic terms like "data minimization" and "legitimate interests of the data controller" do little for public imagination or comprehension. By contrast, loyalty is clear, it is easy to understand, and it is potentially robust enough to counterbalance spurious industry claims about the importance of "innovation" or the idea that commercial data processing carries First Amendment value. GDPR-style ideas like requiring companies to undergo data protection impact assessments can feel wonky and feeble, but every person in America likely knows how it feels to be betrayed.

---

31 *See, e.g.*, H.R. 117, 192d Gen. Ct., §2(a) (Mass. 2021) ("A covered entity shall be prohibited from taking any actions with respect to processing facial recognition data or designing facial recognition technologies that conflict with an end user's best interests.").

32 *See, e.g.*, U.N. Convention for the Protection of Human Rights and Fundamental Freedoms art. 8, Nov. 4, 1950, 213 U.N.T.S. 221; Charter of Fundamental Rights of the European Union arts. 7–8, 2000 O.J. (C 364).

33 Bilyana Petkova, *Privacy as Europe's First Amendment*, 25 EUR. L.J. 140, 154 (2019).

34 For an extended version of an argument along these lines, see Hartzog & Richards, *Privacy's Constitutional Moment*, *supra* note 5, at 1727–32.

If companies owe us duties of loyalty, then "innovative" uses of data to exploit us start to resemble betrayal and fraud, and claims of First Amendment protection for manipulative uses of data look appropriately laughable. Loyalty also has the virtue of placing the obligation for ethical data processing right where it belongs, ensuring those to whom we expose our data vulnerabilities do not betray us. A duty of loyalty in privacy law would be important not just as a set of rules, but as an *idea* capable of rallying democratic support for strong rules.

Finally, loyalty can be good for business. At a U.S. Senate hearing in 2020, Senator Brian Schatz expressed the idea that duties of loyalty are only needed for bad businesses, because good businesses know that the best way to make money over the long term is to be loyal to their customers.[35] On the other hand, if disloyalty is permitted by the law, the pressures on business to show quarterly profits create strong short-term and short-sighted incentives to cheat and behave in disloyal ways. This in many respects is the story of the contemporary digital economy, a story that data loyalty offers the potential to change for the better.

## III. A MODEL FOR LEGISLATING DATA LOYALTY

One undeniable virtue of creating a duty of data loyalty is that it would not be necessary to invent it from whole cloth. Loyalty duties have a long and established pedigree in our law, most famously in the law of fiduciaries. A duty of data loyalty could draw heavily from this tradition and its proven ability to protect against the power imbalances in relationships in a fair, principled, and meaningful way.

Fiduciary law scholars have identified a two-step process lawmakers use to implement loyalty obligations in such a fair and just way.[36] Lawmakers initially articulate a primary, general duty of loyalty. Next, courts and lawmakers go about the task of creating and refining what have been referred to as "subsidiary" duties that are more specific and sensitive to context. These subsidiary duties target the most opportunistic contexts for self-dealing and typically result in a mix of overlapping open-ended rules, maxims, more specific standards, and highly specific rules.

---

35 *See Revisiting the Need for Federal Data Privacy Legislation: Hearing Before the S. Comm. on Com., Sci., & Transp.*, 116th Cong. (2020) (statement of Sen. Brian Schatz).

36 *See, e.g.*, Robert H. Sitkoff, *Other Fiduciary Duties: Implementing Loyalty and Care, in* THE OXFORD HANDBOOK OF FIDUCIARY LAW 419 (Evan J. Criddle, Paul B. Miller & Robert H. Sitkoff eds., 2019) ("The duties of loyalty and care, which we might call the *primary* fiduciary duties, are typically structured as broad, open-ended standards that speak generally. . . . By contrast, the *other* fiduciary duties, which we might call the *subsidiary* or *implementing* fiduciary duties, are typically structured as *rules* or at least as *more specific standards* that speak with greater specificity.").

Thus, we propose that a duty of data loyalty should be implemented on two levels. The first level is a broad and general prohibition on substantial conflicts with the trusting party's best interests. This would prevent the most egregious forms of disloyalty across the board, and it would also serve to orient the company's incentives generally against betrayal rather than micromanaging specific instances. The second level of a duty of loyalty would be more specific and, where necessary, restrictive. This would involve the articulation of specific and substantive subsidiary duties targeting particular contexts and actions that provide clear rules and less wiggle room to ensure accountability and keep the frameworks from becoming watered down. Though this two-step approach, a duty of data loyalty could provide both general applicability as well as sensitivity to individual contexts.

### *A. First, a General Catchall Duty*

We propose a general rule of data loyalty as follows:

*Organizations shall not process data or design systems and tools in ways that significantly conflict with trusting parties' best interests that are implicated by their exposure.*

Let's break this proposed duty down a little.

#### 1. A No-Conflict Rule for Data and Design

Organizations gain a power advantage over trusting parties in two different ways: collecting and processing data and controlling our mediated experiences.[37] If the duty of loyalty is to accomplish anything, it should prohibit the conflicted design of digital tools and data processing. Avoiding conflicts is loyalty's core mandate and the logical starting point for lawmakers, judges, industry, and civil society. A general rule against conflicted design and data processing could serve as the foundation for a host of regulatory regimes, self-regulatory efforts, and guidance to the public to encourage and nurture its trust.

A general no-conflict rule has the remarkable advantage of directing lawmakers (and trusted parties themselves) to interrogate not just actions but *motives* and *gains*.[38] Established fiduciary no-conflict rules

37 *See, e.g.*, RICHARDS, *supra* note 13; HARTZOG, *supra* note 22.

38 *See* Paul B. Miller & Andrew S. Gold, *Fiduciary Governance*, 57 WM. & MARY L. REV. 513, 557–58 (2015) (quoting Lionel Smith, *The Motive, Not the Deed*, *in* RATIONALIZING PROPERTY, EQUITY AND TRUSTS: ESSAYS IN HONOUR OF EDWARD BURN 53, 67 (Joshua Getzler ed., 2003) ("[T]he motives of the fiduciary are the crucial element in determining whether

> do not require the fiduciary to act in any particular way but are instead thought to establish boundaries within which the fiduciary may reasonably be expected to act loyally, at least to the extent that the rules isolate biasing factors that might induce the fiduciary to subjugate the interests of beneficiaries to the interests of others.[39]

## 2. People Over Profits

Some lawmakers are reluctant to adopt duties of data loyalty because they fear creating a conflict with the duties of loyalty that directors of organizations owe to shareholders.[40] This is an illusory conflict and, at most, is resolvable by lawmakers without substantially remaking corporate law.[41] The supposed conflict between trusting parties and shareholders has been wildly overstated.[42] Fiduciary law scholar Andrew Tuch explains that "imposing user-regarding

---

the fiduciary has acted loyally, and the requirement of motive is quite specific—the fiduciary 'must act (or not act) in what *he perceives to be* the best interests of the beneficiary.'")).

39 *Id.* at 557.

40 One of the most repeated critiques levied against the idea of imposing duties of data loyalty on companies is Lina Khan and David Pozen's claim that relational rules might create conflicting loyalties. The authors assert that "[t]he tension between what it would take to implement a fiduciary duty of loyalty to users, on the one hand, and these companies' economic incentives and duties to shareholders, on the other, is too deep to resolve without fundamental reform." Khan & Pozen, *supra* note 3, at 529, 534 ("[T]he information-fiduciary proposal could cure at most a small fraction of the problems associated with online platforms—and to the extent it does, only by undercutting directors' duties to shareholders, undermining foundational principles of fiduciary law, or both.").

41 *See, e.g.*, Balkin, *The Fiduciary Model*, *supra* note 3, at 23 ("Management's fiduciary obligations to shareholders *assume* that the corporation will attempt to comply with the legal duties owed to those affected by the corporation's business practices, even if this reduces shareholder value.").

42 *See* Andrew F. Tuch, *A General Defense of Information Fiduciaries*, 98 WASH. U. L. REV. 1897, 1902 (2021) (arguing that Khan and Pozen "significantly overstate the threat that corporate and fiduciary law poses for the information fiduciary model."). Tuch argues that corporate law only imposes duties on *directors*, not corporations, and the information fiduciaries proposal imposes duties on *corporations*, not directors. *See id.* at 1909. Relational duties would not create a set of inconsistent obligations among a *single* fiduciary. *See id.* at 1910. The issue of parallel fiduciary obligations owed by corporations as a whole to clients and directors to shareholders is routine. *See id.* Not only is the "likelihood of fiduciary breach that Khan and Pozen point to in claiming tension between Balkin's proposal and corporate law . . . theoretically remote," it is "in practical terms, nonexistent." *Id.* at 1915. Additionally, if lawmakers obligate a duty of loyalty, then directors are bound to privilege it over shareholder interests. *See id.* at 1916–17 ("Delaware law altogether avoids tension with regimes such as Balkin's. Delaware corporate law requires directors to exercise their discretion within legal limits imposed on the corporation; it does not license or excuse non-compliance with corporate obligations, even if directors believe that doing so would maximize shareholder value. And Delaware law offers no suggestion that a corporation's duties or responsibilities should be diluted or otherwise shaped by the content of directors' duties. Instead, case law indicates that directors must act 'within the law.'" (footnotes omitted)).

obligations on corporations will not create untenable frictions between duties to users and duties to shareholders. . . . [T]he primary criticism—that Delaware corporate law undermines the information fiduciary regime—should be dismissed."[43]

If lawmakers were to adopt data loyalty rules, then corporate law would in fact demand that directors adhere to them first and foremost.[44] In other words, the loyalty that directors owe to shareholders takes a backseat to legal obligations placed upon the corporation, including duties of loyalty to customers.[45] In fact, if a duty of data loyalty owed by platforms to people is made positive law, a director that acts with the intent to act in conflict with users' best interests or fails to act in the face of a known loyalty obligation may be liable for breach to shareholders of their fiduciary obligation as well as their duty to users.[46]

If data loyalty is going to work, then trusting parties must be prioritized over other loyalties owed by organizations, such as loyalty duties owed by firms to shareholders. Prioritizing trusting parties over shareholders would resolve any lingering "divided loyalty" concerns, as well as conflicting loyalties between users and third-party vendors. Self-interested actions would be allowed, but only if they didn't conflict with trusting parties' best interests regarding their data and mediated experiences. And of course, it is an elementary principle of U.S.

---

43 *Id.* at 1902 ("The criticism rests on a partial understanding of corporate law doctrine and theory. The criticism sees conflicting obligations where none exist and identifies strategies for resolving these apparent conflicts that are unknown to corporate law. . . . I also argue that Khan and Pozen's arguments are not merely mistaken but, if accepted, may do harm. Applying their case to financial conglomerates—more apt analogues for social media companies than the '[d]octors, lawyers, accountants, and the like' to whom scholars often draw their comparison—shows that Khan and Pozen's arguments, if accepted, would have pernicious effects on broad spheres of corporate regulation." (quoting Khan & Pozen, *supra* note 3, at 506)).

