

Mr. Ryan Calo, Law Professor, University of Washington

Questions Submitted by Members of the Senate Committee on Commerce, Science, and  
Transportation

Enlisting Big Data in the Fight Against Coronavirus

April 9, 2020

**Chairman Wicker**

1. Many national and local governments around the world are seeking to use new technology to combat this unprecedented pandemic. Earlier this week, the German government launched an app that allows users to “donate” personal data collected by their fitness trackers or other health devices to help authorities analyze the spread of COVID-19. Authorities in Moscow have launched an app intended to be downloaded by those who test positive for COVID-19. Yet this app raises privacy concerns, as it would allow officials to track residents’ individual movements.

As governments seek to use new technologies in the fight against COVID-19, it is imperative that privacy rights be protected. Are there specific examples of app-based programs you can recommend to policymakers that are both useful in the fight against COVID-19 and respectful of individual privacy rights?

Mr. Chairman, I have been impressed with the attention to privacy and security of several teams engaged in the development of software applications (“apps”) in the United States and Europe to combat coronavirus. In my testimony I mention the Decentralized Privacy-Preserving Proximity Tracing from Europe. There are similar efforts at MIT, Stanford, and my own institution the University of Washington. Nevertheless, perfect privacy and security is likely unattainable, and many experts have expressed skepticism that truly privacy-friendly contact tracing apps will be effective. In some cases, such apps may even do more harm than good by providing false reassurance to participants or enabling malicious political or economic fraud.

Consider, for example, the platform for anonymous contact tracing recently announced by Apple and Google. The platform supports apps that would allow participants to upload their health status without disclosing their identity or location and nevertheless alert other participants who may have come into contact with them. This approach is privacy-conscious but faces a number of practical limitations. First, the arrangement is voluntary, such that participants may encounter other infected persons who are not using the app. Alternatively, they may encounter infected participants who are asymptomatic but still contagious. Second, the arrangement is self-reported—or at least the platform contains no mechanism by which to verify health status. This could lead to false positives, including by malicious actors as I discuss in my testimony. And third, the arrangement involves self-help. We would have to trust participants to self-quarantine and reset that quarantine another 14 days whenever they receive a new notification.

Mr. Ryan Calo, Law Professor, University of Washington

The reality is that no clever app, standing alone, will get us out of the Hobson's choice American's face between sheltering in place and risking infection. The foreign jurisdictions that have been successful in containing the virus use technology and information as part of a comprehensive strategy that includes wide-spread, rapid-results testing, aggressive contact tracing by health officials, and mandatory quarantines of the exposed. Other uses cases involve digital badges that verify a person is immune from coronavirus coupled with government checkpoints. This kind of response takes significant government resources and will inevitably have a toll on civil liberties.

2. Much of the discussion surrounding the collection of private data to fight the spread of COVID-19 presents two goals – effectiveness and privacy protection – as mutually exclusive factors that need to be balanced. On one side of the balance, it is assumed that greater amounts of personal data, in more granular form, will allow authorities to track the spread of the virus more effectively. On the other side of the balance is protection of individual privacy, which is believed to be threatened by greater surveillance of individuals by the government.

Is this an accurate view of the situation? Are privacy and effectiveness always part of a trade-off, such that the most effective public health measures will come at the expense of privacy, and vice versa? Or do you believe that the most effective policies for combatting COVID-19 can also respect individuals' privacy?

I believe that *individual* surveillance for purposes of combatting the spread of COVID-19 will inevitably involve some trade offs to privacy and civil liberties if they are to be effective. Such trade offs may be worthwhile, assuming we safeguard privacy and civil liberties by promoting accountability and limiting mission creep or secondary use of that data. But trade offs will exist. There are, however, many important ways that we can use information in an aggregated or anonymized form to make better decisions about coronavirus or to study the virus itself. Although I am skeptical of the Apple-Google contact-tracing platform for the reasons I describe in my testimony and answers, I applaud the Google COVID-19 Community Mobility Report as shedding light on social distancing compliance across the country and the world. Moreover, I believe that health researchers absolutely need access to data to better understand the virus—again, with appropriate safeguards in place to avoid needless privacy harms.

3. Professor Calo: In the United States, the mobile advertising industry and technology companies are collecting consumers' smartphone location data to track the spread of COVID-19 and compliance with social distancing measures. The location data is purported to be in aggregate form and anonymized so that it does not contain consumers' personally identifiable information.

Mr. Ryan Calo, Law Professor, University of Washington

How can the use of anonymized, de-identified, and aggregate location data minimize privacy risks to consumers? And, what additional legal safeguards should be imposed on the collection of this data to prevent it from being used or combined with other information to reveal an individual's identity?

My own view is that aggregated data can be a useful tool in combatting coronavirus through better-informed health policy. It is important to note, however, that location information in particular can be self-identifying even if stripped of conventionally personally identifiable information. Where a person is and goes is unique to them and very telling, as the Supreme Court reminded us recently in the *Carpenter* case involving historic location information held by common carriers. Members of this Committee may be surprised, moreover, at how clever security researchers and some adversaries are at re-identifying supposedly anonymized data. But generally speaking, aggregated data can and should inform health policy where appropriate.

4. Professor Calo: As technology companies share anonymized location data with the U.S. government to support COVID-19 response efforts, to what extent should purpose limitation principles apply to the use and analysis of this data? And, when the pandemic finally passes, what should be done with any anonymized or de-identified data – and identifiable data, if applicable – collected by technology companies and the government for the purpose of addressing the public health crisis?

I am generally concerned with the prospect of mission creep and secondary use. I favor explicit safeguards that ensure that data collected and shared for the purpose of combatting coronavirus is only used for that purpose by government or industry, and that government powers accrued in connection with the pandemic are limited in time and scope to addressing public health crises. Personally I would like to see these safeguards enshrined in federal privacy legislation, which I know this Committee has championed under your and the Ranking Member's leadership.

Thank you for these excellent questions and again for the opportunity to testify.

**Sen. Thune**

5. More and more Americans all throughout the country are turning to online video services to conduct their jobs, education, and social interactions in an effort to practice social distancing. For instance, Zoom Communications had more than 200 million daily users last month. It was found that thousands of Zoom's calls and videos have been exposed to other users online and log-in information has been stolen resulting in many individuals' personal information being compromised.

Did Zoom's privacy policy clearly outline what types of information its platform would collect on individuals? If not, what transparency requirements should be in place for companies like Zoom?

Mr. Ryan Calo, Law Professor, University of Washington

Senator Thune, I have expressed concerns that Zoom in particular was not prepared from a privacy and security perspective to handle this wholesale migration of work, learning, and play to its platform. I would note that most household name technology companies—from Twitter to Google to Uber to Facebook—are presently under a consent decree with the Federal Trade Commission for privacy and security lapses. Without opining on the specifics of Zoom’s privacy policies or practices, I would encourage this body to ask the FTC to conduct an audit of Zoom and other, similar platforms to determine industry best practices and take any action necessary to ensure compliance under Section V of the FTC Act.

Americans are connecting with each other via online services across all 50 states. Would a patchwork of state laws benefit consumers and better protect their privacy? Should the United States enact a national privacy standard to safeguard consumer’s information?

I believe that a federal privacy law is past due in the United States and applaud this Committee for its robust engagement with these issues. Whether a privacy law would improve on the approach taken by individual states would of course depend on the strength of that law and the gains in compliance from nation-wide uniformity.

6. Without a federal privacy law in place, the American people must rely on the promises of tech companies that all have varying degrees of commitment to maintain consumers’ privacy.

How do we ensure that organizations are actively engaging in data minimization and strategic deletion practices after data is used or transferred?

