Connecticut Debate Association
State Finals, March 24, 2018
Joel Barlow High School

Resolved: Individuals should be paid for the use of their personal data.

Should tech companies pay us for our data?

The Los Angeles Times, By ANDREAS WEIGEND, FEB 14, 2017

Should tech companies pay us for our data?

Data are the most important resource of the 21st century. The capacity to transform raw data into decision-making recommendations is changing the world in ways that rival the Industrial Revolution.

Much of the data being created and shared are about our personal lives: where we live, where we work, where we go; who we love, who we don't and who we spend our time with; what we ate for lunch, how much we exercise and which medicines we take; what appliances we use in our homes and which stories grab our attention.

And the companies that collect and analyze these data are generating billions in revenue. According to a report by Oracle and the MIT Technology Review, the most successful data companies, including Amazon, Google and Uber, treat data as an asset — something that can be bartered, sold and monetized.

If ... Facebook divvied up every cent of its profits among its users, each would receive about $5 for 2016.

If companies are banking on your personal data for their business, shouldn't you get a piece of their profits? To many observers, that seems only fair.

Microsoft Research scientist Jaron Lanier and others argue that users should receive “nanopayments” for data created by them. One Dutch technology pundit, Jathan Sadowski, has even said that a lot of data collection is a “form of theft” — an appropriation “without consent and compensation.”

But it's pointless to chase nanopayments, and not just because the big technology companies might not like the concept.

Many of the products and services provided by "data refineries" — including Amazon's recommendations, Google's traffic predictions and Uber's surge pricing — depend on information from millions of individuals. One person's contribution isn't worth much. If, for instance, Facebook divvied up every cent of its profits among its users, each would receive about $5 for 2016.

Was having unlimited access to a communication and networking platform — one that is constantly evolving, with new features including livestreaming and encryption in Messenger — worth more to you than a venti caramel macchiato? If so, you've already been "paid" for your data.

Of course the nanopayments crowd is quick to point out that not all users are equal; some people create more valuable content than others. If only the truly deserving received payment, that might add up to more sizable figures for them.

The question then becomes: Who's truly deserving?

Let's say you take a photo of a group of friends at a birthday party and post it on Facebook. Many people in your social network then repost it. The value of the post to Facebook comes from the traffic that the photo inspires as well as the data about relationships and interests embedded in people's interactions with it. Should you alone get a kickback from Facebook? Or should you split it with everyone tagged in the photo? How about with those who comment, tag or like it? How much does a comment or like depreciate in value over time?

Not only would it be costly to implement algorithms to make these calculations, it's often not clear what would be fair.

Besides, focusing on the value of our data ignores the benefits we receive. I know a man — let's call him Joe — who decided to try out Facebook but didn't want to share data about himself. He signed up under a false name and didn't make friends. Unsurprisingly, Joe didn't find anything very useful on Facebook. Because there were no data for Facebook to analyze, its algorithms couldn't deliver a personalized News Feed.

The value of data comes from improving the decisions we make. You have to give data to get this benefit.

But while Joe missed out on the joys of Facebook, Facebook missed out on essentially nothing. Since most people are not like Joe — most people share their data — Facebook has plenty of information to work with as it refines its services. Again, any one individual's data have a dollar value of close to zero.

Rather than focusing on how much our information is worth and asking for a paltry financial handout in return, we should demand something far more valuable: the right to experiment with our data and the settings that determine what

data refineries show us. This "seat at the controls" will ensure that we can make the best decisions for ourselves, while still benefiting from these companies' recommendations.

Andreas Weigend was the chief scientist at Amazon. He is now a lecturer at UC Berkeley and the director of the Social Data Lab. He is the author of "Data for the People."

It's Time to Tax Companies for Using Our Personal Data

The New York Times, By Saadia Madsbjerg, Nov. 14, 2017

Our data is valuable. Each year, it generates hundreds of billions of dollars' worth of economic activity, mostly between and within corporations — all on the back of information about each of us.

It's this transaction — between you, the user, giving up details of yourself to a company in exchange for a product like a photo app or email, or a whole ecosystem like Facebook — that's worth by some estimates $1,000 per person per year, a number that is quickly rising.

The value of our personal data is primarily locked up in the revenues of large corporations. Some, like data brokerages, exist solely to buy and sell sets of that data.

Why should companies be the major, and often the only, beneficiaries of this largess? They shouldn't. Those financial benefits need to be shared, and the best way to do it is to impose a small tax on this revenue and use the proceeds to build a better, more equitable internet and society that benefit us all.

The data tax could be a minor cost, less than 1 percent of the revenue companies earn from selling our personal data, spread out over an entire industry. Individually, no company's bottom line would substantially suffer; collectively, the tax would pull money back to the public, from an industry profiting from material and labor that is, at its very core, our own.