44 Tuch argues that "[u]nder the information fiduciary model, corporate law would require compliance with user-regarding obligations, creating incentives for directors to favor users' interests over those of shareholders." *Id.*

45 *Id.* at 1917–18 ("Reflecting corporate law's attitude toward legal compliance, former Harvard Law Dean Robert Clark identifies the corporation's purpose as to 'maximize the value of the company's shares, subject to the constraint that the corporation must meet all its legal obligations to others who are related to or affected by it.' . . . Even the most ardent advocates of shareholder primacy have not suggested that corporate law requires, or should require, corporations or directors to maximize shareholder value in violation of a corporation's legal obligations." (quoting ROBERT CHARLES CLARK, CORPORATE LAW 17–18 (1986))).

46 *See id.* at 1918–19 (citing *In re* Walt Disney Co. Derivative Litig., 906 A.2d 27, 67 (Del. 2006); *see also* Stone v. Ritter, 911 A.2d 362, 369–70 (Del. 2006) ("The failure to act in good faith may result in liability [for directors] because the requirement to act in good faith 'is a subsidiary element[,]' i.e., a condition, 'of the fundamental duty of loyalty.'") (citing Guttman v. Huang, 823 A.2d 492, 506 n.34 (Del. Ch. 2003)).

constitutional law that a federal duty of loyalty would take precedence over any state duties by operation of the Supremacy Clause.

Data loyalty would still allow companies to profit and flourish. The "best interests" polestar of loyalty, by design, accommodates all kinds of self-serving behavior. It simply makes self-serving behavior allowable only in instances where it aligns with the best interests of the primary trusting party.[47]

3. The Collective Best Interests of Trusting Parties

There are a few different ways to deal with inevitable conflicts between trusting parties as well. The first would be to impose a reasonableness and fairness approach, or a duty of impartiality between people who expose themselves to organizations.[48] In trying to accommodate the best interests of billions of individuals, whose "best interests" might differ from person to person, lawmakers could also follow tort law's move to a more objective standard: the reasonable user. Not only would a reasonable user standard help companies better determine the scope of their duties, but it would also inject a normative element into the analysis.

Our proposal adopts a collective approach to "best interests," to better avoid conflicts between trusting parties and help free privacy law from its overly individualistic focus. Allowing lawmakers and regulators to focus on the collective best interests of "trusting parties," they can better respond to systemic harms detected sporadically by individuals but strongly at the group level. We recommend that lawmakers specifically prioritize interests that are held collectively by groups of users, with certain individually held interests holding sway only to the extent they do not conflict with collective user interests.[49]

A more collective best interests approach would be an improvement over the individual self-determination model, which does not compel people to consider the common good or threats to

47 *See* John H. Langbein, *Questioning the Trust Law Duty of Loyalty: Sole Interest or Best Interest?*, 114 YALE L.J. 929, 932 (2005) ("[A] transaction prudently undertaken to advance the best interest of the beneficiaries best serves the purpose of the duty of loyalty, even if the trustee also does or might derive some benefit. A transaction in which there has been conflict or overlap of interest should be sustained if the trustee can prove that the transaction was prudently undertaken in the best interest of the beneficiaries.").

48 Andrew S. Gold, *The Fiduciary Duty of Loyalty*, *in* THE OXFORD HANDBOOK OF FIDUCIARY LAW, *supra* note 36, at 398 ("[C]onflicts among best interests obligations [owed to multiple beneficiaries] are unavoidable. Where such conflicts exist, one answer is to find that loyalty must manifest itself as fairness and reasonableness. Another answer is to impose a duty of impartiality," which would demand "due regard" (though not necessarily equality). (footnotes omitted)).

49 *Id.* (discussing the hierarchy of obligations approach to how "common shares might ordinarily benefit from fiduciary obligations while preferred shares will only benefit in exceptional [circumstances]").

groups they are not a part of. When people give consent to data practices, they usually aren't motivated to reflect upon how their decision will affect vulnerable groups that they are not a part of.[50] This is similar to some people's indifference to public health when they "choose" not to wear a mask during a pandemic.

A reasonable user approach would also be consistent with the parallel duty of care and sensitive to the fact that tech companies deal in bulk and batched relationships. A reasonableness, context-sensitive approach would require loyalty obligations that are proportional to risk of abuse. The duty would be the most robust where the volume of data collected is highest and organization's power over people is the greatest. Because this duty of loyalty would be new and novel for privacy law and would need to be tailored to the unique characteristics of modern information relationships, lawmakers have the ability to craft a unique and fitting approach that borrows from how duties of loyalty operate in other contexts without being bound by it.

4. Limited to Trusting Parties' Exposure

In our previous work on trust, we defined the concept of trust as the willingness to make oneself vulnerable to the actions of others.[51] Our proposed general duty of loyalty would be limited to the extent of that vulnerability. Specifically, the "best interests" should be limited to the interests affected by the entrustment of data, labor, and attention, instead of an overall well-being standard. Organizations would be directed to ask what interests were implicated by the affordances of the data and design of user interfaces. So while it might be disloyal for a company to design a system that leveraged trusting parties' geolocation to allow pharmaceutical companies to target people when they are currently in the hospital (and thus vulnerable), it would probably not be disloyal for that company to generally allow pharmaceutical companies to place advertisements on their app or website. Systems that allow for such microtargeted advertising based on highly detailed profiles rather than isolated contexts make exploitation of vulnerable parties easier and compound incentives for companies to engineer exposure for financial gains.[52]

In conjunction with a duty of care, a duty of loyalty animates a number of different broad subsidiary duties, such as duties of candor,

50 *See, e.g.*, Richards & Hartzog, *Pathologies of Digital Consent*, *supra* note 5, at 1498; Evan Selinger & Woodrow Hartzog, *The Inconsentability of Facial Surveillance*, 66 LOY. L. REV. 33, 44 (2020).

51 Richards & Hartzog, *Taking Trust Seriously*, *supra* note 5, at 448.

52 *See, e.g.*, JOSEPH JEROME & ARIEL FOX JOHNSON, ADTECH AND KIDS: BEHAVIORAL ADS NEED A TIME OUT (2021).

good faith, nondelegation of key services, and confidentiality.[53] But legislatures and courts often go further and create or delegate authority for the creation of a series of clearer subsidiary obligations that are more like rules than vague standards.[54]

This two-tiered approach allows lawmakers to tailor rules to specific relationships to avoid specific foreseeable conduct while maintaining flexibility for new and changed rules in the future.[55] As applied to privacy law, it would allow lawmakers to target large platforms or social media companies that presented specific problems of gatekeeping for third parties or self-dealing due to two-way markets without applying the same specific rules to traditional e-commerce or media streaming companies bound by a general duty of loyalty. Companies not bound by specific subsidiary rules would still be bound by a general duty of loyalty.

### *B. Second, Rules for Subsidiary Implementing Duties*

Lawmakers can create specific subsidiary rules to help resolve objections that a duty of data loyalty is just too vague.[56] Enacting

---

53 *See* Whitt, *supra* note 3, at 94–95 ("Additional fiduciary obligations recognized by courts of equity over many centuries include the duty of candor, duty of good faith, duty not to delegate the services to others, and the duty of confidentiality. Typically they are subsumed as 'subsidiary' or 'implementing' obligations under either the duty of care or of loyalty. However, in some legal quarters the duty of confidentiality has been deemed an important supportive component of the 'primary' fiduciary duties. . . . [T]he duty of confidentiality deserves special status in the digital environment as an 'enabling' obligation that strengthens the more well-established fiduciary duties of care and of loyalty." (footnotes omitted)).

54 Robert Sitkoff explains that

> [t]he duties of loyalty and care, which we might call the *primary* fiduciary duties, are typically structured as broad, open-ended standards that speak generally. . . . By contrast, the *other* fiduciary duties, which we might call the *subsidiary* or *implementing* fiduciary duties, are typically structured as *rules* or at least as *more specific standards* that speak with greater specificity.

Robert H. Sitkoff, *Other Fiduciary Duties: Implementing Loyalty and Care*, *in* THE OXFORD HANDBOOK OF FIDUCIARY LAW, *supra* note 36, at 419.

55 Sitkoff gives the prudent investor rule as an example of a how subsidiary rules develop in trust law. *Id.* at 421 ("Structurally the prudent investor rule is an elaborated standard that, by focusing on risk-and-return and diversification, gives specific content to the open-ended, primary duty of care, called prudence in trust parlance, as applied to the investment function of trusteeship . . . . [W]ithin the fiduciary fields that do include an investment function, the prudent investor rule encompasses the accumulated learning on what the duty of care requires in fiduciary investment. In consequence, rather than start from scratch in every fiduciary investment matter, fiduciaries, beneficiaries, and courts may look to the elaboration within the prudent investor rule to discern the application of the duty of care.").

56 In a hearing on the future of transatlantic data flows called by the U.S. Senate Committee on Commerce, Science, & Transportation, Senator Wicker asked of a panelist who advocated for a duty of loyalty in privacy law, "[w]here is there a working duty of loyalty

legislation should also either provide for subsidiary duties or delegate rulemaking authority for future subsidiary rules. These subsidiary data loyalty rules might take a page from and model information privacy versions of nonprivacy fiduciary duties such as disclosure, consent, accounting for property (access and portability rights), confidentiality, and the full suite of fair information practice principles. This would apply some of the most significant obligations compelled by the GDPR. A duty of loyalty, combined with a duty of care, could spur on specific rulemaking for concepts like data minimization and legitimate basis requirements that would be bound together by an antibetrayal ethos.

But lawmakers need not stop there. One of the most important subsidiary duties to stem opportunistic behavior would be a robust prohibition on abusive trade practices. As we detailed in prior work, companies leveraging people's own cognitive and resource limitations against them to wrongfully extract data and labor is an endemic problem online.

Subsidiary rules prohibiting abusive trade practices would prohibit entrustees from materially interfering with the ability of trusting parties to understand the terms of the relationship and the risk associated with exposure and engagement.[57] Rules against abuse would also prohibit entrustees from taking unreasonable advantage of trusting parties' lack of understanding about the material risks, costs, or conditions of the entrustees' service or the inability of trusting

---

in place in law somewhere that we can look to? When we're able to be specific in those instances, then we're getting somewhere. But beyond that, it's hard actually to define [a duty of loyalty]." *The Invalidation of the EU-US Privacy Shield and the Future of Transatlantic Data Flows: Hearing Before the S. Comm. on Com., Sci., & Transp.*, 116th Cong. (2020) (statement of Sen. Roger Wicker, Chairman, S. Comm. on Com., Sci., & Transp.). Senator Wicker is the sponsor of one of the most prominent proposals for an omnibus federal privacy law in the United States. The Senator actually expressed tentative support for a duty of loyalty, even though such a duty does not explicitly appear in the bill he sponsored. And in full disclosure, the panelist was one of the authors of this Essay. Also, thank you for reading so deeply in our paper—and in its footnotes. *See also* James Grimmelmann, *When All You Have Is a Fiduciary*, LAW & POL. ECON. PROJECT (May 30, 2019), https://lpeproject.org/blog/when-all-you-have-is-a-fiduciary [https://perma.cc/V5PB-4D6B] (arguing that when applied to digital platforms "the rule against self-dealing is either absurdly under-inclusive, absurdly over-inclusive, or both").