In my own work, I have emphasized the broad powers of the Federal Trade Commission to police against privacy and security abuses. Infusing the FTC with adequate resources—including a bureau with deep technical expertise, akin to the FTC Bureau of Economics but for Technology—strikes me as crucial. The specific safeguards you mention could come about in at least three ways: the FTC could functionally require data minimization and retention limits as necessary to avoid unfairness and deception under Section V of the FTC Act, the FTC could be empowered to promulgate rules requiring data minimization and retention limits, or Congress could pass a law that so requires.

7. The country of Israel, through its internal security service, has reportedly used smartphone location based contact tracing to notify citizens via text that they have been in close proximity to someone infected with COVID-19, and ordering them to self-isolate for 14 days. A recent opinion piece in the Scientific American urged democratic governments to quickly follow Israel’s lead (see [“As COVID-19 Accelerates, Governments Must Harness Mobile Data to Stop Spread”](#)).

Mr. Ryan Calo, Law Professor, University of Washington

Please provide your thoughts on smart-phone location based contact tracing in light of the extraordinary privacy and other civil liberties concerns such an approach raises for U.S. citizens.

In my testimony and answers, I have stressed that digital contact tracing, to be effective, will require backing by government and therefore will implicate civil liberties. The American people through their representatives may decide that these extraordinary times call for invasive measures in order to slow and contain the spread of coronavirus. For example, some Americans may embrace testing and reporting requirements, mandatory quarantine, and “badges” that indicate who is free of coronavirus or possess antibodies against it. I am not an elected official and so it is hard for me to speak on anyone’s behalf but my own. What I want to emphasize is that effective technical measures to address COVID-19 are going to require significant investment of government resources and palpable trade offs to civil liberties. Moreover, should America follow the lead of Israel and other nations, I believe Congress should explicitly provide accountability mechanisms to guard against overreach and to limit the measures in time and scope to addressing the present health crisis.

According to the [Wall Street Journal](#), MIT is developing a contact tracing app for COVID-19 patients and others who have not been infected by COVID 19 that can be voluntarily downloaded to a person’s smart-phone. Please provide your views on this approach to contact tracing.

I believe voluntary, self-reported, and self-help approaches to digital contact tracing such as MIT’s are likely to prove ineffective and could perhaps do more harm than good. Imagine that a person downloads a given contact tracing app and attempts to use it to accomplish the apparent goal of leaving their home. In one scenario, they come into contact with a person who is infected but doesn’t also use the app, so they are not warned. In another, they come across a person who uses the app, is infected, but is one of the up to 25% of people who show no symptoms of the virus. In yet another, they are warned not to go near their local polling place on election day because a brazen political operative has downloaded the app and falsely reported being infected. Or they are warned to avoid their favorite restaurant or grocery store because a desperate, unscrupulous competitor has falsely reported being infected.

As I have emphasized in my testimony and answers, no clever app standing alone is going to get us out of the present health crisis. In order to be effective, digital contact tracing would have to be accompanied by significant investment of government resources in the form of wide-spreading, rapid testing, investigation of positive cases of COVID-19, and the imposition of mandatory quarantine for the infected and exposed. Perhaps Americans will embrace this expenditure and trade off to civil liberties in exchange for containment of coronavirus. But I am skeptical that apps like MIT’s will form a meaningful part of the pandemic response.

Mr. Ryan Calo, Law Professor, University of Washington

8. COVID-19 has caused private companies to seek out and utilize health data in an effort to protect users, employees, and the general public from the spread of the virus. Both Apple and Alphabet have released websites to help users self-screen for exposure to COVID-19. This data will be used to help public health officials. However, these tools also allow technology companies access to user's health information which the companies could in turn profit from in the future.

How are technology companies balancing the need for timely and robust reporting to prevent the spread of the virus with the confidentiality and privacy of the participants?

This is a great question but I am not sure I know the answer. There are federal laws, such as the Health Insurance Portability and Accountability Act (HIPAA), that require certain safeguards for covered entities in collecting, processing, and sharing protected health information. Even those are being relaxed to a degree during this pandemic; the Health and Human Services Office for Civil Rights (OCR) has announced that it will not enforce the HIPAA Privacy Rule even against covered entities engaged in "good faith" efforts at COVID-19 testing. Meanwhile, it is not clear what activities by tech companies are covered by HIPAA. In that case, technology companies are held to the promises they make and not much more.

What safeguards are in place to ensure data collected as part of the fight against COVID-19 are not sold to business partners or used for the development of other commercial products?

I agree with Ranking Member Cantwell that an accountability framework is needed to avoid "secondary use," which is the consumer privacy term for taking data collected for one purpose and redeploying it for another without adequate permission. At present, technology companies such as Apple and Alphabet (parent company to Google) are held bar to the promises the company has made to consumers by the Federal Trade Commission, state attorneys general, and the plaintiffs, and not much more.

It may be worth noting that the issues here are not limited to consumer data as such. Apple and Google have created a platform with an application programming interface ("API") nominally designed for digital contact tracing of coronavirus through Bluetooth. Yet this infrastructure supports many commercial functions as well (e.g., the start up Tile for locating lost items).

9. Anonymization techniques are also critical for safeguarding consumers' privacy. Truly anonymized data can protect a consumer's personal information, like their geolocation, political opinions, or religious beliefs. How do companies guarantee that every dataset they are storing contains truly anonymous data? And is the ability to re-identify data a part of the discussion in data-sharing arrangements?

Mr. Ryan Calo, Law Professor, University of Washington

Companies use a wide variety of anonymization techniques and some data-sharing agreements contain obligations not to re-identify. Neither the techniques nor including the contract language is obligatory under law, although standards such as differential privacy have emerged as best practice. Security research has shown that complete anonymity is rare, however, especially where the underlying data involves location over time. My own privacy research suggests that even if data were irreversibly anonymized, there can still be harms to the consumer. Imagine, for example, that a consumer has a gambling problem or is trying to quit smoking. If an advertiser is able to reach that consumer on the basis of their vulnerability, does it matter that the advertiser does not know the person's individual identity? The same could be true of COVID-19 status.

Thank you for your great questions.

**Sen. Blunt**

Your Committee has prioritized drafting federal privacy legislation for the purpose of creating clear, baseline definitions and standards for data collection, storage, and use across industry sectors. Similarly, the bills before this Committee attempt to create definitions to meet appropriate levels of consent and transparency for protecting consumers' privacy and security. I applaud these efforts.

In relation to COVID-19, the end users of specific data sets, like location data, are more likely to be governmental entities than commercial entities. Big data can be an incredible tool to better understand the spread of the virus, and the impact on communities across the country. Data can help identify resource deficits, inform governments and health care professionals to employ countermeasures at the appropriate time, and provide insight to the downstream economic effects of this pandemic.

However, the U.S. commercial entities that would collect this data have very few guardrails on the collection and distribution of this data. Similarly, there are few requirements or regulations at federal or state level that require or even identify methodologies for anonymizing or pseudonymizing data. De-identifying data may result in greater data privacy and data security for consumers or individual citizens, but relies heavily on all of the entities involved in the collection and storage of that data making decisions based on best practices.

10. What efforts do you recommend that federal agencies undertake to ensure that data being used to track viral spread are upholding the highest possible standards for individual privacy and security?

While a full response to this question may not be possible in the time and space allotted, Senator Blunt, I would emphasize the need for data minimization, adequate security, clear accountability, and express limitations on secondary use or its national security analog, "mission creep." Here are some principles to keep in mind:

Mr. Ryan Calo, Law Professor, University of Washington

- First, federal agencies should have a clear reason for collecting individual information—what is the use case in fighting the pandemic, and is it plausible?
- Second, federal agencies should only collect that information needed to accomplish their stated purposes.
- Third, data should be anonymized prior to passing between government and the private sector unless doing so would thwart a legitimate government purpose.
- Fourth, data should be secured physically and digitally in accordance with its sensitivity. Location and health data are particularly sensitive.
- Fifth, clear rules should be in place to the effect that data can only be used for the purpose collected and that any added emergency powers should only be exercised for combatting the public health crisis.
- And finally, federal agencies should have clear accountability structures in place to guard against abuse of these principles.