This idea is not new. It is, essentially, a sales tax, among the oldest taxes that exist, but it hasn't been done because assigning a fixed monetary value to our data can be very difficult. For a lot of internet businesses, our personal data either primarily flows through the business or remains locked within.

The flow-through type of business is an internet provider, like AT&T, Comcast and Verizon; the latter is a platform provider, like Facebook or Twitter. Google mostly makes platforms, like its search function and Gmail, but it also lays fiber-optic cable, providing the internet to some municipalities. The company also is making self-driving cars.

It's perhaps easiest to consider your data as something real and physical, like a car, which, in a sense, it is. It moves around a real, physical infrastructure, owned and operated by the internet providers, and the information is also stored — or parked — on millions of hard drives in vast buildings in usually cold climes. These are owned and operated by the platform makers, including Amazon, which has a very lucrative arm that does nothing but rent out its server space.

When you use the internet, your information travels through the providers' pipes (roads) and into and out of the platform makers' servers (parking lots). For the most part, the platform makers rely on your data to improve their products. Google uses its immense trove of real human searches to make better artificial intelligence like transcriptions, translations and self-driving cars.

How much, exactly, these new and innovative businesses are worth is unclear even to Google. But it, and all the platform makers, don't just improve products with our data. The companies invite advertisers into the platforms, offering to deliver advertisements to exactly the right kind of person, at exactly the right time, based off the platform makers' intimate knowledge of us, via our data. The internet providers are less nuanced: We pay a monthly fee for access to the pipes the companies lay and the upkeep on them, but the companies also sell our data to brokers, who bundle and sell it to advertisers.

It's this data brokerage industry that should be the primary, initial focus of the data tax. This industry exists solely to collect our information and sell it as a commodity to retailers, advertisers, marketers, even other data brokerages and government agencies.

It's this marketplace that traffics in the actual monetary value of our data, and from it we can begin to map just how much that data might be worth overall.

The data brokerage industry generates more than $150 billion in revenue each year, but revenue is reportedly growing so fast that it's expected to reach $250 billion by 2018. A small tax, say 0.8 percent, on data brokerages based in the United States would generate about $2 billion annually.

Microlevies like this one have been issued successfully before. Over the last decade, the governments of 10 countries, including Chile, France and Niger, have successfully raised more than 1 billion euros in funding from a tax on airline tickets of €1 to €40 depending on the class of ticket.

The money generated has gone toward global H.I.V./AIDS, tuberculosis and malaria eradication programs. In Austria, the government was considering imposing a value-added tax on the big data transactions of social media companies that benefit from personal data, but has been stymied by the complications of assigning a fixed value to such transactions.

Our data is ours, but it also is not ours. We trade it away for so much of our experience on the internet. Money from a data tax could begin to counter this trade imbalance.

The money should go toward improving privacy of our information on the internet, countering identity theft, improving connectivity and internet literacy, all causes that would help create a more equitable internet for all.

Saadia Madsbjerg is managing director of the Rockefeller Foundation

Your Data Is Crucial to a Robotic Age. Shouldn't You Be Paid for It?

The New York Times, By Eduardo Porter, March 6, 2018

Should Facebook pay us for our puppy pictures?

Of course, the idea sounds crazy. Posting puppies on Facebook is not a chore. We love it: Facebook's 1.4 billion daily users spend the better part of an hour on it every day. It's amazing that we don't have to pay for it.

And yet the idea is gaining momentum in Silicon Valley and beyond: Facebook and the other technological Goliaths offering free online services — from which they harvest data from and about their users — should pay for every nugget of information they reap.

The spring break pictures on Instagram, the YouTube video explaining Minecraft tactics, the internet searches and the Amazon purchases, even your speed following Waze on the way to spend Thanksgiving with your in-laws — this data is valuable. It will become more valuable, potentially much more so, in the not-too-distant future.

Getting companies to pay transparently for the information will not just provide a better deal for the users whose data is scooped up as they go about their online lives. It will also improve the quality of the data on which the information economy is being built. And it could undermine the data titans' stranglehold on technology's future, breathing fresh air into an economy losing its vitality.

The idea has been around for a bit. Jaron Lanier, the tech philosopher and virtual-reality pioneer who now works for Microsoft Research, proposed it in his 2013 book, "Who Owns the Future?," as a needed corrective to an online economy mostly financed by advertisers' covert manipulation of users' consumer choices.

It is being picked up in "Radical Markets," a book due out shortly from Eric A. Posner of the University of Chicago Law School and E. Glen Weyl, principal researcher at Microsoft. And it is playing into European efforts to collect tax revenue from American internet giants.

In a report obtained last month by Politico, the European Commission proposes to impose a tax on the revenue of digital companies based on their users' location, on the grounds that "a significant part of the value of a business is created where the users are based and data is collected and processed."