57 We propose that lawmakers adapt language from the Consumer Financial Protection Bureau's authority to regulate abusive trade practices along these lines: "Abusive trade practice" means any conduct by a covered entity that 1) materially interferes with the ability of a trusting party to understand a term or condition of the agreement between covered entities and trusting party relating to the processing of personal data or effect or functionality of a system, tool, or user interface deployed by the covered entity; or 2) takes unreasonable advantage of: a) a lack of understanding on the part of the trusting party of the material risks, costs, or conditions of the covered entity's product or service; or b) the inability of the trusting parties to protect their interests in selecting or using a covered entity's product or service; or c) the reasonable reliance by the trusting party on a covered entity's representation to act in the interests of the trusting party.

parties to protect their interests within the relationship. Finally, anti-abuse rules would prohibit entrustees from taking unreasonable advantage of the reasonable reliance by trusting parties on entrustees' representations to act in the trusting parties' interests.

Lawmakers might also consider rigid prohibitions on specific practices like the deployment of unreasonably dangerous automated tools or the use of personal data to train those automated systems. They could create subsidiary rules for inherently dangerous practices and technologies that, at the systemic level, are in fundamental conflict with the best interests of trusting parties, such as microtargeting, a practice that paves the path for third party abuse and imposes more externalities than benefits for trusting parties; and affect recognition, a fundamentally misguided, mistaken, and oppressive tool.[58] Lawmakers could craft even more rules designed for specific parties such as "social media platforms may not deploy affect recognition technologies on photos or videos submitted by trusting parties." There might also be disclosure mandates, process requirements, prohibitions on conduct, or obligated tasks. Each rule should target specific areas where trusted parties have an incentive to engage in self-dealing.[59]

Lawmakers could, of course, impose all these rules even without couching them within an umbrella duty of loyalty. We have proposed in previous research that trust-building and trust-enforcing rules irrespective of a relationship between the parties could be meaningful complements or the next best thing to broad and strong relational obligations.[60] Many of these rules, such as data protection obligations, should have sibling rules that apply regardless of whether data controllers are in an information relationship with a trusting party. But we believe that a duty of loyalty would act as an important animating force, interpretive guide, and catchall provision that would bring more coherence, flexibility, and accountability through enforcement than these rules would have as stand-alone laws.

Nonetheless, we propose specific subsidiary rules within information relationships to maximize the advantages of a relational approach to privacy. Scholars and lawmakers have identified different contexts where the incentives for self-dealing by the powerful party in

---

58 For an exploration on the dangers of affect recognition systems see, e.g., KATE CRAWFORD, ATLAS OF AI (2021); Kate Crawford, *Artificial Intelligence Is Misreading Human Emotion*, THE ATLANTIC (Apr. 27, 2021), https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/ [https://perma.cc/T6AV-J25T]; LUKE STARK & JESSE HOEY, THE ETHICS OF EMOTION IN ARTIFICIAL INTELLIGENCE SYSTEMS (2021).

59 *See* Andrew S. Gold, *The Fiduciary Duty of Loyalty*, *in* THE OXFORD HANDBOOK OF FIDUCIARY LAW, *supra* note 36, at 401 ("Different opportunism risks will then justify different loyalty content and approaches to legal decision-making.").

60 *See* Richards & Hartzog, *Taking Trust Seriously*, *supra* note 5.

an information relationship are overwhelming, making these contexts ripe for subsidiary data loyalty rules.[61] We synthesize these contexts into five main areas: Entrustees should be loyal when *collecting* information, being sure to collect only information for purposes that do not conflict with a trusting party's best interests. Entrusees should be loyal when *personalizing*, i.e., treating people differently based upon personal information or characteristics. Entrustees should be loyal *gatekeepers*, avoiding conflicts when allowing government and other third-party access to trusting parties and their data. Entrustees should be loyal when trying to *influence* trusting parties, such as when they leverage personal data and digital tools to exert sway over people to achieve particular results. Finally, entrustees should be loyal in the ways they *mediate* interactions between users of their platform, specifically in the creation and administration of systems that govern how people are allowed to interact with each other. These contexts often overlap and involve issues like discriminatory microtargeting, harmful amplification of misinformation, failure of process for content moderation, and abusive dark patterns. We propose that lawmakers create subsidiary loyalty rules and standards to mitigate these kinds of disloyal behaviors.

1. Loyal Collection

A duty of loyalty should attach the moment a trusted party invites disclosure and makes the decision to collect personal information. In this way, data loyalty could embolden the fair information principle of data minimization. This principle holds that data collectors should only identify the minimum amount of personal information needed to fulfill a legitimate purpose and collection and hold that much information and no more.[62] Combined with the storage limitation principle, which holds that organizations should not keep data longer than they need it for their stated purpose, data minimization is a central pillar in data protection regimes around the world, but it too often fails to find traction.[63]

61 *See, e.g.*, Balkin, *The Fiduciary Model*, *supra* note 3; Scholz, *supra* note 3, at 197; Dobkin, *supra* note 3, at 17 (identifying four major ways of breaching an information fiduciary duty: "manipulation, discrimination, third-party sharing, and violating a company's own privacy policy"); Barrett, *supra* note 3, at 1100 ("[A]n information fiduciary framework should also address manipulation and discrimination in order to ensure that people are protected from the full array of modern digital threats that they face.").

62 *See Principle (c): Data Minimisation*, INFO. COMM'RS OFF., https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/#data_minimisation [https://perma.cc/6TTU-BJ8H].

63 *See id.*; Council Regulation 2016/679, art. 5(e) 2016 O.J. (L 119) 1 (EU) (providing the GDPR's storage limitation principle).

Data loyalty could provide a normative vision for when companies have exceeded their duty to minimize collection and retention—when it conflicts with a trusting party's (or collective trusting parties') best interests. Under general data protection frameworks that impose data minimization requirements, organizations must typically ensure that the data they are processing is adequate (sufficient to fulfil the stated purpose), relevant (has a relevant link to that purpose), and limited to what is necessary (collecting and holding only that which is needed for that purpose).[64] A duty of loyalty could provide a value-laden baseline that requires an examination of not just the purpose of the collection but also elevates the interests of those affected by the collection. While parties at an arm's length might act opportunistically in collecting as much data as possible, trusted parties remain loyal by leaving all data that, if collected, would conflict with the trusting parties' best interests on the table.

2. Loyal Personalization

The modern Internet routinely and systemically treats people differently based upon personal information or characteristics. Targeted and behavioral advertising is the most prominent example of this, but first-party product and streaming recommendations, news feeds, default settings, layouts, and more are all designed automatically to look and act differently based on people's personal characteristics. Some of this personalization, such as targeted recommendations for networked connections based upon intentionally revealed data such as where you work or attended high school, would probably be loyal. Other personalization systems, however, such as those that wrongfully discriminate or have a disparate impact on protected, marginalized, or vulnerable groups of people, would likely conflict with that trusting collective's best interests. Subsidiary rules built around the concept of loyal personalization could firmly and clearly address a systemic problem in a way that traditional data protection frameworks have been unable to mitigate.

3. Loyal Gatekeeping

Entrustees have a remarkable ability to facilitate third party access to trusting parties and their data. They can do so through their APIs, advertiser portals, fusion centers, and government backdoors. This access is the source of most major platforms' power. And everyone wants a piece of the users. Advertisers clamor for their attention. Data brokers and companies training AI models lust for their data. And

64 Council Regulation 2016/679, art. 5(c) 2016 O.J. (L 119) 1 (EU) (providing the GDPR's data minimization principle).

governments demand evidence. Entrustees have financial incentives to build portals and facilitate access for third parties. Some access granted by trustees to third parties is not in conflict with trusting parties' best interests. For example, contextual advertising usually doesn't significantly leverage people's own data or limitations against them, nor does it usually expose trusting parties to significant privacy harms. Protocols for interoperability to help people transfer data from one place to another also serve the interests (and wishes) of trusting parties.

However, certain lax gatekeeping practices would be disloyal because of how they endanger trusting parties by obscuring risk and breaking promises while facilitating access to third parties for organizational gains or to avoid costs. The three most resonant privacy scandals in the past decade, the government surveillance revelations by Edward Snowden, the FBI's request that Apple help it bypass encryption protections, and Cambridge Analytica's massive Facebook data exfiltration, all involved gatekeeping issues. Subsidiary rules built around the concept of loyal gatekeeping would help resolve longstanding debates around what obligations trusted organizations have regarding third-party access through portals, APIs, interfaces, and the automated scraping of websites. And in combination with a duty of confidentiality, subsidiary rules could also help clarify when sharing a trusting party's data with third parties is disloyal.

4. Loyal Influencing

Technologies are artifacts built to act upon the world. Every single design decision made in the creation of a website or app is meant to facilitate a particular kind of behavior.[65] The structure of digital technologies will affect people's choices even if the effect is not intended by designers. When designers create a drop-down menu, privacy settings, "I agree" buttons, and any other feature that implicates people's privacy, they are influencing them. They can't avoid it.[66] Given their power, they should be loyal in exercising their influence.

The most prominent example of disloyal influence involves organizations leveraging "dark patterns" or "malicious interfaces"

65 *See, e.g.*, LANGDON WINNER, THE WHALE AND THE REACTOR: A SEARCH FOR LIMITS IN AN AGE OF HIGH TECHNOLOGY 94 (1986).

66 *See* Cass R. Sunstein, *The Ethics of Nudging*, 32 YALE J. ON REGUL. 413, 421 (2015) ("Human beings . . . cannot wish [choice architecture] away. Any store has a design; some products are seen first, and others are not. Any menu places options at various locations. Television stations come with different numbers, and strikingly, numbers matter, even when the costs of switching are vanishingly low; people tend to choose the station at the lower number, so that channel 3 will obtain more viewers than channel 53.").

which are user interface elements meant to influence a person's behavior against their intentions or best interests.[67] Companies deploy effort traps to make deleting an account confusing and difficult. They make "cancel" buttons hard to see and press, they obscure important details in tiny fonts or walls of boilerplate, and they leverage our deeply entrenched and empirically validated overconfidence regarding risk, deference for conformity, endowment effects, status quo bias, and other biases and mental shortcuts to manipulate us to their ends. Jamie Luguri and Lior Strahilevitz have observed that "dark patterns are strikingly effective in getting consumers to do what they would not do when confronted with more neutral user interfaces."[68]

---

67 *See, e.g.*, HARTZOG, *supra* note 22, at 148, 162; Jamie Luguri & Lior Jacob Strahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEGAL ANALYSIS 43 (2021); Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014); Gregory Conti & Edward Sobiesk, *Malcious Interfaces and Personalization's Uninviting Future*, IEEE PRIV. & SEC., May/June 2009, at 72, 73; JOHANNA GUNAWAN, DAVID CHOFFNES, WOODROW HARTZOG & CHRISTO WILSON, TOWARDS AN UNDERSTANDING OF DARK PATTERN PRIVACY HARMS (2021); Harry Brignull, *Dark Patterns: Deception vs. Honesty in UI Design*, A LIST APART (Nov. 1, 2011), https://alistapart.com/article/dark-patterns-deception-vs-honesty-in-ui-design/ [https://perma.cc/4VBK-HEEG]; COLIN M. GRAY, YUBO KOU, BRYAN BATTLES, JOSEPH HOGGATT & AUSTIN L. TOOMBS, THE DARK (PATTERNS) SIDE OF UX DESIGN (2018); Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty & Arvind Narayanan, *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 PROCEEDINGS ACM HUM.-COMPUT. INTERACTION 81 (2019); ARUNESH MATHUR, JONATHAN MAYER & MIHIR KSHIRSAGAR, WHAT MAKES A DARK PATTERN. . . DARK? (2021); Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp & Stefan Pfattheicher, *Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns*, 4 PROC. ON PRIV. ENHANCING TECHS. 237, 248 (2016); Ari Ezra Waldman, *Cognitive Biases, Dark Patterns, and the 'Privacy Paradox'*, 31 CURRENT OP. PSYCH. 105, 105, 107–09 (2020).