In short, I agree with the opening statement of the Ranking Member to the effect that the response to the pandemic should operate within a clearly stated framework to help ensure privacy and accountability.

11. Does data lose any utility when it is de-identified or anonymized? Is it possible to have large data sets that are not tied to individual's identities, but which would still be useful for governments or public health-related end users?

Data does not lose all utility merely by being de-identified or anonymized. Neither process is infallible, but best practice suggests aggregating or de-identify personal data whenever doing so preserves the basic utility of that data. Moreover, large data sets could help policymakers make wiser decisions during the pandemic. I mention one example—Google's regional reports on social distancing—in my testimony. Please note, however, that while large data sets and the techniques used to make sense of them can be useful, they are not infallible. My testimony also recommends humility in applying big data to addressing pandemic and gives cautionary examples involving artificial intelligence to predict flu trends. I also point to work by a variety of scholars indicating the ways that big data can disproportionately harm vulnerable or minority individuals and communities.

12. It is important to me that as government entities access commercially-collected or publicly available data, that those efforts are giving reasonable consideration to protecting individual privacy and security. Are there any technologies that offer the opportunity to collect data that would be useful to a governmental pandemic response efforts, without resorting to surveillance methods that jeopardize individual privacy – like those which have been used recently by foreign governments?

Mr. Ryan Calo, Law Professor, University of Washington

Yes and no. There are techniques—such as differential privacy, or the data collective model developed at the University of Washington using Microsoft technology—that help assure that governments can only ask certain types of questions about data sets and not others. But when it comes to, for example, tracking down who has been exposed to coronavirus at an individual level and forcibly quarantining them as other nations have done, then the trade offs to privacy and civil liberties seem significant and inevitable. Still, judicial supervision and other accountability measures can be put into place that help to mitigate the impact to civil liberties.

Thank you for the opportunity to answer your excellent question.

**Sen. Cruz**

13. A little over two weeks ago, the Johns Hopkins Center for Health Security published a report titled *“Modernizing and Expanding Outbreak Science to Support Better Decision Making During Public Health Crises: Lessons for COVID-19 and Beyond.”* Although full of thought provoking ideas, one of the most notable was a recommendation to establish a “National Infectious Disease Forecasting Center,” similar to the National Weather Service. Much like the National Weather Service, this new infectious disease forecasting center would have both an operational role—providing the best modeling and forecasting to policy makers and public health professionals before, during, and after a disease outbreak—as well as a research role—providing a venue for academic, private sector, and governmental collaboration to improve models and encourage innovation.

What do you all think of this idea, and what do you all think the positives and negatives would be if such a concept was operationalized?

Senator Cruz, I was not specifically aware of this recommendation and appreciate your calling my attention to it. There can be tremendous utility to greater accuracy in forecasting outbreaks of deadly viruses such as COVID-19. In my testimony, I describe several other efforts. The positives include the ability to respond more rapidly to health crises, to better direct limited resources where they are needed, and generally to promote knowledge generation and sharing among academia, government, and private industry.

The main negative I envision (apart from the budgetary impact) is that forecasting complex phenomena is notoriously difficult. I address some of the limitations of artificial intelligence and other techniques of data analysis on page three of my testimony. A model can appear to work well for a time, leading policymakers to rely upon it to make critical decisions around resource allocation, only to fail down the line. Such was the case with Google Flu Trends, which applied complex mathematical to user search terms to successfully predict the incident of flu around the time of H1N1, only to break down in accuracy just a few years later.

It is also important to note that a lack of political will is sometimes the greater hurdle than a lack of information. But in general I see a lot of upside to our Johns Hopkins colleagues’ proposal.

Mr. Ryan Calo, Law Professor, University of Washington

14. One of the big reasons weather forecasting works, if not the biggest, is how many observations—things like water temperature, barometric pressure, radio profiles of the atmosphere, etc.—are fed into the weather model. Now while collecting ocean temperatures from buoys, or pressure readings from weather balloons, doesn't really raise privacy concerns, collecting health observations almost certainly would.

How can we thread the needle—either in this concept or private sector modeling—of getting enough of the right kind of data to accurately model infectious disease outbreaks while still protecting the privacy and security of individuals?

I really appreciate this question. Threading that needle is hard, but there are techniques I have mentioned in my testimony and answers that help ensure the privacy and security of individuals during predictive model. One is differential privacy, a system that allows for the sharing and even publication of data sets while mathematically reducing the likelihood of identifying individual members of that data set. But great care is still needed. Weather forecasting is improving in part due to recent gains in artificial intelligence. AI is often described as a “black box” into which individual data disappears. Recently, however, the field of adversarial machine learning has shown how training data can be extracted from AI systems through clever queries. My best advice is to bring in computer security specialists such as my colleagues at the University of Washington's Privacy and Security Lab at the earliest stages of design or procurement to help threat model government and private efforts to model infectious disease.

15. To date the State of Texas has reported thousands of cases of coronavirus, and hundreds of deaths related to complications from infection. To mitigate the risk of infection in Texas and across the country, the administration has restricted international travel, provided more access to medical supplies by involving the powers of the Defense Production Act, and cut red tape to expand access to testing. Congress also passed the CARES Act which provided \$377 billion in emergency loans for small businesses and directed \$100 billion to hospitals and healthcare providers. However, I believe much still needs to be done to finish this fight and recover once this is behind us.

In your expert opinions, what more needs to be done to beat this virus, and how can federal, state, and local governments work with private companies to both mitigate spread of the virus—both now and later this summer or fall—and recover quickly once the threat of this virus has passed?

I fear I am not expert in the right ways to answer this question. My understanding is that widespread, rapid-result testing and intermittent social distancing until a vaccine can be developed will be needed to address the pandemic. But I defer to colleagues with greater experience in public health.

Thank you for the opportunity to answer these great questions.

**Sen. Moran**

16. Many of the discussed proposals related to utilizing “big data” to fight against the spread against coronavirus rely upon the concepts of anonymized and aggregated data to protect the personal identity of individuals that this information pertains to and prevent consumer harms that could result. As such, many members on this Committee have spent significant time and energy drafting federal privacy legislation that tries to account for practices such as these that prevent harmful intrusions into consumers’ privacy while also preserving innovative processing practices that could utilize such information responsibly without posing risks. *That being said, do the witnesses have any policy recommendations for the Committee as it relates to effectively defining technical criteria for “aggregated” and “anonymized” data, such as requiring companies to publicly commit that they will refrain from attempting to re-identify data to a specific individual while adopting controls to prevent such efforts?*

Senator Moran, in my testimony and answers I have referred to techniques such as differential privacy that try to mathematically guarantee that large data sets will only yield certain kinds of answers, but not others. That said, I would not necessarily enshrine a specific technique into legislation because of the rapidly evolving nature of technology and security research. Rather, I would empower the Federal Trade Commission to continue to establish requirements for adequate security and to pursue poor privacy and security practices as violations of the FTC Act. I would also establish a research exception to the Computer Fraud and Abuse Act so that academic and independent researchers can help hold technology companies accountable without fear of legal reprisal.

