Senator Tillis
Questions for the Record
Big Data, Big Questions: Implications for Competition and Consumers
Mr. John Robb, Global Guerillas Report

Q: The term "data" can have multiple meanings, which can sometimes generate confusion in policy discussions. How would you define "data" and "big data", as used in your written testimony?

A: The difference between data and "big data" is scale.

Big data has far more volume, variety, and velocity than traditional data. This complexity changes the way insight is generated from it. Instead of seeking definitive answers (customers are buying x more than y), analysts look for patterns. Patterns that can be useful in spotting opportunities, detecting problems, and shaping solutions. Finding these patterns through intentional analysis, particularly patterns that aren't ephemeral, is very difficult.

Big data is particularly useful for training AIs (machine learning). There's a high correlation between the amount of data used to train an AI and the quality of the AI. AIs trained using big data naturally find unexpected patterns and contextual cues that are useful in solving the problem they are trying to solve.

Q: How would you define consumer and user data, specifically what would be included and excluded from these definitions?

A: Traditional consumer data is mostly static demographic data. How much do you make, where do you live, and what is your education level?

That's radically changed with the arrival of networked user data. This new user data includes;

- Tracking data. Mouse and keyboard clicks. Page and site visits (on a single site and across sites). Physical location. Tracking over time and space. This is now expanding into detecting focus (eyeball tracking), physical measurement (health monitoring to hand movement), sensor data (pictures, videos, and other data collected by sensors in the environment, from a Tesla car to CCTV), etc.

- User-produced data. What people upload (both intentionally and unintentionally). Textual content (texting, blog posts, essays), voice content (podcasts to audio capture from Alexa or an iPhone), images (smartphone pictures to YouTube uploads). This data set is expanding.

- AI-training data (user-produced + tracking data at scale). User data is being used to actively train AIs. For example; Tesla uses user feedback and

experiences to improve their self-driving system. Enlisting customers to actively train AIs is something we will see much more of in the future. Other firms strip mine the open Web (without consent) to build AIs (GPT-4 (text), Stable Diffusion (images), Google translate (text), etc.).

Q: Would this include user-uploaded videos, images, and text?

A: Yes, see above.

Q: Would such content be considered part of the "user" data, even if it includes content that originates from other sources?

A: Tracking data and uploaded data are both "user data." Data brokers aggregate this data for sale to marketing departments/firms and firms building AIs. Sometimes it is anonymized; sometimes, it isn't. Sometimes firms with access to user-tracking data sell it to generate extra revenue.

If the text or image being uploaded is from another source (somebody else took the picture or wrote the text), then it isn't user data, but the tracking data on the upload is. This is why the best approach to compensating people for data use is made in aggregate, at the population scale.

For example, a data ownership bill for the people of North Carolina would aggregate the data provided by people in the state. Data brokers, with a fiduciary duty to the people contributing to the data set, would compete to sign up people attached to this pool.

Data brokers representing individuals would solve the privacy problem in a way that doesn't destroy data (like in Europe). Data brokers, fueled by industry revenue, would have the lawyers and technologists needed to protect data and find new data sources collected by advancing technology. In contrast, privacy-based regulations rely on getting the attention of overwhelmed bureaucrats.

Q: Does it include data in which intellectual property rights, including copyright, trade secret, trademark, or design rights, may subsist?

A: No. There is already an industry that protects that data. However, firms that claim (like many social networks) all of the user-uploaded data and data collected about them, would not be allowed to claim ownership over it. That data would be owned by the individuals in question.

Q: How would you differentiate your proposal for people to own data from existing intellectual property rights-based approaches?

A: Technology has outpaced our laws. The data being generated by individuals isn't adequately protected by existing legal frameworks. Providing a mechanism for protecting the ownership rights of individuals in aggregate is critical to safeguarding against abuse (by corporations and the government) as well as providing a mechanism for participating in the AI-driven economy of the future.

To be precise: the data being strip-mined right now is being used to create the most valuable technological artifacts that have ever existed, without compensation to the people it is coming from. Making it possible for people to participate in the upside potential of this development will be as important as land ownership was to the development of capitalism (one of the most revolutionary aspects of the American Revolution, and why it was so important to early capitalism, was the ability of individuals without a hereditary title to own land).

Q: You advocate for a new form of digital identity. Could you please explain further what you mean by this, and how you envision it working in the current environment? Are there particular technologies that would need to be developed for this to be implemented?

A: Digital identity is necessary for the assignment of rights (of speech, ownership, etc.). It simply connects an online identity (collection of accounts, activities, etc.) to a living, breathing person in the physical world. Typically, this is done using the same approaches used for registering a financial account (government ID, etc.). It can become more sophisticated through the application of AI (as we are about to see with Twitter).

Q: Ms. Slaiman advocates for "a digital regulator to comprehensively the policy questions surrounding digital platforms." Do you agree that this is necessary?

