A HOOVER INSTITUTION ESSAY

The Business of Knowing

PRIVATE MARKET DATA AND CONTEMPORARY INTELLIGENCE

KLON KITCHEN

Aegis Series Paper No. 2110

In January 2021, the United States Defense Intelligence Agency (DIA) acknowledged that it buys Americans' location data generated by their phones. Assuring legislators in a letter that "personnel can only query the US location database when authorized through a specific process," the DIA also argues that Fourth Amendment requirements for a warrant before collecting this information do not apply, because they are purchasing this data as a service and not using the power of law to compel its acquisition.[1]

Employing similar logic, the Department of Homeland Security (DHS), the Internal Revenue Service (IRS), the Federal Bureau of Investigation (FBI), the Drug Enforcement Administration (DEA), and other government organizations are also purchasing private market data (PMD)—data that is generated by consumers, companies, and other entities and that is collected, collated, analyzed, and sold by technology companies and data brokerage services. This, of course, is raising many concerns and questions. "It's critical we uncover how federal agencies are accessing bulk databases of Americans' location data and why," Nathan Freed Wessler, senior staff attorney with the American Civil Liberties Union's Speech, Privacy, and Technology Project, said in a statement.[2] "There can be no accountability without transparency."[3]

Some will assume these practices illustrate a federal government run amok, intent on trampling Americans' constitutionally protected rights under the guise of "national security." Others will view cries of *tyranny*! and warnings about the "deep state" as nothing more than naivete about the realities of a dangerous world or fearmongering for political advantage. But the issue is more complicated, and there is another side of the story. Government access to PMD does implicate liberty concerns, but it also implicates security issues that require serious consideration if this constitutionally induced tension is to be properly balanced.

This paper argues that US government access to at least some private market data—and the limiting of foreign access to this same information—is essential for national security. It also argues, however, for a refined awareness that acknowledges the privacy we have already lost and that implements greater government oversight and accountability. It must also be said that this paper provokes more questions than it

answers. It does not exhaustively assess or explain many of the relevant facts, trends, issues, and implications cited. The aim here is to abstract from nuance and detail to explain how our nation has come to this place, and to emphasize the security implications of our chosen path forward.

The Proliferation of PMD and of Its Value for "Knowing"

In 2018, people created, captured, copied, and consumed 33 zettabytes (ZB) of data—approximately 33 trillion gigabytes or 128,906,250,000 maxed-out iPhone 12s' worth of information.[4] This number jumped to 59 ZB in 2020 and is predicted to hit 175 ZB by 2025. Put another way: Humans currently produce 2.5 quintillion bytes of data every day.[5] If you laid flat 2.5 quintillion pennies, you could cover the earth's surface five times. By 2025, this number is projected to be 463 exabytes every day. Again, for reference: If a gigabyte is the size of the earth, an exabyte is the size of the sun—and you can fit about 1.3 million earths in the sun.

To put it into even more accessible metrics, in *every minute* of *every day* in 2020, users uploaded 500 hours of video to YouTube, sent 41 million messages on WhatsApp, uploaded 147,000 photos to Facebook, installed TikTok 2,704 times, submitted 69,000 applications on LinkedIn, and hosted 208,000 Zoom meetings.[6] *Every* minute. *Every* day. And this is only the beginning.

As fifth generation (5G) and subsequent telecommunications networks that can transport even more data come online, the oft-promised "Internet of Things" (IoT)—a world where the internet is not just a place you go on your phone, tablet, or laptop, but where it is everywhere, connecting almost everything, and is assumed the way one assumes air-conditioning when you walk into a building—is projected to include more than 30.9 billion IoT devices globally by 2025.[7] We are not just awash in data; we are drowning in it, and the flood is rising exponentially.

That does not mean, however, that we are not leveraging this data. Quite the opposite in fact; whole economies are being built on this information that, as we will see, is becoming a critical national resource. But data are not most valuable in isolation. Data's true utility is realized when data are collected, collated, analyzed, and wrung dry of their attendant insights. These services are being offered by a growing number of technology companies and data brokers, and they are redefining economies and modern notions of what can be known and hidden about ourselves.