17. Consumer data has tremendous benefits to society, as is clearly evident in the fight against the COVID-19 outbreak. Big data and the digitized processes and algorithms that technology companies are developing have led to an entirely new sector of the global economy. *Are you satisfied that the technology industry is striking an appropriate balance between producing services that better our ability to solve problems, as is clear in the fight against COVID-19, versus their production of products that increase their bottom line and generate profit? Are you satisfied that the United States government is striking an appropriate balance between supporting these companies in addressing COVID-19 versus ensuring we conduct adequate oversight of the industries’ activities?*

Due to the prospect that technology companies will place profit over privacy, I support federal privacy legislation such as the bills this Committee has considered. With respect to COVID-19, I believe the government should develop a plan to address the pandemic and then enlist technology companies where appropriate to help federal and local government carry out that plan. I worry that some government officials seem to be calling upon technology companies such as Google, Apple, and Facebook to come up with unspecified solutions, rather than describing a

Mr. Ryan Calo, Law Professor, University of Washington

specific government need. By way of contrast, the government has identified a clear need for ventilators and so has directed an auto manufacturer go carry through on plans to build ventilators. Government officials did not simply charge the auto industry with coming up with whatever technologies it felt would help in the pandemic.

18. Consumer trust is essential to both the United States government and to the companies whose products we use every day. We need to work to maintain that trust and ensuring that the big data being used to analyze the COVID-19 outbreak was collected and processed in a manner that aligns with our principles is important to my constituents. *How can we adequately ensure that the data being used to address COVID-19 is sourced and processed in a manner that ensures consumer trust is not being violated, while allowing the innovation and success we've seen continue to grow?*

In my testimony and answers, I have emphasized my concern that data collected for the purpose of combatting COVID-19 will later be monetized. I gave the example of COVID-19 immunity or its absence influencing the cost of insurance or the commercial offers a consumer encounters online. The consumer privacy term for repurposing data in this way is “secondary use.” I argue that Congress should explicitly prohibit secondary use absent affirmative consent by the data subject. This limitation will help engender trust.

It may be worth noting that the issues here are not limited to consumer data as such. In an apparent response to government calls for innovation around COVID-19, Apple and Google have created a platform with an application programming interface (“API”) nominally designed for digital contact tracing of coronavirus through Bluetooth. Yet this infrastructure supports many commercial functions as well (e.g., the start up Tile for locating lost items) that these companies could develop and monetize in the future.

19. It is important to remember that the internet is a global network and that no matter how secure we make our networks, they remain vulnerable to bad actors, corruption, and misguided influence from around the world. *Can you comment on the practices we've seen used by companies and international partners to ensure the data used to address COVID-19 is both accurately sourced and stored in a manner that is secure?*

This is an excellent question but, given the rapidly evolving nature of the international response to COVID-19, I am not familiar with individual practices of foreign companies in this context.

Thank you for the opportunity to address your great questions.

Mr. Ryan Calo, Law Professor, University of Washington

**Sen. Blackburn**

20. How do you see HIPAA interacting with your worldview of the tech industry?

This is a great question, Senator Blackburn, but I am not sure I know the answer. The Health Insurance Portability and Accountability Act (HIPAA) requires certain safeguards for covered entities in collecting, processing, and sharing protected health information. Even those are being relaxed to a degree during this pandemic; the Health and Human Services Office for Civil Rights (OCR) has announced that it will not enforce the HIPAA Privacy Rule against even against covered entities engaged in “good faith” efforts at COVID-19 testing. Meanwhile, it is not clear what activities by tech companies are covered by HIPAA. In that case, technology companies are held to the promises they make and not much more.

21. How do you envision working with the CDC to develop the updated surveillance system (which was given \$500 million in the recently passed CARES Act) while protecting health information and thereby allow CDC to use their expertise – epidemiology that inherently seeks to protect health information – with big tech’s powerful data collection and analysis tools?

In my testimony and answers, I have emphasized the importance of information and technology in responding to health crises. Insights from better “surveillance” (a term of art in public health) can help detect outbreaks earlier, direct limited resources to where they are most needed, and generally promote the generation of knowledge about infectious disease.

However, I would note again that data-driven analysis has its limitations. A model can appear to work well for a time, leading policymakers to rely upon it to make critical decisions around resource allocation, only to fail down the line. Such was the case with Google Flu Trends, which applied complex mathematical to user search terms to successfully predict the incident of flu around the time of H1N1, only to break down in accuracy just a few years later. Moreover there exists a well-documented tendency for the costs and benefits of artificial intelligence and big data to fall unevenly across the population.

In response to Senator Blunt’s question, I identified a series of principles that government should follow with respect to public health surveillance especially in partnership with privacy industry. These include: (1) having a clear sense of the legitimate purpose for which data is being collected or changing hands, (2) only collect information needed to accomplish that purpose, (3) anonymize data unless doing so would not thwart a legitimate government purpose, (4) (3) physically and digitally secure data, and (5) establish clear rules around secondary use, mission creep, and consequences for abuse.

Generally I agree with the Ranking Member in her opening statement where she calls for a clear privacy accountability framework.

Mr. Ryan Calo, Law Professor, University of Washington

22. Today we are giving into state surveillance for the sake of saving thousands of lives that might otherwise be lost to coronavirus. The CDC is already relying on data analytics from mobile ad providers to track the spread of the disease. How can we ensure the data collection will only be done for the limited purposes of the emergency, with safeguards to ensure anonymity? On retention time, when should the data be deleted? Who has the right to that deletion – the federal government or the individuals themselves? Most importantly, what duty do tech companies owe to protect consumer privacy, even during a global pandemic?

A full response to this excellent question may not be possible in the space and time allotted. In my testimony and answers, I have emphasized the inevitable trade offs between individualized, government-backed contact tracing and civil liberties. But the trade off may be worth the gains. The American people through their representatives may decide that these extraordinary times call for invasive measures in order to slow and contain the spread of coronavirus. For example, some Americans may embrace testing and reporting requirements, mandatory quarantine, and “badges” that indicate who is free of coronavirus or possess antibodies against it. I am not an elected official and so it is hard for me to speak on anyone’s behalf but my own. What I want to emphasize is that effective technical measures to address COVID-19 are going to require significant investment of government resources and palpable trade offs to civil liberties.

There may be ways to mitigate some of these harms. I have referred repeatedly to the importance of guarding against secondary use and mission creep, i.e., the persistence and migration of surveillance powers created in one context such as a global pandemic or terrorism into a new context such as narcotics trafficking. As I mentioned in my testimony, paraphrasing Justice Robert Jackson, a problem with emergency powers is that they tend to kindle emergencies. Specific accountability measures include: (1) limits on data retention, (2) sunset provisions for new surveillance powers, (3) prohibitions on secondary use absent affirmative consent from the data subject, (4) penalties for abuse, and (5) judicial oversight. But none of these are a panacea or likely entirely to avoid harms to privacy and civil liberties.

23. Mr. Calo: The US Department of Energy (DOE) has established a public-private consortium to focus its resources in high-performance computing, big data, and artificial intelligence on combatting the COVID-19 pandemic. From your perspective, how can these extraordinary capabilities be leveraged to accelerate our understanding of the SARS-CoV-2 virus, speed the development of treatments and vaccines for COVID-19, and contribute to ending this pandemic?

Artificial intelligence—the set of techniques aimed at approximating some aspect of human or animal cognition with machines—can be a powerful tool in combatting COVID-19. For example, AI can help test plausible drug compounds for treatment or spot patterns in morbidity

Mr. Ryan Calo, Law Professor, University of Washington

that affect how doctors and nurses respond to the novel coronavirus. AI may even be useful in allocating hospital resources by anticipating outbreaks and correlating hospital capacity with likely need. In my testimony, I refer to a United Nations and World Health Organization report detailing a wide variety of use cases for AI and some concrete examples. I am a big believer in data science, such as that conducted at the eScience Institute at the University of Washington where I work.