A: If the digital regulator is overseeing the launch and establishment of a new industry (data brokers, etc.) and the technology standards that support it, then yes. If its intent is to build a regulatory enforcement body that is focused on privacy regulations and increasingly restrictive content moderation (as we see in the EU), then no. That would be a disaster (for example: China would win the AI race and the US and EU would suffer economic impoverishment due to it).

Q: Given the many issues beyond privacy and competition that address and implicate digital policy—including cybersecurity, national security, consumer rights, free speech, and intellectual property concerns—what existing agency would be the best situated, in your view, to carry out this role?

A: Something that looks like a cross between the SEC and Consumer protection. It would need to be focused on the individual, and their rights, while being able to manage the needs of growing a new industry (data brokers that represent the data pools that individuals join). This industry has the potential to rival finance

in size over time (percentage ownership or royalty rights from big AIs could be worth tens of trillions in the not-too-distant future).

Q: Is it important to you that the regulator should be politically accountable?

A: Of course. However, if the political focus is on the micromanagement of user data collection to shape society, then no. The goal of political discussions on this issue should be focused on creating the *minimum* viable rule set for a prosperous and successful society.

Senator Blackburn
Questions for the Record to John Robb
The Global Guerrillas Report

Q: It's imperative for the U.S. to get a national consumer privacy law in place—the EU and China have already done so. Given consumer concerns about how their data is being used online, what should that regime look like? What are the obstacles the United States faces in getting to that point?

A: There are three systems in place:

1. EU privacy laws use aggressive regulatory oversight to limit and destroy data.
2. China assigns ownership of data rights to the government and assigns loyal corporations the right to fully gather and exploit it.
3. US doesn't have a centralized approach. With few exceptions, it lets corporations do whatever they want in regard to data collection and exploitation.

Here's how this will play out:

- The EU approach is that it will prevent the development of the AIs and products/services that will drive economic growth in the future. Their approach to data is likely to result in economic impoverishment long term.

- The Chinese approach will generate economic growth and success. However, it will also be used for networked authoritarianism by the Chinese government. It will provide the government with complete control over the entire population in real time. From behavior to perception (through control of augmented reality).

- The US approach will yield some economic success, but it will be a system completely controlled by big corporations and a few wealthy individuals. Almost all of the economic success generated by this approach will concentrate in the hands of the corporations. Furthermore, the control

corporations have over data will allow them to dominate politics (in short, corporate-run network authoritarianism).

The solution?

The solution that allows the US the ability to succeed economically and avoid authoritarianism is to provide people with digital rights and data ownership. That approach would create an industry response to user data (an industry of data brokers/banks, much like the financial industry, which has a fiduciary responsibility to protect this data and maximize its returns). It ensures that the data needed to fuel development is available to corporations while allowing the people who provide this data a means of participating in the great wealth created by it.

_____

Senator Chuck E. Grassley
Questions for Mr. John Robb:

Q: How important is the amount of data that a company has to their ability to effectively monetize that information?

A: A few thoughts on this:

- Data is key to success in a networked economy.
- The more you have and the better its quality, the more success you will have. NOTE: there is a high correlation between the amount of data you have available for an AI to train with and the quality of the AI.
- At the level of the economy, if corporations don't have access to data, they won't be able to match the products available from China that do have access.

Q: How difficult can it be for a startup or small business to collect enough data to be able to compete with companies that have large amounts of data?

A: It's hard, but it becomes impossible if the big companies control the small company's access to data. Big companies are using access to data as a weapon against the competition and as a means of extracting monopoly rents from their platforms. Apple and Google are good examples of this. This is already bad in the smartphone industry, it is going to become a catastrophe when augmented reality arrives (soon).

Q: Some commentators argue that the amount of data currently possessed by large incumbent companies forecloses the ability of new entrants to compete.

But, new data is being created every day and what data is important in the future may not be what is being collected today. If so, why isn't there an opportunity for additional companies to enter the market?

A: I agree. Data aggregation and privacy regulations (or "concerns') are being used by big companies to dominate marketplaces. New data is created, but these companies control the flow. For example, this control has allowed big companies in the smartphone market to charge a ~30% tax on transactions in the smartphone economy. It's extortionate.

Q: There is debate over whether Big Data should be regulated through the lens of consumer protection and privacy or whether antitrust laws should be used to address competition concerns with the collection of data. Do you have an opinion about the best approach to address this issue or should we be looking at a combination of different approaches?

Best approach: Data ownership for individuals. They contribute all of the data collected about them to big pools. These pools are managed by data brokers who have a fiduciary duty to protect this data and maximize the returns generated by it.

The data relationship between an app running on a smartphone or an augmented reality headset would be with the data brokers representing the individuals using the device. The device company wouldn't be able to use its control over data to extort monopoly rents. Instead, that benefit would flow to the individuals who contributed the data, providing them participation in the economy being built on this data.

A combo of SEC/banking (industry focus) and consumer protection agency would work best. The goal would be to set up a data brokerage/banking industry (with the individual as the client) that is so profitable that it could hire the people needed to enforce it.