Users' data is a valuable commodity. Facebook offers advertisers precisely targeted audiences based on user profiles. YouTube, too, uses users' preferences to tailor its feed. Still, this pales in comparison with how valuable data is about to become, as the footprint of artificial intelligence extends across the economy.

Data is the crucial ingredient of the A.I. revolution. Training systems to perform even relatively straightforward tasks like voice translation, voice transcription or image recognition requires vast amounts of data — like tagged photos, to identify their content, or recordings with transcriptions.

"Among leading A.I. teams, many can likely replicate others' software in, at most, one to two years," notes the technologist Andrew Ng. "But it is exceedingly difficult to get access to someone else's data. Thus data, rather than software, is the defensible barrier for many businesses."

We may think we get a fair deal, offering our data as the price of sharing puppy pictures. By other metrics, we are being victimized: In the largest technology companies, the share of income going to labor is only about 5 to 15 percent, Mr. Posner and Mr. Weyl write. That's way below Walmart's 80 percent. Consumer data amounts to work they get free.

"If these A.I.-driven companies represent the future of broader parts of the economy," they argue, "without something basic changing in their business model, we may be headed for a world where labor's share falls dramatically from its current roughly 70 percent to something closer to 20 to 30 percent."

As Mr. Lanier, Mr. Posner and Mr. Weyl point out, it is ironic that humans are providing free data to train the artificial-intelligence systems to replace workers across the economy. Commentators from both left and right fret over how ordinary people will put food on the table once robots take all the jobs. Perhaps a universal basic income, funded by taxes, is the answer?

How about paying people for the data they produced to train the robots? If A.I. accounted for 10 percent of the economy and the big-data companies paid two-thirds of their income for data — the same as labor's share of income across the economy — the share of income going to "workers" would rise drastically. By Mr. Weyl and Mr. Posner's reckoning, the median household of four would gain $20,000 a year.

A critical consideration is that if people were paid for their data, its quality and value would increase. Facebook could directly ask users to tag the puppy pictures to train the machines. It could ask translators to upload their translations. Facebook and Google could demand quality information if the value of the transaction were more transparent. Unwilling to enter in a direct quid pro quo with their users, the data titans must make do with whatever their users submit.

The transition would not be painless. We would need to figure out systems to put value on data. Your puppy pictures might turn out to be worthless, but that college translation from Serbo-Croatian could be valuable. Barred from free data, YouTube and Facebook might charge a user fee for their service — like Netflix. Alternatively, they might make their money from training A.I. systems and pay some royalty stream to the many people whose data helped train them.

But whatever the cost, the transformation seems worthwhile. Notably, it could help resolve one of the most relevant questions coming into focus in this new technological age: Who will control the data?

Today, the dominant data harvesters in the business are Google and Facebook, with Amazon, Apple and Microsoft some way behind. Their dominance cannot really be challenged: Could you think of a rival search engine? Could another social network replace the one all your friends are on? This dominance might matter less if companies had to pay for their users' data.

Google and Facebook and Amazon would not be able to extend the network effects that cemented their place at the top of the technology ecosystem to the world of A.I. Everybody wants to be on Facebook because everybody's friends are on Facebook. But this dominance could be eroded if rivals made direct offers of money for data.

Companies with different business models might join the fray. "This is an opportunity for other companies to enter and say look, we will pay you for this data," Mr. Posner said. "All this is so new that ordinary people haven't figured out how manipulated they are by these companies."

The big question, of course, is how we get there from here. My guess is that it would be naïve to expect Google and Facebook to start paying for user data of their own accord, even if that improved the quality of the information. Could policymakers step in, somewhat the way the European Commission did, demanding that technology companies compute the value of consumer data?

In any event, there is probably a better deal out there, in your future, than giving Facebook free puppy pictures.

A Way to Own Your Social-Media Data

The New York Times, By Luigi Zingales and Guy Rolnik, June 30, 2017

The European Union imposed a 2.4 billion euro ($2.7 billion) fine on Google last Tuesday for manipulating its search engine results to favor its own comparison shopping service. It is just the latest institution to recognize the increasing monopolization of the technology industry.

Google has about a 90 percent market share in searches, while Facebook has a penetration of about 89 percent of internet users. Economists have a fancy name for this phenomenon: "network externalities." In traditional product markets, one customer's choice (for example, a particular car tire) does not directly affect other individuals' preferences for that product, and competition generally ensures that consumers enjoy the best products at the lowest possible price.

In the market for social media, by contrast, when one customer uses Facebook over Myspace, it has a direct (and positive) impact on other customers' preferences for the same social network: I want to be in the social network where my friends are. These markets naturally tend toward a monopoly.