Nevertheless, I have cautioned in my testimony and answers repeatedly that a healthy dose of humility is needed when applying AI to problems of global health. First, AI models can appear to work for a time and then breakdown because of a change of conditions. Such was arguably the case with Google Flu Trends, which correctly anticipated flu during around the time of H1N1 but is not in use today because of subsequent failures to predict flu. Second, AI models can hold erroneous or biased assumptions, or be trained upon biased data. There is a well-documented tendency of AI to inure disproportionately to the detriment of vulnerable or minority populations, who may not be well represented in training data.

In my testimony, I invite the Committee to consider a hypothetical: “Imagine, for example, that public health officials were to allocate coronavirus resources on the basis of data trends from connected thermometers like Kinsa Health (retail cost: \$35.99 – \$69.99) or connected pulse oximeters like iHealth Air (retail cost: \$69.99). Only communities where sufficient numbers of consumers were aware of such devices and could afford them would receive an early warning or stockpiled support.”

24. Foreign countries like South Korea, Taiwan, Singapore, and Israel swiftly mobilized collection of cell phone location data to track the spread of the virus and map out infection hot zones. Israel just released an app that allows the public to track whether they have may visited a location that put them into contact with an infected individual. Is it even possible to adopt similar measures while still balancing protections for privacy and civil liberties?

I believe that individuated contact tracing that is mandatory and backed by the government inevitably involves trade offs to privacy and civil liberties. But I am also highly skeptical that any other form of contact tracing—such as the voluntary, crowd-sourced app that Apple and Google’s platform hopes to support—will be effective in lifting the necessity of social distancing. Indeed, some experts attribute the success of South Korea, Taiwan, Singapore, and Israel to wide-spread, rapid testing and early social distancing, more so than digital contact tracing per se. Nevertheless, should the United States decide to follow the lead of these foreign jurisdictions, we can and should put safeguards into place such as judicial oversight and express limitations on secondary use and mission creep discussed above. Thank you for the opportunity to answer these wise questions.

Mr. Ryan Calo, Law Professor, University of Washington

**Sen. Lee**

25. To date, what specific data (or types of data) are companies (or your company) currently collecting for COVID-19 related purposes? What specific data (or types of data) are governments and health officials seeking for COVID-19 related purposes?

Senator Lee, I do not have access to the full range of data collected by companies but it strikes me that individual health status (the presence or absence of the virus, virus symptoms, or antibodies), incidence of mild, severe, or deadly disease, location information, and hospital capacity are among the most salient categories of information.

26. Most tech companies currently claim that the data being gathered is being “anonymized” so that a specific person is not identifiable.

What specific steps are companies (or your company) taking to anonymize this data?

I do not have access to this information.

Certain data may not necessarily be considered personally identifiable, but with enough data points, you could identify a specific person. How can we ensure that data is truly anonymous and is not traceable back to an individual person?

Representatives from industry and technologists may be better positioned to answer this question. In my testimony and answers, I have referred to several techniques such as differential privacy that try to guarantee mathematically that a data base will yield useful insights without identifying individuals. But anonymization is notoriously difficult.

Can effective contact tracing be conducted with “anonymized data”? Or will it require personally identifiable information?

This remains an open question. The platform recently announced by Apple and Google supports contact tracing applications that neither identify participants personally nor store location information. Nevertheless, security experts have already begun to identify clever ways potentially to re-identify participants, and have cast doubt on the efficacy of a digital contact tracing system that keeps no record of location information. Generally speaking, the individuated, government-backed contact tracing occurring in some foreign jurisdictions, whereby government investigators investigate positive cases and enforce mandatory exposure on the exposed, likely could not function in a completely anonymized way.

Mr. Ryan Calo, Law Professor, University of Washington

27. Since the beginning of this COVID-19 crisis, has a federal agency, a state government, or local government requested a company or association to gather any specific consumer data?

To your knowledge, are there any current COVID-19 related data sharing agreements in place between governments and private sector organizations?

To your knowledge, has any federal, state, or local law enforcement used private sector collected data to enforce any COVID-19 related government orders or requirements?

This is a rapidly evolving situation, but I am not personally aware of specific requests by federal or local officials for consumer data or agreements related to data sharing. It would not surprise me were there to be several in place, however.

Thank you for the opportunity to answer these important questions.

**Sen. Scott**

For months, Communist China lied about the Coronavirus data, the spread of the virus, and their response. They silenced critics and those trying to alert the Chinese people to this public health crisis. The lack of usable data coming out of Communist China cost lives and put the world behind on response efforts, including here in the United States.

28. As we work to keep American families healthy, how can we follow the lead of countries with low case counts, like South Korea, using technology and data collection, without infringing on our citizens' rights and privacy?

Senator Scott, I do not doubt that technology and information will play a key role in combatting coronavirus. I believe that individuated contact tracing that is mandatory and backed by the government—which South Korea and other countries are trying—inevitably involves trade offs to privacy and civil liberties. That said, I am also highly skeptical that any other form of contact tracing—such as the voluntary, crowd-sourced app that Apple and Google's platform hopes to support—will be effective in lifting the necessity of social distancing. Some experts attribute the success of South Korea, Taiwan, Singapore, and Israel to wide-spread, rapid testing and early social distancing, more so than invasive tracking of mobile phones per se. Nevertheless, should the United States decide to follow the lead of these foreign jurisdictions, we can and should put safeguards into place such as judicial oversight and express limitations on secondary use and mission creep discussed in my testimony.

Thank you for the important question.

Mr. Ryan Calo, Law Professor, University of Washington

### **Ranking Member Cantwell**

29. Science and technology will be critical drivers of our response to COVID-19, and we have seen many examples of data being used in positive ways – from the University of Washington’s forecasts of hospital needs to Johns Hopkins’ maps of disease spread. These are leading examples of how firms can innovate while protecting other equities, like privacy. What recommendations do you have to encourage further innovation to fight the virus?

Ranking Member Cantwell, like you, I have been greatly impressed with the efforts at the University of Washington and other academic institutions across the state and the country in helping to address the current health crisis. The anticipated infusion of resources from the National Science Foundation (RAPID and EAGER) and the National Institutes of Health will accelerate academic innovation all the more.

I have also been impressed by many of the efforts of American technology firms to leverage aggregated data to shed light on the pandemic and its social impacts. For example, in my testimony I praise the Google COVID-19 Community Mobility Report and have since read reporting by the New York Times on social distancing based on data from the data intelligence firm Cuebiq. Finally, I read the testimony from Kinsa Health Thermometers with great interest.

Consistent with the framework you identify in your opening statement, there may be opportunities to suspend or alter regulations to support COVID-19 specific efforts. For example, I note that the Health and Human Services Office for Civil Rights (OCR) has announced that it will not enforce the HIPAA Privacy Rule against covered entities and their partners engaged in “good faith” efforts at COVID-19 testing.

My expectation as a citizen, however, is that federal and local government will come up with a plausible plan to address the public health crisis and then enlist companies such as Google and Apple as needed to effectuate this plan. The analogy I have in mind is to the recent directive to GM that it carry through on plans to manufacture ventilators due to the obvious national need. Of course, technology companies hold expertise in artificial intelligence and other useful technologies and techniques and should be part of the conversation on how best to combat COVID-19.

How do we encourage technologists to help people transition to regular life while preparing for future pandemic incidents? What are the best practices you have seen in innovating in the fight against COVID-19 that support privacy rights?

Mr. Ryan Calo, Law Professor, University of Washington

My understanding from speaking to subject matter experts is that it will not be easy for Americans to return to regular life short of a vaccine. Jurisdictions that have managed to contain coronavirus have done so primarily through social distancing and widespread, rapid-results testing. In my testimony and answers, I have noted that technology that supports contact tracing and health status verification (digital “badges” or “passports”) may also form an integral part of the American response if the public embraces them. But I have also repeatedly cautioned that, in order to be effective, such efforts involve significant and largely inevitable trade offs to privacy and civil liberties. Voluntary, self-reported, and self-help solutions such as those proposed by MIT faculty and supported through the new Apple-Google platform strike me as privacy conscious but unlikely to be effective in permitting people safely to leave their homes for the reasons I state in the my testimony.