68 Luguri & Strahilevitz, *supra* note 67, at 46 (emphasis omitted). Luguri and Strahilevitz found that

> [r]elatively mild dark patterns more than doubled the percentage of consumers who signed up for a dubious identity theft protection service, which we told our subjects we were selling, and aggressive dark patterns nearly quadrupled the percentage of consumers signing up. In social science terms, the magnitudes of these treatment effects are enormous.

*Id.* They further found that

> the most effective dark pattern strategies were hidden information (smaller print in a less visually prominent location), obstruction (making users jump through unnecessary hoops to reject a service), trick questions (intentionally confusing prompts), and social proof (efforts to generate a bandwagon effect). Other effective strategies included loaded questions and making acceptance the default. . . . In many cases, consumers exposed to dark patterns did not understand that they had signed up for a costly service. These results confirm the problematic nature of dark patterns and can help regulators and other watchdogs establish enforcement priorities.

*Id.* at 47.

Lawmakers have struggled for years to articulate when attempts at persuasion become harmful.[69] But trusting parties do not need to be injured for entrustees to violate a duty of loyalty. Subsidiary rules around disloyal attempts to influence would address the most pernicious and dangerous dark patterns head-on.[70] Lawmakers should focus on how the design is meant to take advantage of a person's limitations or vulnerabilities to benefit the designer in a way that is against the best interests of the trusting party.[71]

5. Loyal Mediation

Certain kinds of organizations design their platforms so that their users interact not just with the organization itself, but with each other. In other words, they mediate people's social and market experiences with other people using their service. Sometimes this is a great experience for people who use these services. But things can go off the rails quickly as companies feel pressured to achieve continual and endless growth. They create systems that reward virality and the most outrageous or venomous hot takes instead of the ostensible purpose of meaningful social interaction and social, emotional, and intellectual nourishment. They optimize their algorithms and interfaces to reward our most impulsive and petty reactions. Amplification of certain kinds of information combined with strategically reduced or increased transaction costs to speak, report harmful and dangerous speech, and hide from other users leads to acute individual harms like harassment[72]

---

69 *See id.* at 104; *see also* Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in A Digital World*, 4 GEO. L. TECH. REV. 1, 3 (2019) ("[A]t its core, manipulation is hidden influence—the covert subversion of another person's decision-making power. In contrast with persuasion, which is the forthright appeal to another person's decision-making power, or coercion, which is the restriction of acceptable options from which another person might choose, manipulation functions by exploiting the manipulee's cognitive (or affective) weaknesses and vulnerabilities in order to steer his or her decision-making process towards the manipulator's ends.").

70 Luguri and Strahilevitz recommend a multi-factor test to help determine when dark patterns cross the line

> that looks to considerations such as (i) evidence of a defendant's malicious intent or knowledge of detrimental aspects of the user interface's design, (ii) whether vulnerable populations—like less educated consumers, the elderly, or people suffering from chronic medical conditions—are particularly susceptible to the dark pattern, and (iii) the magnitude of the costs and benefits produced by the dark pattern.

Luguri & Strahilevitz, *supra* note 67, at 99.

71 Balkin has proposed looking to "techniques of persuasion and influence that (1) prey on another person's emotional vulnerabilities and lack of knowledge (2) to benefit oneself or one's allies and (3) reduce the welfare of the other person." JACK M. BALKIN, HOOVER INST., FIXING SOCIAL MEDIA'S GRAND BARGAIN 4 (2018).

72 *See* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014).

as well as systemic harms like polarization, reduced ability to engage in self-governance, negative public health outcomes, and chilling effects for large groups of vulnerable users.

A duty of loyalty cannot solve all of the complex problems of content moderation or harassment. As we have maintained, a duty of loyalty is merely one important tool in a larger toolkit. But companies do have remarkable power to influence how people using their systems interact with each other.[73] When they use this power in a way that conflicts with the best interests of their users in order to optimize growth, they are being disloyal. Subsidiary rules for loyal mediation are, of course, complicated because of the potentially conflicting interests amongst actors and those potentially adversely affected by the act. One trusting party wants to speak while the other(s) is made worse because of it. This is where our proposed systemic focus and the traditional fiduciary law method of developing a hierarchy of loyalties would help clarify lawmakers' actions.

## CONCLUSION

Duties of data loyalty will take time and effort to meaningfully implement as a part of U.S. privacy law. Data loyalty is a significant and necessary departure from privacy law's ineffective notice and consent approach. But lawmakers can confidently embrace loyalty and other relational duties as part of a holistic approach to mitigating the power and abuses of data collectors. If done clearly, carefully, and with commitment, lawmakers can chart a bold new vision for our privacy rules that is capable of nurturing a sustainable and flourishing future for those who share their personal information as well as those entrusted with it.

73 *Id.* at 25; *see also* Hartzog & Richards, *Privacy's Constitutional Moment*, *supra* note 5, at 1695.

**Developing and Deploying AI Responsibly:**
**Elements of an Effective Legislative Framework to Regulate AI**

**Written Testimony of Brad Smith**
**Vice Chair and President, Microsoft Corporation**

**U.S. Senate Judiciary Committee**
**Subcommittee on Privacy, Technology, and the Law**

**September 12, 2023**

Chairman Blumenthal, Ranking Member Hawley, Members of the Subcommittee, thank you for the opportunity to testify on what I believe is a critical issue for the Congress and the country. I am Brad Smith, the Vice Chair and President of Microsoft Corporation.

I welcome the opportunity today to share some initial thoughts on the proposed Blumenthal-Hawley framework and the critical role legislation must play in ensuring effective oversight of artificial intelligence (AI).

In short, I believe this framework is a strong and positive step towards effectively regulating AI. It reflects the urgency and speed needed to address this fast-moving technology, and it combines strong protection for the public with support for ongoing technology innovation.

Importantly, the framework builds on other federal efforts, like the White House AI commitments unveiled in July and the bipartisan AI Insight Forums, providing the constructive interplay needed between the executive and legislative branches. And it will enable Congress to listen and learn before addressing every detail, while establishing not just a blueprint but the initial foundation on which additional legislative steps can be taken.

As the legislative process moves forward, I hope Congress will include three goals in the list of priorities that deserve the most attention.

***First, Congress should prioritize AI safety and security.*** As I discuss in more detail below, the Blumenthal-Hawley framework addresses these needs in a strong manner, including by proposing a licensing regime under an independent oversight body with a risk-based approach for AI models and uses. Microsoft supports this approach and believes it strikes a sensible balance that can both protect the public and advance innovation, even while we recognize the need to work through a large variety of hugely important details. By also incorporating transparency and security requirements, the Blumenthal-Hawley framework puts Congress on a path to provide the public with the safety standards it deserves.

***Second, Congress should ensure that AI is used in a manner that complies with longstanding legal protections for consumers and citizens.*** This should include the protection of privacy, civil rights, and the needs of children, as well as safeguards against dangerous deepfakes and election interference. The Blumenthal-Hawley framework rightly addresses these issues head-on, without shying away from the critical issues that Congress must consider if it decides, as we believe it should, to replace Section 230 with new approaches that are a better fit for AI technology.

The framework also focuses, as we believe it should, on the differing roles of AI developers and AI deployers, in effect creating a pragmatic regulatory architecture that reflects relevant AI technology architecture. The Blumenthal-Hawley framework makes room for considering in a practical way where current laws may be sufficient if enforced well, as well as for the possibility of new rules where needed. In sum, it strikes a sensible balance based on a practical understanding of the relevant technology, the need to advance innovation, and the imperative to protect the nation's citizens.

***Third, Congress should ensure that AI is put to good use to build a government that can better serve our citizens.*** We should not overlook the opportunity to put AI to use as a valuable tool, even while we protect against its potential abuse as a technological weapon. Some national governments in other countries are moving quickly to use AI to improve healthcare, strengthen education, make public services more accessible, and advance public sector efficiencies. In the United States, state leaders like Governor Newsom in California, Governor Burgum in North Dakota, and Governor Youngkin in Virginia are taking an early lead in using AI to build better state governments. At the federal level, we should consider the role of legislation and oversight to advance similar goals, as well as the strengthening of the country's national defense and security. I hope the Blumenthal-Hawley framework will expand to make more room to address not only new risks associated with AI, but new opportunities as well, especially to build a government that can better serve the public.

In the sections that follow, I focus in greater detail on the principles that Microsoft believes should guide the development of legislation to govern AI, many of which are embraced by the Blumenthal-Hawley framework. This includes ensuring that legislation promotes accountability for both AI development and deployment, builds on existing work, and reflects the technical architecture of AI itself. I will also discuss priority areas where federal regulation and oversight seem particularly appropriate, including requiring "safety brakes" for highly capable AI models used in critical infrastructure, and requiring developers of AI systems to know their customer, their cloud, and their content.[1]

## 1. Promote accountability in AI development and deployment

When we at Microsoft adopted our six ethical principles for AI in 2018, we identified one principle that should serve as the bedrock for all the others: accountability. Accountability means that AI systems must be subject to effective oversight by humans. It also means that the people who develop and deploy these systems must remain accountable to everyone else, affording those impacted by harmful AI systems protection under the rule of law.

We therefore welcome the decision by Senators Blumenthal and Hawley to make accountability a centerpiece of their proposed regulatory framework. It recognizes that to govern AI effectively, we need to act in a targeted way and place the right expectations on the right stakeholders to address the risks of greatest concern. Developers and deployers of AI systems must work together to strengthen AI safety and apply special care in the highest risk scenarios; AI systems that are used to make consequential decisions about the most vulnerable members of our communities should be subject to greater oversight than the AI systems that help us find the next emerging musician based on our prior playlist.

As this Committee considers how best to ensure that the United States remains a world leader in responsible innovation, we would encourage you to focus not only on accountability in the development of AI, but also on accountability in its deployment. Promoting accountability in deployment means risk will be managed more effectively, and that AI can be put to work more broadly to help people in their day-to-day lives and make progress on our greatest societal challenges.

Throughout history, we have seen that countries that lead in the uptake of new technologies are often the ones that fare best, sometimes even over countries that may be more technologically innovative. Consider, for example, the printing press. Few inventions in history have had such profound effects on the world. Yet although the printing press was invented in Germany in the

[1] Further details on these ideas can be found in Microsoft, *Governing AI: A Blueprint for the Future* (May 25, 2023), at https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw.

1400s, it was the Dutch and the English who first truly embraced printing and books. Indeed, by 1500, citizens in the Netherlands were reading more books per capita than anyone else in the world. It is certainly no coincidence that England and the Netherlands also soon found themselves at the forefront of economic innovation and global commerce.