Historically, there have been two main government interventions to reduce this risk of monopoly power. The first is price regulation. When railway companies gained excessive market power in the late 1800s, the United States government created the Interstate Commerce Commission and gave it the power to set maximum prices. (In the long run, the remedy turned out to be worse than the disease, but that's another story.)

The second is antitrust. When Standard Oil, in the early 1900s, controlled 90 percent of oil refinery capacity in the United States, the federal government used its antitrust power to break it up into more than 30 smaller companies. A similar breakup was imposed 70 years later on AT&T.

Still, there is a problem with traditional antitrust policy when looked at through the lens of network externalities: It focuses only on consumers' benefits from competition. But consumers love Google and Facebook since they do not pay a dime for their services.

What many users do not fully appreciate is that they do pay for these services, in the form of very valuable information. And those who appreciate this cost have no choice: There is no major search engine that does not store our past searches or collect information on our activities, and there is no significant social media platform that does not retain our preferences. That is the cost of using these technologies. Lack of competition also means lack of choice, which is ultimately lack of freedom. But what can be done?

For a 21st-century problem, we suggest a 21st-century solution: a reallocation of property rights via legislation to provide more incentives to compete. In fact, the idea is not new. Patent law, for example, attributes the right to an invention to the company a scientist works for, to motivate companies to invest in research and development. Similarly, in the mobile industry, most countries have established that a cellphone number belongs to a customer, not the mobile phone provider. This redefinition of property rights (in jargon called "number portability") makes it easier to switch carriers, fostering competition by other carriers and reducing prices for consumers.

The same is possible in the social network space. It is sufficient to reassign to each customer the ownership of all the digital connections that she creates — what is known as a "social graph." If we owned our own social graph, we could sign into a Facebook competitor — call it MyBook — and, through that network, instantly reroute all our Facebook friends' messages to MyBook, as we reroute a phone call.

If I can reach my Facebook friends through a different social network and vice versa, I am more likely to try new social networks. Knowing they can attract existing Facebook customers, new social networks will emerge, restoring the benefit of competition.

Today Facebook provides developers with application-program interfaces that give them access to its customers' social graph, Facebook Connect and Graph A.P.I. Facebook controls these gates, retaining the right to cut off any developer who poses a competitive threat. Anticipating this outcome, very few developers invest seriously in creating alternatives, eliminating even the threat of competition.

By guaranteeing access to new customers' data and contacts, a Social Graph Portability Act would reduce the network externality dimension of the existing digital platforms and ensure the benefits of competition.

Google and Facebook will no doubt display their enormous lobbying power to kill this idea in its infancy. But they would do so at their own risk. If their monopoly is not curbed by competition, it will eventually be curbed by regulation, and experience suggests that will be worse, not only for consumers, but also for Google and Facebook themselves.

As the uproar over United Airlines' treatment of a passenger it ejected from a flight has shown, people's tolerance for companies' market power is running low. A "social graph to the people" revolution is in the making; Congress better be in front of it or find itself overwhelmed.

Luigi Zingales and Guy Rolnik are professors at the University of Chicago Booth School of Business.

Your Location Data Is Being Sold—Often Without Your Knowledge

The Wall Street Journal, By Christopher Mims, March 4, 2018

Location-based ads are growing, which means the industry has more ways than ever to track you

As location-aware advertising goes mainstream—like that Jack in the Box ad that appears whenever you get near one, in whichever app you have open at the time—and as popular apps harvest your lucrative location data, the potential for leaking or exploiting this data has never been higher.

It's true that your smartphone's location-tracking capabilities can be helpful, whether it's alerting you to traffic or inclement weather. That utility is why so many of us are giving away a great deal more location data than we probably realize. Every time you say "yes" to an app that asks to know your location, you are also potentially authorizing that app to sell your data.

Dozens of companies track location and/or serve ads based on this data. They aim to compile a complete record of where everyone in America spends their time, in order to chop those histories into market segments to sell to corporate advertisers.

Marketers spent $16 billion on location-targeted ads served to mobile devices like smartphones and tablet computers in 2017. That's 40% of all mobile ad spending, research firm BIA/Kelsey estimates, and it expects spending on these ads to double by 2021.

The data required to serve you any single ad might pass through many companies' systems in milliseconds—from data broker to ad marketplace to an agency's custom system. In part, this is just how online advertising works, where massive marketplaces hold continuing high-speed auctions for ad space.

But the fragmentation also is because of a very real fear of the public backlash and legal liability that might occur if there were a breach. Imagine the Equifax breach, except instead of your Social Security number, it's everywhere you've

been, including your home, your workplace and your children's schools.

The fix, at least for now, is that with most individual data vendors holding only parts of your data, your complete, identifiable profile is never all in one place. Giants like Facebook and Alphabet 's Google, which do have all your data in one place, say they are diligent about throwing away or not gathering what they don't need, and eliminating personally identifying information from the remainder.