30. Frequently, data used to combat COVID-19 is described as “anonymized” or “aggregated” or “de-identified,” and these terms are meant to convey that data will be used or shared in a privacy-protective manner. How do you define “anonymized,” “aggregated,” and “de-identified” data? What are the best practices to ensure that the data remains anonymous?

To me, the key distinction is whether the information has been aggregated, in the sense of combined with other information to tell a larger story about populations or trends rather than individuals, or merely de-identified or anonymized. The latter could involve, for example, stripping out conventionally personally identifiable information but still associating data with a unique identifier. Often such techniques allow for later re-identification—for example, if my travel history is unique enough to narrow down to me even without my name. A person better versed in techniques of anonymization could give more details. My understanding is that techniques exist—such as differential privacy—that can help assure that only certain kinds of questions can be asked of a data set. Sharing of data can also be conditioned contractually on a promise not to re-identify.

Thank you for these excellent questions, your leadership in the area of consumer privacy, and again for the opportunity to address this Committee.

**Sen. Blumenthal**

31. Privacy for America, a coalition of advertising associations including IAB and NAI, have proposed a federal privacy framework that is focused on a set of prohibited data uses, transparency measures, and a limited subset of data rights found under the GDPR and CCPA. However, the Privacy for America framework also provides wide discretion for companies to use particular types of data or engage in particular activities without consent, as well as a self-regulatory safe harbor and broad state preemption.

Mr. Ryan Calo, Law Professor, University of Washington

Would the Privacy for America framework provide Americans the full set of consumer rights and protections necessary to guarantee the privacy, security, and equitable use of their personal data, and the enforcement regime necessary to deter and punish the misuse of their information? Please elaborate on why or why not.

Senator Blumenthal, in my view, nothing short of federal privacy legislation that contains concrete safeguards against violating the privacy expectations of consumers and empowers the Federal Trade Commission, state attorney generals, and (ideally) individual litigants to police against abuse will be adequate. We have had self-regulation in most corners of consumer privacy for as long as I can remember and it has not been effective, which is why Americans report being more and more worried about privacy year after year. However, I am confident in the ability of an FTC—infused with more resources and with access to technology expertise on par with its Bureau of Economics—to continue to serve as the nation’s leading privacy and security watchdog.

Thank you for your question and your leadership on this issue.

**Sen. Markey**

32. As the coronavirus pandemic continues to affect every facet of American life and cause immense pain to individuals, families, and communities across our country, it is imperative that both the public and private sectors are laser focused on identifying and implementing innovative, evidence-based solutions. However, every step of the way, we must ensure that these solutions do not breach individuals’ civil liberties and cherished rights, including the right to privacy. Recent reports have indicated that private companies and government entities are using information about individuals’ location to combat the pandemic. Such information can be analyzed to determine how effective “stay at home” orders have been, for example. However, misuse of or inappropriate access to this type of information can have significant consequences for regular Americans. Mr. Calo, why is information about an individual’s location so sensitive, particularly when it is analyzed in conjunction with information about health status?

Senator Markey, I am hard-pressed to name two more sensitive categories of personal information than location and health status, let alone in combination. Location information has ramifications for personal security and, over time, lends a detailed picture of an individual’s activities. In the recent *Carpenter* case, the Supreme Court explained that location records “hold for many Americans the ‘privacies of life’” and that a government with access to historic location data “achieves near perfect surveillance.” Meanwhile, an individual’s health status constitutes a roadmap of their vulnerabilities. I have no doubt that abuse of location and health status information by governments or corporations would have significant negative impacts on citizens and consumers.

Mr. Ryan Calo, Law Professor, University of Washington

33. Schools across the country have closed over the past several weeks in order to protect students, faculty, staff, and broader communities during the ongoing pandemic. As a result, millions of students are now relying on software and online tools for their education. Many of these “ed tech” offerings collect vast amounts of students’ data. While the Family Educational Rights and Privacy Act (FERPA) and the Children’s Online Privacy Protection Act (COPPA) provide important safeguards to protect young students’ privacy, serious threats to kids’ educational information remain. Mr. Calo, how does the rise of “distance-learning” and dependence on ed tech during the ongoing coronavirus pandemic have the potential to increase risks of student privacy invasions?

There are myriad ways that a wholesale shift to online learning has the potential to increase the privacy risk to students. For example, online learning environments are mediated—they take place through a technology designed by a third party. This means that a corporation has the ability both to collect minute data about kids and to design every aspect of their interaction with the company, their teacher, and one another. In a series of articles, I have explored how mediation gives companies the means and even the incentive to exploit their asymmetric power over data and design for financial gain. Furthermore, tools such as Zoom—if not properly secured—create a channel into the lives of children that would not exist otherwise, allowing potential contact with bullies, “trolls,” or worse. There is also the concern—articulated by Julie Cohen, Elana Zeide, and others—that children need a measure of privacy for purposes of self-development. All of this is to say that vigilance and oversight will be needed to protect vulnerable populations such as children.

34. The Federal Trade Commission regularly issues guidance to companies to provide “plain-language guidance to help businesses understand their responsibilities and comply with the law.” This guidance can serve as an important resource for companies that aim to serve their customers responsibly and follow best practices. In March, I sent a letter to the FTC and the Department of Education with Senator Blumenthal and Senator Durbin, urging those agencies to jointly issue guidance to ed tech companies in order to protect student privacy. I am pleased that the FTC has heeded our call and took the interim step of providing guidance to ed tech companies and schools on how to comply with COPPA during the ongoing pandemic. I also urge the FTC to go further by collaborating with the Department of Education, informing parents of privacy best practices and encouraging caregivers to be aware of ed tech services that may provide different versions and educational opportunities depending on whether the parent grants permission for data collection and use. Mr. Calo, do you agree that the FTC and the Department of Education should issue this guidance to parents in order to protect students’ privacy during this pandemic and in the future?

I wholeheartedly agree that the Federal Trade Commission should work with subject matter experts such as the Department of Education to issue guidance around ed tech that goes beyond compliance with COPPA. Thank you for the opportunity to answer these important questions.

**Sen. Peters**

35. Earlier this week, I sent a letter to the Centers for Disease Control (CDC) asking them to publicly report all available information about who is able to access COVID-19 tests, which continue to be scarce. Most states and the federal government haven't released demographic data on the race or ethnicity of people who've tested positive for the virus. The CDC has included age and gender data that it has released daily since the pandemic began, but has not released racial or ethnic data. If we have all of the data, or included information on medical providers. If we have all of the data necessary, Congress and the federal government can direct resources to the areas that need it most.

Can you describe how we can best use racial or ethnicity data to help our nation's underserved communities that are being hit hardest by COVID-19 and what are the federal road blocks for obtaining this data?

Senator Peters, I agree that demographic data can help identify and address racial disparities with respect to rates of exposure, infection, and successful treatment of COVID-19. Like you, I have seen reporting to the effect that the coronavirus is having a disparate impact on vulnerable and minority populations. I cannot identify with confidence what all of the roadblocks may be to obtaining this data to the extent it rests in the hands of the federal government. Generally speaking, the Privacy Act of 1974 governs the circumstances under which federal agencies may share records regarding individuals, but does not serve as a prohibition on disclosing overall demographic trends to my knowledge.

36. The one thing that has been absent from this discussion is that neither the federal government nor the private sector have adequately anticipated nor met the demands for personal protective equipment. Even basic things like masks and gloves have been inaccessible. Our nation has unparalleled resources in the supply chain and manufacturing space.

From a data perspective—where have failures been and what improvements do you recommend?