This lesson is worth bearing in mind as we think about how governments can best promote the benefits and responsible use of AI. Although AI innovation is undoubtedly important, those countries that succeed in rapidly *adopting and using* AI responsibly are the ones most likely to reap the greatest benefits. Fortunately, the United States is well placed to lead both responsibly and rapidly because of its existing AI safety initiatives.

**2. Build on existing efforts**

The United States is the home of many of the world's top developers of advanced AI. Given the foresight of U.S. policymakers and regulators, it also has a number of existing AI safety frameworks on which to build—including the White House initiative to secure voluntary commitments from industry, and the NIST AI Risk Management Framework. These initiatives target different aspects of AI development and deployment, making them especially relevant to use in concert to support accountability across both activities. As this Committee considers the proposed Blumenthal-Hawley framework and other efforts already underway in Congress to regulate AI, we encourage you to take account of these existing initiatives.

The White House voluntary commitments focus on three fundamental principles: safety, security, and trust.[2] They target developers of 'frontier' AI models – the most advanced models that exceed the capabilities of currently released systems like OpenAI's GPT-4 – requiring participating companies to take on specific duties designed to advance each of these principles, in particular:

- Ensure that their products are *safe* before offering them to the public. This includes commitments to engage in red teaming of frontier models to identify safety risks, to share information with other companies and governments on emerging risks and threats, and to develop standards and best practices for frontier AI safety.

- Build systems that put *security* first. This includes commitments to invest in cybersecurity and insider threat safeguards, and to provide incentives for third parties to discover and report AI security issues and vulnerabilities.

- Do right by the public and earn people's *trust*. This includes commitments to deploy provenance technologies or watermarks so that people know when they encounter AI-generated audio or visual content; to publicly disclose model or system capabilities, limitations, and appropriate and inappropriate uses, including effects on fairness and bias; to prioritize research into societal risks posed by AI systems; and to develop and deploy frontier AI systems to help address society's greatest challenges.

Microsoft was among the first companies to adopt these commitments, and I am proud to say that we have gone beyond them[3]—a point I will come back to later in this testimony. Among other

[2] *See* White House, *Voluntary AI Commitments*, at https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf.

[3] *See* Microsoft, *Voluntary Commitments by Microsoft to Advance Responsible AI Innovation* (July 21, 2023), at https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2023/07/Microsoft-Voluntary-Commitments-July-21-2023.pdf.

things, we have committed to implement the NIST AI Risk Management Framework (AI RMF) across our own AI development and deployment practices, and to attest to this with our customers.

NIST developed the AI RMF based on the directive that this Congress issued in the National Artificial Intelligence Initiative Act of 2020. The framework, which supports risk management efforts by both developers and deployers of AI technologies, builds on NIST's years of experience in developing similar frameworks to address cybersecurity risks. We commend NIST for having developed the AI RMF, and in particular for having done so in a transparent and consensus-driven process involving input from both the public and private sectors.

We are pleased to see that the Blumenthal-Hawley regulatory framework incorporates transparency measures to promote responsibility and due diligence from the companies developing and deploying AI systems, and we believe the NIST AI RMF could be a useful tool to inform the risk management strategies and practices about which they should be transparent. One place to start could be federal procurement. The Federal Government has a proven track record of using procurement rules to incentivize industry to innovate, improve their products, and embrace industry best practices more generally.

Building on the model of existing rules that require federal contractors to adopt strong cybersecurity practices, Congress could likewise encourage industry adoption of standards based on the AI RMF by requiring federal contractors to self-attest, as a condition of bidding on federal contracts, that they have implemented those standards. Because it's always better to learn to walk before trying to run, Congress could initially focus on the procurement rules for critical decision systems, meaning AI systems that have the potential to meaningfully impact the public's rights, opportunities, or access to critical resources or services. As industry uptake of responsible AI practices increases, the Government could expand the scope of these procurement rules to additional areas as appropriate.

Congress could also consider directing NIST to establish an AI RMF program office to provide federal agencies with guidance on the framework and to promote its adoption. This Office could also provide resources on AI trustworthiness to officials responsible for procuring such systems. In tandem, the General Services Administration and OMB could be directed to develop voluntary, standard contract language for agencies to use in those procurements.

While the White House voluntary commitments and the NIST AI RMF provide a strong foundation, we also need laws and regulation that build on this foundation and complete the AI regulatory architecture.

**3. <u>Require safety brakes for AI that controls or manages critical infrastructure</u>**

Artificial intelligence is not the first-time societies have confronted a valuable new technology that also has the potential for harm if it fails. History is replete with examples.

For instance, the growth of cities led to increasingly taller buildings, which in turn required the use of elevators. People were understandably worried about what might happen if the cables holding an elevator aloft were to fail. Elisha Otis, the inventor of the elevator, solved this by developing a brake that would catch the elevator car before it fell. It wasn't long before city building codes required in the installation of safety brakes on all elevators. And it worked: today, elevators are supremely safe, and we don't give a second thought to stepping into one.

That pattern has repeated itself many times. Governments require circuit breakers in buildings to protect against fires caused by surges in electricity. They require school buses to have emergency

brakes, in case the main brakes fail, and require bus drivers to be trained in how to use them. They require airplanes to have collision avoidance systems installed, and to ensure that pilots are able to make decisions based on those systems in safety-critical situations.

The common thread running through all these examples is that, where a technology failure can cause significant or widespread harm, it is often appropriate to require suppliers to install back-up safety systems, and to ensure that people have the ability to use them if they're ever needed.

The Blumenthal-Hawley framework recognizes this critical role for 'safety brakes'. A clear scenario in which they are needed is when highly capable AI models are used to access, manage, or control critical systems, the failure of which could cause widespread harm—for example, systems controlling the power grid, first responders, transportation traffic, and the like. Laws requiring developers to build safety brakes into such systems, and requiring that deployers can use them effectively, would promote accountability by ensuring that these systems remain under human control at all times.

Although the details of what such a regulatory regime would look like deserve further consideration, we envision it having at least four components:

- First, direct regulators to define the class of high-risk AI systems controlling critical infrastructure that would require safety brakes. At least initially, regulators could focus on highly capable AI systems that: (i) take decisions or actions affecting large-scale networks; (ii) process or direct physical inputs and outputs; (iii) operate at least semi-autonomously; and (iv) would pose a significant potential risk of large-scale harm if they were to fail.

- Second, require AI developers to build safety brakes into the design of these designated AI systems. Although the type of 'safety brake' would likely vary depending on the system and how it was used, all of them should have the ability to detect and avoid unintended consequences, and to disengage or deactivate the AI system in the event of unintended behavior.

- Third, require deployers of designated systems to test and monitor them to ensure that they remain within human control. Although exact requirements will depend on the system, deployers should have to periodically test, verify, and rigorously validate the operation and performance of the system and its components, consistent with safety best practice. Deployers should be accountable for proving that they can operate safety brakes built in by developers and that the system remains under human control at all times.

- Fourth, AI systems that control the operation of designated critical infrastructure should be deployed only in licensed AI infrastructure. Because a point of vulnerability for AI systems can be the computing infrastructure on which they run, we think it is also worth requiring the operators of such infrastructure to be licensed. To obtain a license, the operator would need to design and operate their infrastructure in a manner that allows another point of intervention—in effect, another safety brake in the event that the application-level measures fail.

Today, there are relatively few AI systems that exert the kind of control over critical systems that would necessitate these types of safety brakes. But we shouldn't use that as an excuse for delay. Given the rapid pace at which highly capable AI systems are progressing, we should begin developing the rules of the road now for the future that we know is coming.

**4. KY3C: Know your customer, cloud, and content**

As the Blumenthal-Hawley framework recognizes, it is important to think about how laws can help promote the responsible use of AI *and* guard against its misuse. It's also valuable to leverage regulatory frameworks that have proven themselves to be effective in addressing risks that share characteristics in common with those we are most concerned about for AI.

In the financial services context, 'know your customer' obligations seek to minimize the risk that financial institutions unwittingly facilitate transactions that are being used for illegal ends. They have been critical to advancing U.S. national and economic security, and the integrity of the U.S. financial system, by enabling authorities to identify and tackle money laundering, terrorist financing, and other crimes.

More recently, a refined set of 'know your customer'-inspired techniques is emerging as a best practice in the cybersecurity space. Leading cloud service providers maintain policies and processes to protect against threat actors gaining inappropriate access or engaging in certain abusive activities, including by using digital indicators and machine learning to hunt for threat actors, disable their accounts, and assess new accounts for fraud and abuse risk.

In our view, a version of this framework could be applied with good effect in the AI context, advancing several of the goals of the Blumenthal-Hawley framework. In particular, we think it's worth considering a "KY3C" regulatory framework that would impose obligations on various actors in the AI value chain, requiring them to know their *customers*, their *cloud*, and their *content*. Specifically:

- Know your customer. At least in scenarios involving high-risk or other sensitive use cases, it might be appropriate to require operators of the cloud infrastructure on which a highly capable AI model is running to know the customers who are accessing the model. Depending on the scenario, it might also be appropriate to require AI model developers or operators of highly sensitive AI applications to know their customers, managing access not only to AI datacenter infrastructure but also to the models and broader capabilities for sensitive uses. As in the financial services sector, the goal would be to reduce the risk that AI providers unwittingly let bad actors use powerful AI models or applications running on their infrastructure to engage in theft, fraud, or other misuse.

- Know your cloud. Developers of highly capable AI models, in turn, should have an obligation to "know the cloud" on which their models are deployed. This obligation would help reduce the risk that bad actors can exploit vulnerabilities in that cloud to corrupt the functioning of those models. To satisfy this obligation, developers would have to use licensed AI cloud infrastructure, requiring the cloud provider to meet ongoing regulatory requirements proving that they have taken the necessary steps to protect that infrastructure against malicious attacks and adversarial actors.

- Know your content. Although the creative potential of highly capable AI models is astounding, it also opens the door for the creation of deepfakes, misinformation, and other malicious content. To help reduce the risk of people being deceived or misled, we think the public deserves to "know the content"—that is, to know when they come into contact with images or audiovisual content that has been produced or altered by AI. The law could require deployers of AI systems to accomplish this through the use of provenance technologies or watermarking, at least in scenarios presenting a significant risk of harm.

> Key technical building blocks to enact such a legislative requirement already exist. One of the most important of these is the Coalition for Content Provenance and Authenticity (C2PA), co-founded by companies including Adobe, the BBC, Intel, Microsoft, Sony, and Truepic. As Microsoft's Chief Scientific Officer Eric Horvitz has said, the success of such labeling will require education aimed at media literacy, awareness, and vigilance, together with investments in quality journalism. Although success on these fronts will require work, the health of democracy and meaningful civic discourse will undoubtedly benefit from initiatives such as these that help protect the public against deception or fraud facilitated by AI-generated content.

We appreciate that these proposals will require further thought and discussion, including on how best to balance national security and crime-prevention goals against the need to protect privacy, freedom of expression, and other fundamental human rights. But if we target regulation at those use cases that pose the greatest risk, and put adequate safeguards in place, we think it is possible to balance these interests in ways that benefit both individuals and society.