Yet as the industry and the ways to track us expand, the possibility that our whereabouts will be exposed multiplies. If you've ever felt clever because an app on your phone asked to track your location and you said no, this should make you feel a little less smug: There are plenty of ways to track you without getting your permission. Some of the most intrusive are the easiest to implement.

The spy in your pocket

Your telco knows where you are at all times, because it knows which cell towers your phone is near. In the U.S., how much data service providers sell is up to them.

Another way you can be tracked without your knowing it is through any open Wi-Fi hot spot you might pass. If your phone's Wi-Fi is on, you're constantly broadcasting a unique address and a history of past Wi-Fi connections. Retailers sometimes use these addresses to identify repeat customers, and they can also use them to track you as you go from one of their stores to another.

WeatherBug, one of the most popular weather apps for Android and iPhone, is owned by the location advertising company GroundTruth . It's a natural fit: Weather apps need to know where you are and provide value in exchange for that information. But it also means that app is gathering data on your location any time the app is open—and even when it isn't, if you agreed to always let it track your location. That data is resold to others.

GroundTruth also says it gathers location data from "over a hundred thousand" other apps that have integrated bits of its code. Company President Serge Matta declined to disclose which apps. App makers agree to harvest location data because it grants them access to GroundTruth's mobile advertising network.

This data is what enables marketers like Jack in the Box to push an advertiser's message to potential customers near its restaurants. A typical engagement includes pushing location-based promotions or coupons through mobile ads, says Iwona Alter, chief marketing officer of Jack in the Box.

Every month GroundTruth tracks 70 million people in the U.S. as they go to work in the morning, come home at night, surge in and out of public events, take vacations, you name it.

Anonymize, de-anonymize

Companies like GroundTruth try to ensure they aren't tracking or storing data on individuals. Most of what they sell are anonymous blobs of people who fit particular descriptions—"soccer moms who intend to buy an SUV," for example. But they also occasionally hand off location data to a third party, such as LiveRamp, owned by data broker Acxiom , before it is matched up with potentially personally identifying information, such as your complete shopping history at a retailer. LiveRamp is almost like an escrow company for data.

Companies like Acxiom could be prime targets for hackers, said Chandler Givens, chief executive of TrackOff, which develops software to protect user identity and personal information. LiveRamp goes to great lengths to mathematically obfuscate our individual identities, said Sheila Colclasure, chief data ethics officer at LiveRamp and Acxiom. But some security researchers fear data brokers like Acxiom might be compromised already, or could be someday.

Acxiom and LiveRamp in the U.S. are governed by federal and state laws that regulate the collection and use of data in the particular businesses their clients are involved in, Ms. Colclasure said. Nearly every year, a bill comes up in the Senate or House that would regulate our data privacy—the most recent was after the Equifax breach—but none has passed. In some respects, the U.S. appears to be moving backward on privacy protections.

There might never be a breach of our location data. But given the drumbeat of hacks of both companies and governments, it's hard to believe hackers aren't at least trying to compromise such a high-value target.

Write to Christopher Mims at christopher.mims@wsj.com

Data Could Be the Next Tech Hot Button for Regulators

The New York Times, By Steve Lohr, Jan. 8, 2017

Wealth and influence in the technology business have always been about gaining the upper hand in software or the machines that software ran on.

Now data — gathered in those immense pools of information that are at the heart of everything from artificial intelligence to online shopping recommendations — is increasingly a focus of technology competition. And academics

and some policy makers, especially in Europe, are considering whether big internet companies like Google and Facebook might use their data resources as a barrier to new entrants and innovation.

In recent years, Google, Facebook, Apple, Amazon and Microsoft have all been targets of tax evasion, privacy or antitrust investigations. But in the coming years, who controls what data could be the next worldwide regulatory focus as governments strain to understand and sometimes rein in American tech giants.

The European Commission and the British House of Lords both issued reports last year on digital "platform" companies that highlighted the essential role that data collection, analysis and distribution play in creating and shaping markets. And the Organization for Economic Cooperation and Development held a meeting in November to explore the subject, "Big Data: Bringing Competition Policy to the Digital Era."

As government regulators dig into this new era of data competition, they may find that standard antitrust arguments are not so easy to make. Using more and more data to improve a service for users and more accurately target ads for merchants is a clear benefit, for example. And higher prices for consumers are not present with free internet services.

"You certainly don't want to punish companies because of what they might do," said Annabelle Gawer, a professor of the digital economy at the University of Surrey in England, who made a presentation at the Organization for Economic Cooperation and Development meeting. "But you do need to be vigilant. It's clear that enormous power is in the hands of a few companies."