There are some 4,000 data brokerage companies around the world, with 87 percent of those companies headquartered in the United States.[8] Just one of these data brokers, estimates the US Federal Trade Commission (FTC), "has 3000 data segments for nearly every U.S. consumer."[9] Another "has information on 1.4 billion consumer transactions

and over 700 billion aggregated data elements."[10] And still another "adds three billion new records each month to its databases."[11] One of the largest of these brokers, Acxiom, has 23,000 servers collecting and analyzing data on more than 500 million consumers worldwide.[12] All of this adds up to an industry worth more than $200 billion that can accurately be described as the beating heart of the "knowledge economy."[13]

A key portion of this industry—and a part that helpfully illustrates just how valuable this information can be—is sometimes referred to as programmatic marketing or the programmatic web. Programmatic marketing is the use of artificial intelligence (AI) and robust data sets to enable highly tailored marketing based on a consumer's demographics, attitudes, and behaviors, as understood by an analysis of their digitized data. Programmatic marketing is why women between the ages of nineteen and thirty-six receive ads for baby clothes after they search for "best folic acid supplements." It is why men who are assessed to have a high likelihood of prostate cancer receive unsolicited online ads for erectile dysfunction drugs. And it is why ads for those shoes you looked at three weeks ago appear as you read the *New York Times* online.

Thomas Davenport, Abhijit Guha, and Dhruv Grewal have explained how companies can better use data and AI for programmatic marketing to improve their bottom lines.[14] They divide these tools into two general types: task automation and machine learning. Task automation applications "perform repetitive, structured tasks that require relatively low levels of intelligence," according to the article.[15] "They're designed to follow a set of rules or execute a predetermined sequence of operations based on a given input, but they can't handle complex problems such as nuanced customer requests."[16] Examples would include a customer relationship manager program that automatically sends an email to new customers or basic consumer service chatbots like Facebook's Messenger bots.

Machine learning algorithms "are trained using large quantities of data to make relatively complex predictions and decisions. Such models can recognize images, decipher text, segment customers, and anticipate how customers will respond to various initiatives, such as promotions."[17]

Summarizing the utility of these applications, the authors are clear about their value:

> AI can streamline the sales process by using extremely detailed data on individuals, including real-time geolocation data, to create highly personalized product or service offers. Later in the journey, AI assists in upselling and cross-selling and can reduce the likelihood that customers will abandon their digital shopping carts. For example, after a customer fills a cart, AI bots can provide a motivating testimonial to help close the sale—such as "Great purchase! James from Vermont bought the same mattress." Such initiatives can increase conversion rates fivefold or more.

> After the sale, AI-enabled service agents from firms like Amelia (formerly IPsoft) and Interactions are available 24/7 to triage customers' requests—and are able to deal with fluctuating volumes of service requests better than human agents are. They can handle simple queries about, say, delivery time or scheduling an appointment and can escalate more-complex issues to a human agent. In some cases AI assists human reps by analyzing customers' tone and suggesting differential responses, coaching agents about how best to satisfy customers' needs, or suggesting intervention by a supervisor.[18]

In many ways, we are only at the forefront of programmatic marketing. As daily life becomes more digitized and as companies become more adept at collecting and leveraging our "digital exhaust," programmatic marketing will represent an unprecedented source of insight into our individual and our collective lives. This data can enable a near-total reconstruction of an individual's identity, location history, interpersonal relationships and networks, entertainment and purchasing preferences and habits, and even future economic, social, and political outcomes.

Facebook is a familiar example of the power and value of data. By creating an account and filling out a basic profile, the social media company learns a user's name, birth date, phone number, email address, contacts, schools attended, current and past occupations, relationship status, hometown, current city of residence, physical address, birth name, personal website, and other social media profiles. As you continue to use the site, Facebook learns where you like to visit, shop, and eat because you check in at these locations or post pictures of your experiences. Even if you do not post your location and even if you decline permission to share your GPS position, the company is able to follow your location by tracking the IP addresses and other information from the devices you use to access the social media service.