**5. Ensure the regulatory framework maps to the technology architecture of AI**

Both the safety brake and the KY3C proposals discussed above are examples of a broader point—namely, that the regulation of AI needs to map to the technology architecture of AI itself. To be effective, the law needs to place different regulatory responsibilities on different actors based on their role in managing different aspects, or layers, of AI technology. This is particularly important when it comes to regulating highly capable and other frontier AI models, which is a critical aspect of any federal regulatory regime. It's also consistent with the Blumenthal-Hawley framework's embrace of the core principle that both developers and deployers should be accountable for AI safety.

There is no single right way to describe this technology architecture, and different engineers might well describe it differently. For our purposes, it's useful to think of AI as involving a technology stack made up of three core layers: the application layer, the model layer, and the infrastructure layer, more commonly known as the "cloud" layer. Each layer of this technology stack, in our view, necessitates a distinct regulatory approach. AI developer accountability at the infrastructure, model, and application layers must be implemented in ways that correspond with the technology stack, just as AI deployer accountability at the application layer must also correspond with the risk management capabilities uniquely available as AI is deployed.

**The applications layer**

At the top of the technology stack are AI-powered applications and services. These are what deliver information and other AI outputs directly to users. The application developer may operate the AI "model"—the prediction or classification engines that power these applications and services—itself, or obtain that functionality from a third party through an "application programming interface," or API. OpenAI's ChatGPT, Microsoft's Bing Chat, and GitHub's CoPilot (an application that generates software code in response to user prompts) are all examples of AI-powered applications. Microsoft also offers APIs to enterprise customers that allow them to build applications on top of AI models that Microsoft makes available from OpenAI and Microsoft itself. This lets enterprises easily build AI solutions to suit their own needs, without having to develop their own AI models or build the sophisticated infrastructure on which these models run.

Because this is the layer at which people directly interact with AI outputs, it is also the layer at which the safety and rights of people will be most impacted. As a result, we need to ensure that the laws and regulations that currently govern conduct apply with equal force to those who provide or deploy

AI applications and services. Simply put, we need to apply and enforce the laws that are already on the books.

For example, the fair performance of AI systems across different demographic groups is a well-known and pressing concern for AI development. It is already unlawful for a bank to discriminate against a mortgage applicant based on race or gender. If a bank decides to use AI tomorrow to help it evaluate loan applications, it will need to ensure that this doesn't lead to such discrimination. We don't need new laws to make this happen; we just need to apply the laws we already have.

Although this may sound simple, the implications are potentially profound. Those who develop and deploy AI applications and services will need to think hard, and proceed carefully, to ensure that they fully comply with existing laws—most of which, however, were drafted before the advent of AI. Regulators will need to know how to apply those laws to AI applications and services, and courts will need to know how to interpret and enforce those laws with regard to AI.

**The model layer**

The next layer down in the AI technology stack consists of pre-trained AI models. These are the foundational technologies that power the top layer of AI applications and services. Most AI models used today are designed to do discrete tasks—such as translate text, or recognize patterns in images—and do not, in our view, require new legislation. However, a class of highly capable AI models is emerging that may require new regulatory approaches, both for the models themselves and for the infrastructure on which they run.

These highly capable AI models are unique in many respects. They are typically trained on huge, internet-scale datasets and include billions of parameters that interact in exceedingly complex ways. Many are also effective out-of-the-box at doing a wide range of tasks—from drafting a mathematical proof or writing a poem, to developing software code for a new application or creating an image of a sunset seen from Mars.

In addition to being multifaceted and powerful, the behavior of these frontier, highly capable AI models can also be hard to predict. In many cases, these models are so complex that the outer bounds of their capabilities can only be determined in practice, such as through controlled releases with users. As a result, harnessing the full potential of these models, while also ensuring that they align with our laws and values, is challenging, and our understanding of how to do this effectively is evolving.

Although developers like Microsoft are addressing these risks through rigorous testing and self-imposed standards, the risks involved are too important, and their scale and potential impacts at present too unknowable, to address them through self-regulation alone. We therefore think it is appropriate for Congress to consider legislation that would impose a licensing regime onto developers of this discrete class of highly capable, frontier AI models, and we are pleased to see that the Blumenthal-Hawley regulatory framework seeks to establish such a regime.

Although the details of this licensing regime again would benefit from further thought and discussion, and there are critical consequences and details to deeply consider, such as the impact to open source models and the importance of continuing to foster an innovative open source ecosystem, we think it should seek to serve three key goals:

- First and foremost, any licensing regime must ensure that the development and deployment of highly capable AI models achieve defined safety and security objectives. In concrete terms, this

may require licensees of these models, among other things, to engage in the pre-deployment testing that the Blumenthal-Hawley regulatory framework proposes. We agree that highly capable models may need to undertake extensive prerelease testing by internal and external experts. In addition, a licensing regime may require developers of highly capable models to provide advance notification of large training runs; engage in comprehensive risk assessments focused on identifying dangerous or breakthrough capabilities; and implement multiple other checkpoints along the way.

- Second, it must establish a framework for close coordination and information sharing between licensees and regulators, to ensure that developments material to the achievement of safety and security objectives are shared and acted on in a timely fashion. The Blumenthal-Hawley framework provides that an independent oversight body not only conducts audits but also monitors technological developments, which may be best accomplished in partnership with licensees. The adoption of controls over model deployments, potentially based on the assessed level of risk and evaluations of how well-placed users, regulators, and other stakeholders are to manage residual risks, may be required. Post-release monitoring may also help ensure that the models are functioning as intended and remain under human control at all times.

- Third, it must provide a footing for international cooperation between countries with shared safety and security goals. Because AI systems and their outputs are not confined by geographic borders, domestic regulation alone will not be enough to secure the benefits of highly capable AI models and guard against their misuse. We believe there is an opportunity for the United States to work with like-minded countries to advance an international framework for AI governance, enabling an AI system evaluated as safe in one jurisdiction to qualify as safe in another. There are many effective precedents for this, such as common safety and security standards set by the International Civil Aviation Organization, which allows for an airplane to fly from Brussels to New York without being re-fitted over the Atlantic. The United States could also work with others to advance a global consensus on the most pressing risks and opportunities around AI and to improve our collective understanding of AI safety.

**The cloud infrastructure layer**

The third layer in the technology stack is the cloud infrastructure layer on which highly capable AI models are developed and run. This infrastructure is more powerful than the datacenters that run more traditional digital applications and services: it typically provides far greater computer power, uses specialized AI chips, and involves significant engineering skills and resources.

Because it provides the technological foundation on which highly capable AI models run, this AI cloud infrastructure layer is a key control point for these models—and also a potential point of vulnerability if they are not managed properly.

We therefore see a role for licensing providers of this cloud infrastructure to ensure that they act responsibly in ensuring the safe and secure development and deployment of highly capable AI models. To obtain a license, an AI datacenter operator would need to satisfy specified technical capabilities around cybersecurity, physical security, and safety architecture. As noted above, it would also include a "know your customer" requirement to guard against these datacenters unknowingly permitting their infrastructure to run AI applications used for criminal or other harmful purposes, and have the capability to provide a "safety brake" on AI systems used to control or manage critical systems.

These requirements and a new generation of export controls can help protect U.S. national security interests and avoid the proliferation of frontier models to adversaries and those intending to cause harm. The need for such regulation will become increasingly clear as AI models on the frontiers of technology become more capable, more autonomous, and more likely to bridge the digital-physical divide. Congress would do well to ensure that we stay ahead of these risks by ensuring the appropriate legislative guardrails and authorities are in place ahead of time.

**6. Conclusion**

Powerful new AI technologies should give all of us grounds for optimism, given their many potential benefits when they are developed and deployed responsibly. At the same time, we must not ignore their potential perils. Industry plays an essential role in promoting the safe and responsible development of AI. But laws and regulations have a vital role to play as well. At their core, these laws should require AI systems to remain subject to human control at all times, and ensure that those who develop and deploy them are subject to the rule of law. We need to place the right expectations on the right stakeholders to address the risks of greatest concern. The Blumenthal-Hawley framework sets consideration of all of this on the right course.

RESPONSES OF BRAD SMITH TO QUESTIONS SUBMITTED BY SENATOR HAWLEY

1. **During the period of time it was available on the public internet, how many times was the Microsoft Celeb database accessed by individuals located within the People's Republic of China?**

The MS Celeb database was retired in 2019. Microsoft does not retain data indicating the location of individuals who accessed the site prior to that point.

2. **How much money has Microsoft invested in artificial intelligence development in the People's Republic of China?**

Microsoft incorporates artificial intelligence (AI) into nearly all our products and services. We account for AI-related expenditures at the business unit level, not by geographic region. Microsoft does conduct certain AI work in China, subject to careful controls that go beyond what U.S. export controls and other applicable laws and regulations require. Microsoft Research Asia (MSRA), for example, does not engage in collaborative research with researchers or students who are attending or employed by Chinese military institutes, and such students and researchers are ineligible for internships, fellowships, or academic research grants. Likewise, MSRA does not partner with any academic in China who is employed by a university on the Bureau of Industry and Security's Entity List or Unverified List. It also does not undertake any such collaborations in China involving sensitive AI use cases like facial recognition or synthetic media.

In general, Microsoft believes that U.S. national security interests are best advanced by U.S. technology leadership on the global stage. Our work in China serves this goal by providing Microsoft access to world-class talent and enabling us to stay at the heart of the global tech ecosystem. The research we do at MSRA is made public and freely available globally, advancing the state of the art in computer science. At the same time, we are clear-eyed about the risks of working in China and vigilant about protecting U.S. national security.

Ensuring U.S. technology leadership also requires accelerating domestic innovation. Microsoft is proud to be at the forefront of U.S. AI innovation. We invest substantial sums annually both in AI development and in building the secure, state-of-the-art data centers required to run frontier models. We are also strong supporters of the National AI Research Resource (NAIRR), which will ensure that government researchers and academics have access to the compute power necessary to conduct cutting-edge research.

3. **How many China-based Microsoft employees are members of the Chinese Communist Party?**

Microsoft does not ask employees in China or elsewhere to provide information about affiliations with political parties.

4. **To the extent you are aware, has the personal data of American citizens ever been accessed by China-based Microsoft employees? If not, please provide a full explanation of all security measures employed by Microsoft to prevent such access.**

Microsoft is an industry leader in protecting personal data. We live up to our commitment to protect customer privacy by providing products, information, and controls to help customers make informed choices about their data. We publish a Microsoft Privacy Report twice a year to give customers the latest information on what personal data we collect, how it may be used, and how to manage and control their data.

At times, our global engineering teams must access customer environments where data resides to improve the quality and reliability of our products and services and to address customer service issues. All such access is governed by our global privacy policies. Microsoft does not grant its engineers outside the United States standing access to Bing Chat data that is linked to individual names, for example. When necessary to complete engineering tasks, however, such engineers may obtain access to chat data that is not associated with any account registration information (including the user's name).

5. **How does Microsoft verify that users of its Bing chatbot meet the purported age requirement of 13?**

At Microsoft, we recognize we have a responsibility to ensure that younger users' experiences with technology are safe, age-appropriate, and advancing their best interests. While we believe there are real opportunities for young people to make the most of generative AI and other technologies, including to support their learning and access to information, we also recognize that young people may be vulnerable to unique harms and we take steps to address those risks.