Maurice Stucke, a former Justice Department antitrust official and a professor at the University of Tennessee College of Law, who also spoke at the gathering, said one danger was that consumers might be afforded less privacy than they would choose in a more competitive market.

The competition concerns echo those that gradually emerged in the 1990s about software and Microsoft. The worry is that as the big internet companies attract more users and advertisers, and gather more data, a powerful "network effect" effectively prevents users and advertisers from moving away from a dominant digital platform, like Google in search or Facebook in consumer social networks.

Evidence of the rising importance of data can be seen from the frontiers of artificial intelligence to mainstream business software. And certain data sets can be remarkably valuable for companies working on those technologies.

A prime example is Microsoft's purchase of LinkedIn, the business social network, for $26.2 billion last year. LinkedIn has about 467 million members, and it houses their profiles and maps their connections.

Microsoft is betting LinkedIn, combined with data on how hundreds of millions of workers use its Office 365 online software, and consumer data from search behavior on Bing, will "power a set of insights that we think is unprecedented," said James Phillips, vice president for business applications at Microsoft.

In an email to employees, Satya Nadella, Microsoft's chief executive, described the LinkedIn deal as a linchpin in the company's long-term goal to "reinvent productivity and business processes" and to become the digital marketplace that defines "how people find jobs, build skills, sell, market and get work done."

IBM has also bet heavily on data for its future. Its acquisitions have tended to be in specific industries, like its $2.6 billion purchase last year of Truven Health, which has data on the cost and treatment of more than 200 million patients, or in specialized data sets useful across several industries, like its $2 billion acquisition of the digital assets of the Weather Company.

IBM estimates that 70 percent of the world's data is not out on the public web, but in private databases, often to protect privacy or trade secrets. IBM's strategy is to take the data it has acquired, add customer data and use that to train its Watson artificial intelligence software to pursue such tasks as helping medical researchers discover novel disease therapies, or flagging suspect financial transactions for independent auditors.

"Our focus is mainly on nonpublic data sets and extending that advantage for clients in business and science," said David Kenny, senior vice president for IBM's Watson and cloud businesses.

At Google, the company's drive into cloud-delivered business software is fueled by data, building on years of work done on its search and other consumer services, and its recent advances in image identification, speech recognition and language translation.

For example, a new Google business offering — still in the test, or alpha, stage — is a software service to improve job finding and recruiting. Its data includes more than 17 million online job postings and the public profiles and résumés of more than 200 million people.

Its machine-learning algorithms distilled that to about four million unique job titles, ranked the most common ones and identified specific skills. The job sites CareerBuilder and Dice are using the Google technology to show job seekers more relevant openings. And FedEx, the giant package shipper, is adding the service to its recruiting site.

That is just one case, said Diane Greene, senior vice president for Google's cloud business, of what is becoming increasingly possible — using the tools of artificial intelligence, notably machine learning, to sift through huge quantities of data to provide machine-curated data services.

"You can turn this technology to whatever field you want, from manufacturing to medicine," Ms. Greene said.

Fei-Fei Li, director of the Stanford Artificial Intelligence Laboratory, is taking a sabbatical to become chief scientist for artificial intelligence at Google's cloud unit. She sees working at Google as one path to pursue her career ambition to "democratize A.I.," now that the software and data ingredients are ripe.

"We wouldn't have the current era of A.I. without the big data revolution," Dr. Li said. "It's the digital gold."

In the A.I. race, better software algorithms can put you ahead for a year or so, but probably no more, said Andrew Ng, a former Google scientist and adjunct professor at Stanford. He is now chief scientist at Baidu, the Chinese internet search giant, and a leading figure in artificial intelligence research.

Rivals, he added, cannot unlock or simulate your data. "Data is the defensible barrier, not algorithms," Mr. Ng said.

Yahoo Says Hackers Stole Data on 500 Million Users in 2014

The New York Time, By Nicole Perlroth, Sept. 22, 2016

SAN FRANCISCO — Yahoo announced on Thursday that the account information of at least 500 million users was stolen by hackers two years ago, in the biggest known intrusion of one company's computer network.

In a statement, Yahoo said user information — including names, email addresses, telephone numbers, birth dates, encrypted passwords and, in some cases, security questions — was compromised in 2014 by what it believed was a "state-sponsored actor."

While Yahoo did not name the country involved, how the company discovered the hack nearly two years after the fact offered a glimpse at the complicated and mysterious world of the underground web.

The hack of Yahoo, still one of the internet's busiest sites with one billion monthly users, also has far-reaching implications for both consumers and one of America's largest companies, Verizon Communications, which is in the process of acquiring Yahoo for $4.8 billion. Yahoo Mail is one of the oldest free email services, and many users have built their digital identities around it, from their bank accounts to photo albums and even medical information.