If you use Facebook Messenger to chat or to call your friends, the company says it does not record the content of those interactions, but it does know how often you speak with a contact and for how long. As you post and share content, the company learns even more about your religious, social, and political views, where and how you consume media, and what content you find most engaging. The company then combines this information with other "partner data," including information from other apps and even offline actions and purchases. And all of this is applied to more than 2.85 billion monthly active users globally—continually adding to and refining the Facebook social graph: a sophisticated graph of the social relations and interactions between all of the entities on the social network.

All of this data collection translates into meaningful value for Facebook. In 2020, Facebook generated nearly $84.2 billion in ad revenues—nearly 90 percent of the company's total revenue—and the company accounts for nearly 10 percent of all

digital advertising globally.[19] And this is just one company in a growing constellation of businesses who specialize in data generation, collection, and utilization. In fact, global programmatic advertising spending has almost doubled in the last four years and is expected to reach $155 billion by the end of this year.[20]

The simple but profound truth illustrated in this example is that modern marketing is fundamentally an "intelligence" operation. Governments around the world employ millions of people tasked with collecting, understanding, predicting, and shaping human behavior and events; but the private sector is pioneering this art and science and is functionally disrupting the state's monopoly on this critical capability. Even more, the data itself is overwhelmingly being generated and held in the commercial sector, where it is in some ways easier and in some ways harder to acquire.

The Need to "Know" Everything and the Promise of AI for National Security

Knowledge has always been a means to power. The more one knows, the better one can understand a situation, a challenge, an opportunity, or a risk. The gathering of knowledge, then, has always been a defining feature of national security and of American national security, specifically. After all, it is very difficult to defend against threats or to seize opportunities if you do not know about them.

In this vein, before the United States became a nation, General George Washington wrote of the "advantage of obtaining the earliest and best Intelligence of the designs of the Enemy," and charged Nathaniel Sackett with the creation of what would eventually become the Culper Spy Ring.[21] This and other intelligence operations were so successful that, at the end of the Revolutionary War, British Major George Beckwith concluded, "Washington did not really outfight the British. He simply out-spied us."[22] The value of intelligence to American security has persisted ever since.

The US intelligence community budget was $85.8 billion in 2020, spread across eighteen member departments and agencies, with at least 263 discrete intelligence organizations being established or restructured since 2001.[23] This sprawling enterprise is arrayed against an equally diverse set of issues, according to the Office of the Director of National Intelligence, including Russia, China, North Korea, Iran, Western isolationism, biological/chemical/nuclear WMDs, outer space, cyberspace, artificial intelligence, quantum computing, automation, nanotechnology, biotechnology, global inequality, violent extremism, migration, urbanization, climate change, pandemics, and transnational crime.[24] In fact, it is not hyperbole to assert that the United States has the largest, most diverse set of national interests—and, therefore, corresponding intelligence requirements—of any nation in the history of the world. This unprecedented interest and capacity also create an unending demand for information.

Importantly, it is essential to understand that the US intelligence community is tasked with much more than the anti-terrorism operations that are featured in pop culture. American policy makers lean on intelligence to inform their decisions on a much broader set of national security issues that increasingly intersect with an even broader array of facts and topics. Explaining this reality back in 2014, the DIA's then chief analytic methodologist, Josh Kerbel, observed the following:

> Today, however, the [intelligence community] no longer has the luxury of watching a single discrete entity that demands classified collection in order to obtain relevant data. There is a much more expansive range of interconnected and complex challenges. These challenges—economic contagion, viral political and social instability, resource competition, migration, climate change, transnational organized crime, pandemics, proliferation, cyber security, terrorism, etc.—are interdependent phenomena, not discrete "things." . . . Intelligence analysts must be capable of thinking creatively—holistically and synthetically across traditional boundaries. The long-held emphasis on reductive thinking that breaks issues into discrete pieces—reinforced by the compartmentalization associated with classified information—is no longer sufficient.[25]

Kerbel's point is that modern intelligence must account for the growing interconnectedness of the world and of its attendant challenges. This, he argues, requires the intermingling of unclassified and classified data "holistically and synthetically" to enable complex understanding of complex problems. Intelligence must evolve, and it is.