To access the full Bing chat experience, individuals must sign up for a Microsoft account. Consistent with the U.S. industry standard and the Children's Online Privacy Protection Act (COPPA), we have long required an age gate before a Microsoft account may be created in the United States. Individuals identified as under the age of 13 are prevented from creating an account unless their parent provides verified consent, and in those cases, the account comes with a variety of parental controls and safety settings. Not only do Microsoft child accounts come with these additional safety settings and parental controls, they also are prohibited from accessing the full Bing chat experience, as well as the recently released Copilot feature for Windows.

While our approach to age verification represents the industry best practice, this is an area where industry and policymakers should continually reassess current practices and seek improvement. Identifying child users is important to keeping them safe and ensuring that their experiences are age-appropriate. However, age assurance methods require key trade-offs between accuracy, privacy, and other critical interests, and getting

that balance right is important. These topics are subject to active discussion and study by civil society, policymakers, academia, and industry in jurisdictions around the world. We are committed to partnering with others on these issues and are already working diligently to identify innovations to keep kids safe while enabling the public to access information and protecting personal privacy.

For instance, the Xbox Live service recently launched an age verification test pilot in August 2023 with a small set of U.S. Xbox beta testers over the age of 18. The pilot asks beta testers to match a selfie and a government identification to verify that they are 18+. Xbox launched this test pilot to better understand consumer preferences in this area, what consumers' concerns would be in using an age verification mechanism, and how to build a verification experience in a privacy-protective and legally compliant manner. While we are excited about the potential for innovation in this area, more research is needed to ensure that age assurance mechanisms like these, which require verification of either facial features or other personally identifiable information from customers, are respective of privacy, protective of data, and can be implemented safely.

We believe there is a need for research to help understand how younger users interact with generative AI more broadly, what its impact might be, and whether (for example) it might be used in ways to make children's online experiences safer. This is work we are actively exploring in partnership with others.

Finally, we continue to support efforts in Congress and in state capitals to pass new laws and expand upon the protections contained in COPPA.

**6. Please provide a full explanation of Microsoft's rationale for setting the Bing chatbot age requirement at 13, rather than at 18.**

Prior to the launch of Microsoft's new Bing experience, we proactively analyzed the potential benefits and risks the new features could entail for child and teen users. We recognized that the benefits of AI are not risk-free, and an essential component of Microsoft's Responsible AI program is to identify potential harms, measure their propensity to occur, and build mitigations to address them. While this approach is important for all users of our technology, it is critical for our youngest users.

We determined there are real opportunities for young people to use generative AI in positive ways to assist with their learning, enhance their access to important educational information, and position them to seize future economic opportunities. However, we also recognized that younger users have unique needs, and child users may not always understand the limits of the technology with which they are engaging, particularly when that technology is relatively novel. Accordingly, like we do with our other services, we aimed to take a balanced approach that enables young people to enjoy the benefits of

technology while protecting their safety and wellbeing. We have developed and implemented a strong safety architecture for the Bing chatbot that addresses important child safety needs. In addition, based on an analysis of the potential risks and consistent with our wider approach to safety, Microsoft child accounts (i.e., accounts for those under 13, for which we have special obligations under COPPA) are currently prohibited from accessing the full new Bing experience.

7. **How many times has the Chinese government requested user data from Microsoft's Chinese operations—whether pursuant to China's "national security law," or for some other reason?**

Microsoft does not provide any government with direct or unfettered access to our customers' data, and we do not provide any government with our encryption keys or the ability to break our encryption. If a government seeks customer data, it must follow applicable legal process. All requests must include information about specific accounts and identifiers. Microsoft's legal compliance team reviews all requests to ensure they are compulsory and valid and rejects those that are invalid. Between January 2021–December 2022, which includes our most recent reporting period, Microsoft received 25 legal requests from Chinese law enforcement authorities; Microsoft rejected all of them and did not release customer data. These statistics are reported in our biannual transparency report, titled Law Enforcement Request Report (available at https://www.microsoft.com/en-us/corporate-responsibility/law-enforcement-requests-report).

Separately, while we have chosen not to offer several consumer-facing products and services in China that we offer elsewhere in the world, we have licensed certain cloud services technology for enterprise customers in China to a third-party services provider in China. This third-party operates the local data centers in China and provides subscription and billing services, as well as technical support. The entity is contractually obligated to follow Microsoft's lawful access policies in the event it receives a government request for data.

STATE OF CALIFORNIA Gavin Newsom, Governor

**CALIFORNIA PRIVACY PROTECTION AGENCY**
2101 Arena Blvd
Sacramento, CA 95834
www.cppa.ca.gov



September 12, 2023

The Honorable Richard Blumenthal, Chair
The Honorable Josh Hawley, Ranking Member
Subcommittee on Privacy, Technology, and the Law
United States Senate Committee on the Judiciary
Washington, DC 20510

**Re: Privacy, Technology, and the Law Subcommittee hearing, "Oversight of A.I.: Legislating on Artificial Intelligence"**

Dear Chair Blumenthal and Ranking Member Hawley:

The California Privacy Protection Agency[1] (CPPA or Agency) sincerely thanks the United States Senate Subcommittee on Privacy, Technology, and the Law for exploring artificial intelligence (AI) policy in its hearing, "Oversight of A.I.: Legislating on Artificial Intelligence." We are pleased to provide this information about our work to implement and enforce the California Consumer Privacy Act (CCPA), and particularly our role with respect to AI and automated decisionmaking technology (ADMT), as you consider these issues.

The CCPA directs the Agency to work with policymakers in other jurisdictions to ensure consistency in privacy protections.[2] We are grateful for the collaboration thus far and look forward to continuing to work with you to ensure this consistency, and to make it easier for consumers to exercise their rights and for businesses to comply.

**The California Consumer Privacy Act**

In 2018, California became the first state in the nation to adopt a comprehensive commercial privacy law, the California Consumer Privacy Act.[3] That measure went into effect on January 1, 2020, and the Attorney General began enforcing it on July 1, 2020. In November 2020, California voters ratified Proposition 24, the California Privacy Rights Act, which amends and expands the CCPA, including by creating the first authority with full administrative powers focused on privacy and data protection in the United States, the California Privacy Protection Agency. On April 21, 2022, rulemaking authority under the CCPA formally transferred to the Agency, and on March 29, 2023, the Agency finalized its first substantive package of regulations to further implement the CCPA.[4] Along with the Attorney General, the Agency is vested with the authority to undertake enforcement to protect Californians' privacy.

[1] Established in 2020, the California Privacy Protection Agency was created to protect Californians' consumer privacy. The CPPA implements and enforces the California Consumer Privacy Act. It is governed by a five-member board that consists of experts in privacy, technology, and consumer rights.
[2] Cal. Civ. Code § 1798.199.40(i).
[3] Cal. Civ. Code § 1798.100 et seq., https://cppa.ca.gov/regulations/pdf/cppa_act.pdf.
[4] See, California Privacy Protection Agency, California Consumer Privacy Act Regulations, https://cppa.ca.gov/regulations/consumer_privacy_act.html.

The CCPA as amended includes specific notice requirements for businesses, grants new privacy rights to consumers, and imposes corresponding obligations on businesses. The rights granted to consumers include the right to know what personal information businesses have collected about consumers and how that information is being used, sold, and shared; the right to delete personal information that businesses have collected from consumers; the right to stop businesses' sale and sharing of personal information; and the right to non-discrimination in service, quality, or price as a result of exercising their privacy rights. As of January 1, 2023, California consumers also have the right to correct inaccurate personal information that a business maintains about them, and the right to limit a business's use and disclosure of sensitive personal information about them to certain business purposes, among other protections. The CCPA provides additional protections for children under 16.

**The CCPA and Automated Decisionmaking Technology**

The Agency is currently engaged in pre-rulemaking activity on cybersecurity audits, risk assessments, and automated decisionmaking technology. The CCPA directs the Agency to issue regulations that (1) govern access and opt-out rights with respect to businesses' use of automated decisionmaking technology (ADMT), including profiling and (2) require "businesses' response to access requests to include meaningful information about the logic involved in those decisionmaking processes, as well as a description of the likely outcome of the process with respect to the consumer."[5] "Profiling" is defined in the CCPA as "any form of automated processing of personal information to evaluate certain personal aspects relating to a natural person and in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements."[6]

The CCPA also directs the Agency to issue regulations requiring businesses whose processing of consumers' personal information presents significant risk to consumers' privacy or security to (1) perform a cybersecurity audit on an annual basis; and (2) submit to the CPPA on a regular basis a risk assessment with respect to their processing of personal information.[7]

On February 10, 2023, the California Privacy Protection Agency issued an Invitation for Preliminary Comments on Proposed Rulemaking on Cybersecurity Audits, Risk Assessments, and Automated Decisionmaking.[8] The Agency received a total of 57 comments, totaling over 1,000 pages. Those comments are posted on the Agency's website.[9] The Agency has released draft language for discussion by the Agency Board with respect to cybersecurity audits[10] and risk assessments.[11]

---

[5] Cal. Civ. Code § 1798.185(a)(16).
[6] Cal. Civ. Code § 1798.140(z).
[7] Cal. Civ. Code § 1798.185(a)(15).
[8] California Privacy Protection Agency, Invitation for Preliminary Comments on Proposed Rulemaking, Cybersecurity Audits, Risk Assessments, and Automated Decisionmaking (Feb. 10, 2023), https://cppa.ca.gov/regulations/pdf/invitation_for_comments_pr_02-2023.pdf.
[9] California Privacy Protection Agency, Preliminary Rulemaking on Cybersecurity Audits, Risk Assessments, and Automated Decisionmaking, https://cppa.ca.gov/regulations/pre_rulemaking_activities_pr_02-2023.html
[10] California Privacy Protection Agency, Draft Cybersecurity Audit Regulations for California Privacy Protection Agency September 8, 2023 Board Meeting, https://cppa.ca.gov/meetings/materials/20230908item8.pdf
[11] California Privacy Protection Agency, Draft Risk Assessment Regulations for California Privacy Protection Agency September 8, 2023 Board Meeting, https://cppa.ca.gov/meetings/materials/20230908item8part2.pdf.

The latter includes draft definitions of artificial intelligence and automated decisionmaking, because a potential threshold under consideration for conducting a risk assessment is training artificial intelligence or ADMT. In addition, another potential threshold under consideration for conducting a risk assessment is the use of ADMT in furtherance of a decision that results in the provision or denial of financial or lending services, housing, insurance, education enrollment or opportunity, criminal justice, employment or contracting opportunities or compensation, healthcare services, or access to essential goods, services, or opportunities.

The draft definition of artificial intelligence is:

> [A]n engineered or machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments. Artificial intelligence includes generative models, such as large language models, that can learn from inputs and create new outputs, such as text, images, audio, or video; and facial or speech recognition or detection technology.