Changing Yahoo passwords will be just the start for many users. They'll also have to comb through other services to make sure passwords used on those sites aren't too similar to what they were using on Yahoo. And if they weren't doing so already, they'll have to treat everything they receive online with an abundance of suspicion, in case hackers are trying to trick them out of even more information.

The company said as much in an email to users that warned it was invalidating existing security questions — things like your mother's maiden name or the name of the street you grew up on — and asked users to change their passwords. Yahoo also said it was working with law enforcement in their investigation and encouraged people to change up the security on other online accounts and monitor those accounts for suspicious activity as well.

"The stolen Yahoo data is critical because it not only leads to a single system but to users' connections to their banks, social media profiles, other financial services and users' friends and family," said Alex Holden, the founder of Hold Security, which has been tracking the flow of stolen Yahoo credentials on the underground web. "This is one of the biggest breaches of people's privacy and very far-reaching."

The Yahoo hack also adds another miscue to what has been a troubled sale of a long-troubled company. In July, Verizon said it would acquire the internet pioneer, roughly a month before Yahoo security experts started looking into whether the site had been hacked. It is unclear what effect, if any, the breach will have on Yahoo's sale price.

In a statement on Thursday, a Verizon spokesman, Bob Varettoni, said his company learned of the breach of Yahoo's systems only two days ago and had "limited information and understanding of the impact."

It is unclear whether security testing — such as a test to see if security experts could break into the Yahoo network — was performed as part of Verizon's due diligence process before it agreed to the acquisition.

But such security is often overlooked by investors, even though breaches can result in stolen intellectual property, compromised user accounts and class-action lawsuits. To date, no law requires such security checks as part of due diligence.

"Cybersecurity can absolutely affect a valuation, and these are important questions that investors need to be asking," said Jacob Olcott, vice president of BitSight Technologies, a security company.

Yahoo said it learned of the data breach this summer after hackers posted to underground forums and online marketplaces what they claimed was stolen Yahoo data. A Yahoo security team was unable to verify those claims. But

what they eventually found was worse: a breach by what they believe was a state-sponsored actor that dated back to 2014.

A potential breach of Yahoo's systems was first reported by the tech news site Recode early Thursday morning.

The first sign that something was amiss appeared in June, when a Russian hacker who goes by the user name Tessa88 started mentioning, in underground web forums, a new trove of stolen Yahoo data, Mr. Holden said. In July, Tessa88 supplied a sample of the stolen collection to people in the so-called underground web for authentication.

The sample contained valid Yahoo user accounts, but it was unclear whether the data was from a breach of a third-party service or Yahoo itself. And it was not clear whether it came from a recent Yahoo breach or a previous incident in 2012, when the internet service acknowledged that more than 450,000 user accounts were compromised.

Then, in August, a second hacker who goes by the alias Peace of Mind began offering a large collection of stolen Yahoo credentials — including user names, easily cracked passwords, birth dates, ZIP codes and email addresses — on a site called TheRealDeal, where hackers can buy and sell stolen data, Mr. Holden said.

TheRealDeal uses Tor, the anonymity software, and Bitcoin, the digital currency, to hide the identities of buyers, sellers and administrators who are trading attack methods and stolen data.

After looking into that data, Yahoo did not find evidence that the stolen credentials came from its own systems. But it did find evidence of a far more serious breach of its systems two years earlier.

Two years is an unusually long time to identify a hacking incident. According to the Ponemon Institute, which tracks data breaches, the average time it takes organizations to identify such an attack is 191 days, and the average time to contain a breach is 58 days after discovery.

Security experts say the breach could bring about class-action lawsuits, in addition to other costs. An annual report by the Ponemon Institute in July found that the costs to remediate a data breach is $221 per stolen record. Added up, that would top Yahoo's $4.8 billion sale price.

Thursday afternoon, Senator Mark R. Warner, a Democrat from Virginia and former technology executive, issued a statement that said the "seriousness of this breach at Yahoo is huge."

He weighed in with a call for a federal "breach notification standard" to replace data notification laws that vary by state. Senator Warner added that he was "most troubled" that the public was only learning of the incident two years after it happened.

Michael J. de la Merced contributed reporting in San Francisco.

Facebook Faces Growing Pressure Over Data and Privacy Inquiries

The New York Times, By Cecilia Kang, March 20, 2018

WASHINGTON — Federal regulators and state prosecutors are opening investigations into Facebook. Politicians in the United States and Europe are calling for its chief executive, Mark Zuckerberg, to testify before them. Investors have cut the value of the social networking giant by about $50 billion in the past two days.

They are all focused on the same thing: whether Facebook mishandled users' data.