But what is *intelligence*? It is necessarily more than data. It is, instead, data leveraged and applied. For national security purposes, it is not enough to *know* a fact. That fact must have context so that it is properly understood. Its relevance to mission requirements and the opportunities and risks created by its acquisition and use must also be assessed. Finally, information must be *actionable*, that is, it must enable action that improves—or at least is thought to improve—the national security. In this sense, intelligence is not a single piece of information but is instead the product of data being pooled together in a manner that provides insights and then enables action.

A definition of *intelligence* from the Central Intelligence Agency (CIA) is clarifyingly simple: "Reduced to its simplest terms, intelligence is knowledge and foreknowledge of the world around us—the prelude to decisions and action by U.S. policymakers."[26] If data leveraged and applied provides "knowledge and foreknowledge of the world around us,"[27] then there is good reason to believe we are on the cusp of a golden age of intelligence—because, as we have seen, we are awash in data about our world.

But the US intelligence community faces a two-sided challenge in this regard: First, it cannot adequately process and use the data it has; and second, it is struggling to gain access to important nonclassified data sets—such as private market data—that could provide material advantage. The first is a technical challenge while the second is a political and legal one.

When it comes to better leveraging the data it has, the intelligence community, like the private sector, is placing its hopes in AI. Former Director of National Intelligence Dan Coats and former Principal Deputy Director of National Intelligence Susan Gordon outline the intelligence community's plight clearly:

> Closing the gap between decisions and data collection is a top priority for the Intelligence Community (IC). The pace at which data are generated and collected is increasing exponentially—*and the IC workforce available to analyze and interpret this all-source, cross-domain data is not . . . the IC must adapt to the rapid global technological democratization in sensing, communications, computing, and machine analysis of data. These trends threaten to erode what were previously unique USIC capabilities and advantages*; going forward, we must improve our ability to analyze and draw conclusions from IC-wide data collections at scale.[28]

Put simply: The intelligence community believes that emerging technologies are essential for the production of timely and valuable intelligence and that a failure to leverage these tools risks its irrelevance and the nation's security. To this end, the intelligence community has developed the Augmenting Intelligence using Machines (AIM) strategy, which explains how it intends to develop and to utilize *a*rtificial intelligence, process *a*utomation, and intelligence community officer *a*ugmentation (AAA) technologies to achieve its mission. As the intelligence community explains:

> The AIM initiative will enable the IC to fundamentally change the way it produces intelligence. We will achieve superiority by adopting the best available commercial AI applications and combining them with IC-unique algorithms and data holdings to augment the reasoning capabilities of our analysts. Simply stated, our goal is the following: "If it is knowable, and it is important, then we know it."[29]

The AIM strategy then provides four "primary investment objectives" that are essential for success. First, the IC must lay a digital foundation for long-term "science and technical intelligence." This involves the mundane, but critically important, acts of building cloud computing and other infrastructure, normalizing data standards, expanding government understanding of commercial offerings and supply chains, and baselining US and foreign AI capabilities and programs.

The second objective calls for the IC to expand its use of commercial and open-source AI. Agile and rapid acquisition is deemed critical for this requirement. Relatedly, the third AIM objective focuses on breaking down data-sharing barriers within the IC, with a special emphasis on the development of AI solutions that can ingest and process data from across all intelligence sources.

The fourth and final objective sets the stage for long-term thriving by requiring ongoing research and investment in AI models that go beyond simply "fusing" information, but that actually enable human analysts to better discover goals and intent or to extract entity information from incomplete or multimodal data.[30]

The reader need not fully understand each of these objectives, or even the larger AIM strategy. What is important to understand is that the intelligence community believes it must take significant and sustained action if it is to be effective going forward. Massive investments, new partnerships, and fundamental changes to established methodologies are deemed critical for future national security. If the director of the National Geospatial-Intelligence Agency was correct, for example, when he publicly estimated that the current acceleration of collection will require more than eight million imagery analysts by 2037 (an impossible demand to meet), it is easy to understand why the intelligence community feels such urgency and is placing such hope in the promise of artificial intelligence.[31]

But even if the intelligence community is able to meet the technical challenge of better leveraging all the data it has, it still faces the political and legal challenge of getting greater access to data that would significantly improve its ability to protect the nation—particularly data that is generated, collected, and analyzed in the private marketplace.