This definition was crafted to align with the National Institute of Standards and Technology's (NIST) Artificial Intelligence Risk Management Framework's[12] and the European Parliament's version of the EU AI Act's definition.[13]

The draft definition of ADMT is:

> [A]ny system, software, or process—including one derived from machine-learning, statistics, other data-processing techniques, or artificial intelligence—that processes personal information and uses computation as whole or part of a system to make or execute a decision or facilitate human decisionmaking. Automated Decisionmaking Technology includes profiling. "Profiling" means any form of automated processing of personal information to evaluate certain personal aspects relating to a natural person and in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements.

The Agency expects to release draft language with respect to ADMT opt-out and access rights, and to begin the formal rulemaking process with respect to these topics, in the coming months.

**Additional California efforts on AI and ADMT**

California is working on other fronts as well to explore AI and ADMT policy. California Governor Newsom has recently issued an executive order to direct certain state entities, among other steps, to

---

[12] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework at iii (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[13] Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts at 137 (May 16, 2023) https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf.

evaluate potentially beneficial use cases for deployment of generative AI tools by the state and to issue general guidelines for public sector procurement, uses, and required trainings for use of generative AI.[14]

Further, California's Civil Rights Council is currently undertaking formal rulemaking to amend its existing employment regulations to address automated decisionmaking. The proposed regulations state that "[i]t is unlawful for an employer or other covered entity to use selection criteria (including a qualification standard, employment test, automated-decision system, or proxy) if such use has an adverse impact on or constitutes disparate treatment of an applicant or employee or a class of applicants or employees on a basis protected by the Act."[15] Thus, California joins New York City and Illinois in taking steps to regulate the use of AI in hiring decisions.[16]

Finally, the California legislature is currently considering AB 302, which would require the state to compile a comprehensive inventory of all high-risk automated decision systems currently being used or are under consideration for adoption by state entities.[17]

California is a world leader in the development of AI technologies. Appropriately, California is also continuing develop innovative guardrails that support technological advances while seeking to ensure that these technologies are used safely. This innovation is also consistent with interoperability. As the Agency works to draft regulations with respect to AI and automated decisionmaking technology, it is carefully consulting other models to harmonize with laws in other jurisdictions. In addition to working with other jurisdictions to ensure consistent protections, the Agency is also in regular contact with state, federal, and international regulators to help ensure consistent enforcement of the law.

We hope that our work in implementing the CCPA is helpful to you as you consider AI and ADMT. Please do not hesitate to contact me if I can be of any assistance.

Sincerely,

Ashkan Soltani
Executive Director

cc: The Honorable Dick Durbin, Chair
The Honorable Lindsey Graham, Ranking Member
Members, United States Senate Committee on the Judiciary

---

[14] Executive Department, State of California, Executive Order N-12-23 (Sept. 6, 2023), https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf.
[15] Civil Rights Council Proposed Modifications to Employment Regulations Regarding Automated-Decision Systems (Feb. 10, 2023), https://calcivilrights.ca.gov/wp-content/uploads/sites/32/2023/02/Attachment-C-Proposed-Modifications-to-Employment-Regulations-Regarding-Automated-Decision-Systems.pdf.
[16] New York City Department of Consumer and Worker Protection, Automated Employment Decision Tools (Updated) (Jul. 5, 2023), https://rules.cityofnewyork.us/wp-content/uploads/2023/04/DCWP-NOA-for-Use-of-Automated-Employment-Decisionmaking-Tools-2.pdf; 820 ILCS 42, Artificial Intelligence Video Interview Act (2020), https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68.
[17] AB 302 (2023), https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB302.




September 11, 2023

Chairman Richard Blumenthal
Ranking Member Josh Hawley
Senate Judiciary Committee
Subcommittee on Privacy, Technology, & the Law
Washington, DC

Dear Chairman Blumenthal, Ranking Member Hawley, and Committee Members:

We write to you to express our support for the US AI Act, which we understand you will introduce this week at the hearing *Oversight of A.I.: Legislating on Artificial Intelligence*.[1]

The Center for AI and Digital Policy (CAIDP) is an independent research organization, based in Washington, DC.[2] We advise national governments and international organizations regarding artificial intelligence and digital policy. We previously testified on AI policy before the House Oversight Committee on *Advances in AI: Are We Ready For a Tech Revolution?*[3]

CAIDP strongly supports AI policies that advance democratic values and promote broad social inclusion based on fundamental rights, democratic institutions, and the rule of law, and therefore welcome your introduction of the US AI Act. For the last several years, we have published a detailed report – the ***AI and Democratic Values Index*** – that evaluates national AI policies and practices in 75 countries based on a dozen key indicators.[4] We have specifically urged the United States to move forward comprehensive legislation for the governance of AI.[5]

---

[1] Senate Judiciary Committee, Oversight of A.I.: Legislating on Artificial Intelligence, September 12, 2023, https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-legislating-on-artificial-intelligence
[2] CAIDP, https://www.caidp.org
[3] Testimony and statement for the record of CAIDP President Merve Hickok, *Advances in AI: Are We Ready For a Tech Revolution?* House Committee on Oversight and Accountability: Subcommittee on Cybersecurity, Information Technology, and Government Innovation (Mar. 8, 2023), https://oversight.house.gov/wp-content/uploads/2023/03/Merve-Hickok_testimony_March-8th-2023.pdf;
[4] *Artificial Intelligence and Democratic Values* (CAIDP 2023)
[5] Marc Rotenberg and Merve Hickok, *Artificial Intelligence and Democratic Values: Next Steps for the United States,* Council on Foreign Relations, August 22, 2022, https://www.cfr.org/blog/artificial-intelligence-and-democratic-values-next-steps-united-states;




We previously submitted several statements to your Subcommittee with related recommendations.[6]

We understand from the summary of the AI Act and the comments of the Sponsors that the US AI Act will: [7]

- Establish a licensing regime administered by an independent oversight body
- Ensure legal accountability for AI harms
- Defend national security and international competition
- Promote transparency
- Protect consumers and kids

We endorse the provisions as set in the preliminary Bipartisan Framework for the U.S. AI Act, and particularly the section to Promote Transparency, which will (1) require developers to disclose essential information about AI systems, (2) give users an affirmative right to know when they are interacting with an AI system, (3) promote watermarking and other technical measures to identify deepfakes, and (4) establish a public database about AI model information. We also believe that individuals must have the right to contest adverse decisions made by an AI system. That will require access to the logic and the date that produced the outcome.

You have also proposed a strong and effective independent enforcement agency that will establish licensing requirements and include the registration of information about Al models and be conditioned on developers maintaining risk management, pre-deployment testing, data governance, and adverse incident reporting programs. The oversight body will have the authority

---

Marc Rotenberg and Merve Hickok, *Regulating A.I.: The U.S. Needs to Act, The New York Times*, March 6, 2023, https://www.nytimes.com/2023/03/06/opinion/letters/alex-murdaugh.html#link-7f9159ab

[6] CAIDP, *Statement to US Senate Judiciary Committee on Oversight of AI.I.: Principles of Regulation* (July 21, 2023), https://www.caidp.org/app/download/8468551763/CAIDP-EJ-Statement-SJC-AI-07212023.pdf1 CAIDP, *Statement to US Senate Judiciary Committee on Oversight of A.I.: Rules for Artificial Intelligence* (May 11, 2023), https://www.caidp.org/app/download/8457859963/CAIDP-Statement-SJC-AI-05112023.pdf; *CAIDP Advises US Senate on AI and Human Rights* (June 13, 2023), https://www.caidp.org/app/download/8462575863/CAIDP-SJC-06132023.pdf?t=1694461301

[7] *Bipartisan • Framework for U.S. Al Act, Senator Richard Blumenthal & Senator Josh Hawley, Chair and Ranking Member of the Subcommittee on Privacy, Technology, and the Law* https://x.com/SenBlumenthal/status/1700147410880569475?s=20




to conduct audits of companies seeking licenses and will monitor and report on technological developments and economic impacts of AI. We also strongly recommend impact assessments prior to deployment for any AI system, not only sophisticated general-purpose AI models, that implicate civil rights or public safety.

We further support a liability regime that ensures legal accountability for harms, though we ask you to consider that a liability regime to ensure compliance with the fairness and safety obligations set out in the legislation will be more effective. Recognizing the danger of catastrophic risk with AI system, by the time harms emerge it may already be too late. Just as restaurants may be fined for maintaining unsafe kitchens without any harm to particular diners and drivers may be fined for exceeding speed limits before damage to property or life occurs, AI developers should also be subject to fines if they fail to comply with necessary safety standards.

We support your proposal to address AI risks to consumers and kids though we recommend a stronger emphasis on data minimization and accountability methods, rather than simply notice about the use of AI systems. There is a growing risk to the American public of unrestricted data collection that will fuel the large AI models of a small number of firms. The safeguards to address that challenge will need to be robust. Limits on data collection is the starting point.

**The Universal Guidelines for AI**

We would also like to call your attention to the Universal Guidelines for AI. In 2018, over 330 leading experts and 60 associations (including the AAAS, the ACM, and the IEEE) endorsed the **Universal Guidelines for AI** (UGAI). The Universal Guidelines for AU seek to maximize the benefits and minimize the risks of AI.[8] The Universal Guidelines for AI set out 12 principles that should be included in the US AI Act. (see attached) We note that several of these principles are incorporated in the draft US AI Act. We ask you to also consider such principles as the Termination Obligation that should be enacted into law. That principle states simply:

> **An institution that has established an AI system has an affirmative obligation to terminate the system if human control of the system is no longer possible.**

[8] Public Voice, *Universal Guidelines for AI (2018),* https://thepublicvoice.org/ai-universal-guidelines/

Center for AI and Digital Policy

We believe this principle, from the **Universal Guidelines for AI** (2018), may be the single most effective safeguard to address concerns about catastrophic risk

**The CAIDP OpenAI Complaint**

We also call your attention to the complaint CAIDP filed with the Federal Trade Commission regarding OpenAI. The FTC has now opened investigation of OpenAI we requested.[9] This is clearly a positive development, but we also need the FTC to prioritize this investigation. It took two years from the time we filed similar complaints with the FTC concerning Google and Facebook before there was a settlement.[10] We can't wait that long this time. AI products are evolving rapidly and being deployed downstream in consumer facing services. As several SJC Members are also members of the Senate Commerce Committee, we urge you to press the FTC on this matter.

Thank you for your attention to our statement. We look forward to the release of the text of the AI Act and would welcome the opportunity to testify at a future hearing.

Sincerely,





Merve Hickok
CAIDP President

Marc Rotenberg
CAIDP Executive Director

[9] Cecilia Kang and Cade Metz, F.*T.C. Opens Investigation Into ChatGPT Maker Over Technology's Potential Harms,* The New York Times, July 13, 2023, https://www.nytimes.com/2023/07/13/technology/chatgpt- investigation-ftc-openai.html; John D. McKinnon and Ryan Tracy, *ChatGPT Comes Under Investigation by Federal Trade Commission,* Wall Street Journal, July 13, 2023, https://www.wsj.com/articles/chatgpt-under-investigation-by- ftc-21e4b3ef.

[10] Federal Trade Commission, *Facebook Settles FTC Charges That It Deceived Consumers By Failing To Keep Privacy Promises*, November 29, 2011, https://www.ftc.gov/news-events/news/press-releases/2011/11/facebook-settles-ftc-charges-it-deceived-consumers-failing-keep-privacy-promises

Center for AI and Digital Policy