In my testimony, I refer to several efforts to use statistics and machine learning to correlate high hospital demand with hospital capacity. Similar efforts could be used to anticipate shortfalls in personal protective equipment, invasive ventilators, and other resources, with the usual caveats that sometimes even initially successful predictive models can break down. I also want to disentangle what is a function of poor information, and what is perhaps a failure of logistics or political will. No amount of information can address the latter.

Mr. Ryan Calo, Law Professor, University of Washington

37. Despite many structural challenges, Taiwan has fared better than many countries in dealing with the COVID-19 pandemic. Stanford Medical School documented 124 distinct interventions that Taiwan implemented with remarkable speed including community initiatives, hackathons, etc. Their “Face Mask Map” a collaboration initiated by an entrepreneur working with government helped prevent the panicked buying of facemasks, which hindered Taiwan’s response to SARS by showing where masks were available and providing information for trades and donations to those who most needed them, which helped prevent the rise of a black market.

What specific initiatives like this should we be implementing here?

There are many challenges surrounding the pandemic that can be mitigated by superior data. Better management of supply chains is one. Another, alluded to in your previous question, is anticipating where there will be shortfalls in hospital capacity or equipment. Yet another is determining where compliance with federal and local guidelines or requirements may be lagging. Ultimately, however, better information needs to be accompanied by the means and political will to respond to problems once they have been identified. It does not help to know about shortages or imperfect social distancing if nothing is done to address them.

38. It was reported that the White House is reaching out to health technology companies about creating a national coronavirus surveillance system to provide a real-time view of where patients are seeking treatment and for what. Essentially, compiling potentially sensitive health information and put it in a database.

Can you provide your thoughts on the White House creating a national surveillance system, the potential pitfalls and how increased government surveillance can affect marginalized communities, particularly communities of color?

I alluded to several scholars in my testimony who have identified the ways in which artificial intelligence has inured disproportionately to the detriment of vulnerable populations and people of color. I did not mention, but will now, the work of law professor and anthropologist Kiara Bridges evidencing the unique privacy challenges that under-resourced communities tend to face. For example, access to public benefits are often predicated on invasive information gathering that wealthier individuals interacting with private providers can avoid.

The reports you mention are concerning. My expectation would be for the government to develop a comprehensive, transparent plan to address the pandemic and enlist technologies only where appropriate to carrying out that plan, rather than issue non-specified requests for creating a surveillance system (that may benefit the companies themselves). Moreover, I am concerned about the prospect of mission creep and its consumer correlate, secondary use. My hope is that this body will consider safeguards against repurposing COVID-19 surveillance for other uses.

**Sen. Baldwin**

39. Emerging reports from many localities demonstrate that COVID-19 is having a disproportionate impact on African Americans and communities of color. For example, in my home state of Wisconsin, Milwaukee County reports that approximately 70% of those killed by coronavirus are African American, despite that community making up only 26% of the county's population.

We know this about Milwaukee County because the local government is proactive about collecting and reporting data on race and ethnicity. Reporting indicates that this disproportionate impact exists in places with significant African American communities, including Chicago, New Orleans, and Detroit. But a lack of consistent, quality data nationwide means we do not yet know just how sizable this disparity is, and what we can do about it.

While I am encouraged that we are drawing on the massive amount of data about Americans held by the private sector to support the COVID-19 response, I worry that it may not include and represent all communities equally. For example, if we use mobility data from mobile phones or particular apps to inform our understanding of adherence to social distancing requirements, I am concerned how it might affect the usefulness of the dataset if members of certain minority communities less likely to own such a device or utilize such an app.

For the members of our panel: how do you think "big data" can support efforts to strengthen our public health knowledge around COVID-19 and race, and how can we ensure that the methods and models through which "big data" supports our understanding of the epidemic take into account differences among communities?

Senator Baldwin, I share your concerns. In my testimony and answers, I have repeatedly alluded to the potential for bias to creep into data-driven responses to the pandemic. In my testimony, I invite the Committee to consider a hypothetical: "Imagine, for example, that public health officials were to allocate coronavirus resources on the basis of data trends from connected thermometers like Kinsa Health (retail cost: \$35.99 – \$69.99) or connected pulse oximeters like iHealth Air (retail cost: \$69.99). Only communities where sufficient numbers of consumers were aware of such devices and could afford them would receive an early warning or stockpiled support." Any partnership between government and industry around AI should be cognizant of the limitations and pitfalls of data analytics.

40. I am also concerned about the impact of "big data" informing our COVID-19 response on rural communities. Again, I worry that some of these data sources may not be well-utilized in rural America – where connectivity is still a significant challenge – and thus may not reflect the reality of the pandemic in those communities. But, I recognize that

Mr. Ryan Calo, Law Professor, University of Washington

this information is vital to developing better predictive models that can inform our current response to COVID-19 and help us prepare for the future.

For the members of our panel: how does “big data” ensure that the different experiences of rural, suburban and urban communities are taken into account when informing models that may guide the COVID-19 response?

This is an important but difficult question. I worry that app-based or data-driven solutions to the pandemic will not necessarily work well for certain populations, including the many Americans living in rural counties. At the University of Washington Tech Policy Lab, we have created various methods to try to address concerns over bias in data and to vet policy ideas for their impact on diverse stakeholders. I would be happy to share documentation on these initiatives—called Data Statements and Diverse Voices, respectively—with the Committee. Another good source of information is AI Now, the NYU-based research institute devoted to the societal impacts of artificial intelligence. But there is no silver bullet for entirely avoiding bias or the uneven distribution of costs of benefits of technology.

41. It is important that public health, and local public health departments in particular, have the data they need to map and anticipate hotspots for infectious disease outbreaks such as COVID-19 or overdose patterns in a community, including data that may be generated by the private sector. It is also important that local health departments have the capability to leverage this information together with that available through traditional public health surveillance efforts. For the members of our panel: how can the private sector coordinate data efforts with public health and ensure that local health departments have the necessary capabilities to make full use of these efforts?

I don't have access to sufficient information to answer this question well. I would note that local authorities not only need access to data, but also the expertise and context to understand the data well enough to inform decision-making. In this, local research universities, such as the University of Wisconsin, could be a great resource.

42. In speaking with experts in Wisconsin working on developing and refining predictive models around COVID-19, I heard that while there is a significant number of both public sector and private sector data sources to inform models, the data is not consistently easy to obtain and incorporate. As we rely on real-time models to inform the COVID-19 effort, as well as look to prepare for future infectious disease outbreaks, it is important that data-sharing be as seamless as possible. For the members of our panel: what are ways we can strengthen the data-sharing infrastructure for government, public health, academic and private sector sources?

I don't have access to sufficient information to answer this question well. Thank you for this and your other important questions.

**Sen. Tester**

43. We're all anxious to use any tools at our disposal to help keep Americans safe during this pandemic. But I think back to some of the powers the government gave itself after 9/11 that maybe went too far. Are there privacy protections the administration has suspended, or may suspend, that threaten our civil liberties?

Senator Tester, this is a very important question. In my testimony and answers, I have expressed concern about the possibility of mission creep. Mission creep refers to the tendency that surveillance and other powers conferred for one purpose—such as combatting terrorism or addressing a deadly pandemic—will come to be used for another. There are strategies that lawmakers can use to increase accountability and guard against mission creep. These include clear rules on what data can be collected and how it can be used, judicial oversight of surveillance, and sunset provisions, which require new surveillance powers to come up for a vote every few months or years. These mechanisms are no panacea and surveillance powers tend to possess a certain inertia. The analog to mission creep in consumer privacy is secondary use. Secondary use refers to collecting data for one person only to use for another. This body should explore prohibitions on secondary use of COVID-19 related data absent affirmative consent from the data subject.