Foreign intelligence agencies like the CIA or the National Security Agency enjoy very broad collection authorities when it comes to non-US citizens. Domestic intelligence agencies like the DHS and the FBI have more constraints—especially when it comes to US citizens—but are still able to conduct extensive surveillance and analysis, when necessary, within existing legal frameworks. The need, then, for greater access to PMD is not primarily driven by tactical demands (though it would be helpful here too) but, instead, by the growing need for deep awareness at scale.

Twenty years after 9/11 the American government is well practiced and well enabled to do the type of "man-hunting" intelligence work that is featured so prominently in popular entertainment. But the return of so-called great power competition with other nations is reminding policy makers that true national security is not contained only within the need to "find, fix, and finish" an individual target—it also includes being

able to understand, predict, and influence whole governments and populations, and private market actors are uniquely capable of collecting and using the data underlying such capabilities.

Specifically, private market data offers an appealing opportunity for the intelligence community to develop at-scale intelligence because it is unclassified, "rich," and recent.

First, private market data is unclassified—meaning it can be easily used and shared. This information is typically freely (if not always knowingly) provided by users in exchange for services, and most terms of service agreements allow the collecting entities to use or to sell this information in whatever way they choose. Anyone who purchases this data, likewise, has minimal constraints on what they can do with this information and whom they can sell it to or share it with. This agility and shareability is very attractive to an American government that is routinely beset by information silos and bureaucratic barriers to essential collaboration. The unclassified nature of this information also allows this data to be intermingled with other datastores, further enabling the data "fusion" and analytic sharing that is called for in the AIM strategy discussed above.

Second, private market data is "rich." This is true in both volume and detail. PMD is frequently collected on a massive scale (remember the FTC findings mentioned earlier) and this is important for identifying trends and gleaning insights at a societal level. Again, we have already considered the extreme detail of this data, so further discussion is not needed. The salient point of this "richness" is that when this volume of highly detailed data is combined with modern and emerging processing capabilities, it yields previously unimagined awareness at the macro, mezzo, and micro levels of the world.

Third, private market data is recent. The "every minute of every day" statistics shared earlier illustrate the volume of new PMD constantly being generated.[32] And that is to say nothing of the metadata—data that gives information about and describes other data—accompanying this content. This constantly refreshing torrent of information can provide insights into virtually every aspect of people's, and a nation's, economic, social, and political life. For an intelligence enterprise tasked with a real-time understanding of geopolitical realities strategically, operationally, and tactically, private market data constitutes an unparalleled pool of insights that is tantalizingly within reach.

The intelligence community's growing "need to know" and the emerging ubiquity of data together capture the proper context for understanding the government's attraction toward private market data. Here are two illustrations of how the government might specifically use this data to advance the nation's security.

Imagine the FBI learns that a known foreign weapons proliferator is attempting to supply a domestic terrorist group with radiological materials so that they can attack the US Senate with a "dirty bomb." It also discovers that this proliferator is attempting to use a known human-smuggling network to infiltrate the United States and to deliver this radiological material to his buyer. Now assume the Bureau has access to a facial recognition tool that scrapes social media and other open-source data sets and is able to identify the ringleader of the human-smuggling network by comparing a partial mirror reflection in a child exploitation video with a Facebook picture from another user that just so happens to capture the criminal in the background, establishing his presence at the time and location of the explicit video. This allows the ringleader to be identified, located, and arrested. Follow-on analysis not only allows law enforcement to disrupt the human-smuggling ring but also to lure the weapons proliferator and the domestic terrorists into a sting that prevents the US Senate attack, liberates scores of women and children, and results in multiple arrests and convictions.