Facebook has built its highly profitable social network off its users, selling advertisements based on their ages, interests and other details. But the scrutiny over the company's vast trove of personal data — following a report that a political consulting firm had improperly obtained information of 50 million users — is taking direct aim at that lucrative formula.

"Investors are reacting to fears of regulation and the consequences of regulation," said Brian Wieser, a senior research analyst at Pivotal Research Group. "The scale of errors can only lead one to conclude these are systemic problems."

So far, most of the social network's top executives have been silent. Mr. Zuckerberg, its founder, and Sheryl Sandberg, his top deputy, have not made any public statements in recent days. The pair did not appear at an employee meeting on Tuesday in Menlo Park, Calif., where the company is based.

At the meeting, employees asked questions about the continuing internal investigation into the use of Facebook data by the political consulting firm Cambridge Analytica. The firm, which was tied to President Trump's 2016 campaign, used the data to target messages to voters.

The meeting, which included Facebook's deputy counsel, Paul Grewal, largely focused on the steps that Facebook was taking to ensure its data could no longer be misused by independent researchers, according to Facebook employees in attendance. Mr. Zuckerberg was expected to address employees on Friday, when the company holds an all-hands meeting.

The company has faced internal dissent over its broader role in spreading disinformation during the presidential

campaign and its response to it. The tensions have prompted the planned departure of Alex Stamos, Facebook's chief security officer, who plans to leave in August.

The pressure on Facebook has been building for years.

It started in the European Union, where regulators have taken an aggressive attitude toward Facebook and other American technology giants for their sway over the region's 500 million people. The company has been the subject of several privacy investigations and charges by European regulators. Europe has approved a new privacy law, which takes effect in May, that will give users of Facebook, Google and other internet services more control over how their data is collected and what Silicon Valley companies know about them.

After the 2016 presidential campaign, lawmakers at home joined the chorus of critics, citing the company's role in Russia's disinformation efforts. The social network was one of the top tools used by Russians to spread false news, and the company's executives have struggled to explain what happened and how they would prevent foreign interference in the future.

The Cambridge Analytica revelations have forced Facebook to scramble to assuage fresh concerns by regulators and lawmakers. The company is sending its representatives to Capitol Hill and arranging conversations with state attorneys general to try to answer questions about how the firm collected the information of Facebook users.

The social networking giant is also facing an investigation by the Federal Trade Commission, which is looking into whether Facebook violated an agreement with the agency, according to a person with knowledge of the inquiry.

The F.T.C. investigation is connected to a settlement the agency reached with Facebook in 2011 after finding that the company had told users that third-party apps on the social media site, like games, would not be allowed to access their data. But the apps, the agency found, were able to obtain almost all personal information about a user.

The current investigation has parallels. The information on the 50 million users was harvested in 2014 by an outside researcher, Aleksandr Kogan. Mr. Kogan, a professor at Cambridge University, paid users small sums to take a personality quiz and download an app, which collected private information from their profiles and from those of their friends. Facebook allowed that sort of data collection at the time.

Then, Mr. Kogan gave the information to Cambridge Analytica, a firm founded by Stephen K. Bannon, the former White House political adviser, and Robert Mercer, the wealthy Republican donor. Passing the information to a third party violated Facebook's policies, the company said last week.

"There are all sorts of obligations under the consent decree that may not have been honored here," said David Vladeck, a former director of consumer protection at the F.T.C.

The company could face fines of $40,000 a day per violation if the agency finds that Facebook broke the agreement.

"We are aware of the issues that have been raised but cannot comment on whether we are investigating," an F.T.C. spokeswoman said in a statement on Tuesday. "We take any allegations of violations of our consent decrees very seriously."

Facebook said it expected to receive questions from the F.T.C. related to potential violations of the 2011 consent decree. "We remain strongly committed to protecting people's information," Facebook's deputy chief privacy officer, Rob Sherman, said in a statement. "We appreciate the opportunity to answer questions the F.T.C. may have."

The F.T.C. inquiry is just one piece of the regulatory backlash. On Tuesday, the New York attorney general, Eric T. Schneiderman, announced that he was joining the Massachusetts attorney general, Maura Healey, in an investigation into whether Facebook had failed to protect the privacy of users in those states. New Jersey's attorney general announced a similar investigation.

Mr. Schneiderman and Ms. Healey sent a letter to Facebook on Tuesday that demanded records of the communications between the company and Cambridge Analytica.

"Consumers have a right to know how their information is used — and companies like Facebook have a fundamental responsibility to protect their users' personal information," Mr. Schneiderman said. "Today's demand letter is the first step in our joint investigation to get to the bottom of what happened."

There have also been growing calls for the top leadership at Facebook to appear before American and British lawmakers to testify about the company and Cambridge Analytica....

Matthew Rosenberg contributed reporting from Washington, and Sheera Frenkel from San Francisco.