44. Lots of folks in my state, especially seniors and veterans, had to suddenly adapt to telehealth without much warning. Understandably, they're drawn to platforms like Skype and Zoom that are among the easiest to use, but may expose sensitive physical and mental health data. What advice would you give them to mitigate threats to their personal information?

This is also an important question. In addition to seniors and veterans relying on telehealth, I would add that other vulnerable populations such as children are using video conferencing at an unprecedented scale.

I have expressed concerns that Zoom in particular was not prepared from a privacy and security perspective to handle this wholesale migration of work, learning, and play to its platform. I would note that most household name technology companies—from Twitter to Google to Uber to Facebook—are presently under a consent decree with the Federal Trade Commission for privacy and security lapses. Without opining on the specifics of Zoom's privacy policies or practices, I would encourage this body to ask the FTC to conduct an audit of Zoom and other, similar platforms to determine industry best practices and take any action necessary to ensure compliance under Section V of the FTC Act.

Thank you for your great questions.

Mr. Ryan Calo, Law Professor, University of Washington

**Sen. Sinema**

45. Some states, including Arizona have limited testing capabilities and therefore limited testing. It is also widely reported that tests around the world have produced inaccurate results. How can we mitigate against inaccurate assumptions related to disease trends in situations in which we have limited or inaccurate data?

Senator Sinema, I am not well-positioned to answer this excellent question. I would note that issues of inaccurate testing would have to be resolved before other techniques such as digital contact tracing could hope to be effective.

46. Many point to travel as a key factor in the spread of COVID-19. Contact tracing for travelers, specifically by plane, is a mechanism that can slow the spread of the virus. The data collected (full name, address while in U.S., email address, and two phone numbers) enables the government to contact individuals who may have come into contact with an individual who has tested positive. Once contact is established, individuals can start self-quarantining. What is the best way to balance the need for this information to slow the spread of the virus and privacy rights?

In my testimony and answers, I have emphasized the inevitable trade offs between individualized, government-backed surveillance to combat COVID-19 and civil liberties. Contact tracing in the wake of travel is one example where the trade off may be worth the gains. The American people through their representatives may decide that these extraordinary times call for invasive measures in order to slow and contain the spread of coronavirus. I am not an elected official and so it is hard for me to speak on anyone's behalf but my own. What I want to emphasize is that effective technical measures to address COVID-19 are going to require significant investment of government resources such as you are describing in your question and palpable trade offs to civil liberties.

There may be ways to mitigate some of these harms. I have referred repeatedly to the importance of guarding against secondary use and mission creep, i.e., the persistence and migration of surveillance powers created in one context such as a global pandemic or terrorism into a new context such as narcotics trafficking. As I mentioned in my testimony, paraphrasing Justice Robert Jackson, a problem with emergency powers is that they tend to kindle emergencies. Specific accountability measures include: (1) limits on data retention, (2) sunset provisions for new surveillance powers, (3) prohibitions on secondary use absent affirmative consent from the data subject, (4) penalties for abuse, and (5) judicial oversight. But none of these are a panacea or likely entirely to avoid harms to privacy and civil liberties.

Mr. Ryan Calo, Law Professor, University of Washington

47. How can big data help resolve challenges within the manufacturing supply chain to spur increased production and distribution of needed testing, personal protective equipment, and other resources to address this pandemic?

There are many challenges surrounding the pandemic that can be mitigated by superior data. Better management of supply chains is one. Another is anticipating where there will be shortfalls in hospital capacity or equipment, which you mention in your question. Yet another is determining where compliance with federal and local guidelines is lagging. I allude to still more in my testimony. Ultimately, however, better information needs to be accompanied by the means and political will to respond to problems once they have been identified. It does not help to know about shortages or imperfect social distancing if nothing is done to address them.

48. This pandemic has caused serious economic harm. Businesses of all sizes and their employees suffer as sales drastically fall or disappear altogether. State, tribal and local governments are under enormous strain as response costs increase and revenues drop.

How can big data assist in the better creation and execution of economic assistance programs like the Paycheck Protection Program, Treasury's lending facilities, business interruption or pandemic risk insurance, and state, tribal and local stabilization funds?

I am not well-positioned to answer this important question. Thank you for the opportunity to answer all of your excellent questions.

**Sen. Rosen**

49. Germany's national disease control center recently asked their citizens to donate data collected by their fitness tracker. This voluntary initiative has consumers download an app on their phones and contribute health information such as pulse rates and temperature that is collected by fitness tracking devices anonymously. Using machine learning, epidemiologists can analyze this data to better understand the spread of the coronavirus across the country and detect previously unknown clusters.

What are the advantages and pitfalls in using voluntarily donated data to improve responses during a pandemic?

Senator Rosen, the advantages of using voluntary data are that subjects have presumably consented to the use of their information. That trust should not be abused, of course, but the voluntariness addresses some of the privacy risks. I detail some of the pitfalls in my testimony and below. They tend to involve the ways voluntary data can be unrepresentative of the overall population or otherwise skewed.

Mr. Ryan Calo, Law Professor, University of Washington

How can we use donated data to support our response to this pandemic and future similar public health issues?

There are myriad use cases for donated data, which depend on the nature of the data and the extent of participation. Categories include symptoms, health status (virus or antibodies), health outcomes, extent of social distancing, consumption or availability of scarce goods, economic conditions, and many others.

What privacy guardrails are needed to ensure that this data is collected and analyzed safely and anonymously?

In response to questions from Senator Blunt and Blackburn, I identified a series of principles that government should follow with respect to public health surveillance especially in partnership with privacy industry. These include: (1) having a clear sense of the legitimate purpose for which data is being collected or changing hands, (2) only collect information needed to accomplish that purpose, (3) anonymize data unless doing so would not thwart a legitimate government purpose, (4) (3) physically and digitally secure data, and (5) establish clear rules around secondary use, mission creep, and consequences for abuse.

What are the gaps we need to consider when analyzing such data?

Voluntary fitness data is likely to reflect a particular population, which may not be representative of a pluralistic society such as the United States. In my testimony, I invite the Committee to consider a hypothetical involving making public health decisions based on aggregate data from a private source : “Imagine, for example, that public health officials were to allocate coronavirus resources on the basis of data trends from connected thermometers like Kinsa Health (retail cost: \$35.99 – \$69.99) or connected pulse oximeters like iHealth Air (retail cost: \$69.99). Only communities where sufficient numbers of consumers were aware of such devices and could afford them would receive an early warning or stockpiled support.” Any partnership between government and industry around AI should be cognizant of the limitations and pitfalls of data analytics.

50. The National Science Foundation (NSF) is the only federal agency whose mission includes supporting all fields of fundamental science and engineering. The research and educational programs backed by NSF are integral to the continued success of our country’s innovation, supporting scientific discoveries that have led to new industries, products, and services. Since 2012, NSF has funded research on the emerging field of data science through its BIG DATA program. Now, NSF’s larger program – “Harnessing the Data Revolution” – will support research, educational pathways, and advanced cyberinfrastructure in the field of data science.

Mr. Ryan Calo, Law Professor, University of Washington

Given NSF's leadership in data science research and development, what role do you think NSF can play in leading public-private partnerships for increased research on big data that could help address the COVID-19 crisis or future pandemics?

The National Science Foundation has been unlocking funding for research related to COVID-19 with its RAPID and EAGER initiatives. This funding will be critical in fueling ongoing innovation by research universities such as my own. Beyond this, NSF officials—who are commonly on rotation from academic positions—have the relationships and legitimacy to convene academia, industry, and government to share knowledge and resources. My long-held view is that conversations should include social scientists and academics from the humanities alongside data scientists and health experts. Only such an interdisciplinary conversation will account for the full societal impacts of efforts to address the pandemic.

Thank you for the opportunity to answer your excellent questions.