Or, consider a larger geopolitical challenge. Imagine the US intelligence community has access to decades of agriculture, climate, and economic trade data that has been collected by dozens of private market sources, including "smart" farm equipment, digitized trading markets, and industry association reporting. Now imagine this data has been pooled and fused by the IC, allowing them to alert the president to a high risk of famine within a partner nation that, if allowed to take hold, would likely result in large-scale death, massive refugee migration into neighboring countries, and the significant weakening—possibly even the downfall—of a friendly government in a strategically important region. But because this warning was possible, international aid and support were mobilized, the crisis was averted, and the improved alliance enabled the United States even greater influence in the region.

Frankly, these two examples are narrow and are relatively simple applications of PMD. Far more sophisticated examples will be possible as more data is made available and as AI capabilities develop. But both of these examples are rooted in real intelligence challenges and demonstrate the potential impact of government access to private market data. Now imagine if the government had failed to detect and disrupt either of these challenges—both could have catastrophic consequences.

The utility of PMD to modern intelligence does not, however, ameliorate the discomfort many feel regarding US government access to this data and the capabilities it is generating. This is why careful oversight will be essential.

Where We Are and What We Must Do

Concerns about the loss of privacy and liberty are well founded, and the American ethos has always suspected the accumulation of power by the state. The Constitution is

primarily a restraining document on the government. It does not exhaustively list all of a citizen's rights; instead, it lists a limited number of specific powers and authorities of the state for the purposes of the common defense and ordered liberty.

But the growing scope of threats to the common defense and to our ordered liberty—alongside the undeniable value of PMD to securing these same objects—suggests that a refinement of the "social contract" is not only in order but is already occurring because the underlying drivers—data proliferation, the declining capacity of the US intelligence community to achieve its mission, and the migration of "intelligence" into the private sector—are only growing stronger. This, then, requires a clear understanding of where we now stand and of what we must now do.

First, Americans have already willingly ceded much of their privacy—at least as it has been popularly understood—to both governmental and corporate powers. I have discussed at length the troves of data that are collected and analyzed and what can be done with these insights. Shoshana Zuboff claims we now live in an age of "surveillance capitalism," which she defines as follows:

> 1. A new economic order that claims human experience as free raw material for hidden commercial practices of extraction, prediction, and sales; 2. A parasitic economic logic in which the production of goods and services is subordinated to a new global architecture of behavioral modification; 3. A rogue mutation of capitalism marked by concentrations of wealth, knowledge, and power unprecedented in human history; 4. The foundational framework of a surveillance economy; 5. As significant a threat to human nature in the twenty-first century as industrial capitalism was to the natural world in the nineteenth and twentieth; 6. The origin of a new instrumentarian power that asserts dominance over society and presents startling challenges to market democracy; 7. A movement that aims to impose a new collective order based on total certainty; 8. An expropriation of critical human rights that is best understood as a coup from above: an overthrow of the people's sovereignty.[33]

You need not fully embrace Zuboff's admittedly dire description to agree with her core claim that society is being reshaped through the generation and collection of private market data.

And people are feeling this change. According to Pew polling, 81 percent of polled Americans believe "they have little/no control" over what data is collected from them.[34] Another 81 percent believe the "potential risks" of data collection "outweigh the benefits."[35] More than three-quarters are "very/somewhat concerned" about how this data is collected.[36] And nearly six in ten say "they have very little/no

understanding" about how this information is used.[37] So, clearly, there is broad-based recognition that large-scale data collection is eroding personal privacy.

But these concerns are not having an obvious impact on people's behavior. The number of American adults who own a smartphone has doubled since 2011 to nearly 85 percent.[38] Social media usage is also booming, with 81 percent of Americans on YouTube, 69 percent on Facebook, 40 percent on Instagram, 31 percent on Pinterest, and 21 percent on Chinese-owned TikTok.[39] Since 2016, Facebook has endured multiple scandals about its data collection and security—including the infamous Oxford Analytica fiasco and reports about it paying 13- to 17-year-olds $20 per month in exchange for nearly unfettered access to their mobile information—and yet its user base and profits have grown vastly during this same time period. In April 2021, Facebook reported more than $26 billion in revenue, which is a 48 percent increase over the previous year.[40]

These and similar statistics do not point to a market failure; they point to a market decision. As concerned as Americans are about the collection and use of their data, they are not sufficiently concerned to deny themselves the conveniences and benefits of the apps and services that harvest this data. This means, as Julia Angwin observed in *Dragnet Nation*, that people have reconciled themselves to a world in which you "can always be found . . . watched in your own home . . . no longer keep a secret . . . be impersonated . . . be financially manipulated."[41] As disquieting as this may be, it is nevertheless a reality. Is it really surprising, then, that the US government sees this market decision and hopes that it too can benefit from this wealth of data—especially when the American people have such high expectations regarding their security?

The second reality we must reckon with is that "the common defense" now requires a greater contribution from the people. As previously stated, the United States has the largest, most diverse set of national interests—and, therefore, corresponding intelligence requirements—of any nation in the history of the world, and Americans have a very low tolerance for national security risk when push comes to shove.

To wit, after observing a decline in US public support for the dropping of two atomic bombs on Hiroshima and Nagasaki, Japan, from 85 percent in 1945 to 46 percent in 2015, Stanford scholars Scott Sagan and Benjamin Valentino wondered if this shift would hold up if Americans faced a similar challenge to World War II—drop the bomb and kill more than 100,000 Japanese or invade Japan and lose several thousand US soldiers.[42] A Stanford news article explains:

> "We wondered what would happen today if Americans were faced with a similar tradeoff," Sagan said. "Has the U.S. public really changed? Or were previous polls misleading guides to real public attitudes about nuclear weapons use?"

> Sagan's findings from a survey experiment conducted in July 2015 involved a representative sample of the U.S. public asked about a contemporary, hypothetical scenario designed to replicate the 1945 decision to drop a nuclear bomb on Hiroshima.
>
> He and Valentino created a news story in which Iran attacked a U.S. warship in the Persian Gulf, Congress declared war, and the president was presented with the option of sending U.S. troops to march into Tehran, which would lead to many American military fatalities, or dropping a nuclear weapon on an Iranian city to try to end the war.[43]

The result?

> Their findings demonstrate that, contrary to the nuclear taboo thesis, a clear majority of Americans would approve of using nuclear weapons first against the civilian population of a nonnuclear-armed adversary, even killing 2 million Iranian civilians, if they believed that such use would save the lives of 20,000 U.S. soldiers.
>
> In addition, contrary to the principle of noncombatant immunity, an even larger percentage of Americans would approve of a conventional bombing attack designed to kill 100,000 Iranian civilians in the effort to intimidate Iran into surrendering, according to Sagan.[44]

Americans feel similar urgency on broader notions of national security. Nearly 70 percent of Americans say "taking measures to protect the U.S. from terrorist attacks" is a top long-range foreign policy goal[45] and 45 percent say China is the United States' greatest enemy.[46] Another 63 percent say "the economic power of China is a critical threat to the vital interests of the U.S. in the next 10 years."[47] Finally, 70 percent of polled Americans say "international issues [are] relevant to their daily lives."[48] What is the upshot of all of this? The people of the United States have a broadly shared concern about their peace and tranquility, and when these are perceived to be credibly threatened, they have high expectations that the government will decisively act.

It should be obvious by now that PMD can greatly enhance the government's ability to meet these expectations and to stay ahead of a constantly expanding list of threats.

But PMD is a broad category, and the IC's access to it is heavily influenced by how it is collected, who collects it, where it was collected, and from whom or what it is collected. These variables must be taken into consideration.

For example, any data collected by a foreign entity—government or nongovernment—from a non-US population, individual, or other target, should be fair game for American

intelligence. There should be no constraint on their ability to buy, steal, or otherwise acquire this data because constitutional protections do not extend beyond our own citizens. Foreign-sourced data that includes US persons' data, including personally identifiable information (PII), should also be easily acquired, but will require special handling that minimizes the US persons' data. Such mitigation efforts are already integrated into the intelligence process and are easily applied here. Domestic private market data and data collection requires more protections.

Domestic intelligence agencies like the FBI and DHS should be given primary responsibility for acquiring and holding PMD from domestic sources that includes US persons' PII. This is in keeping with existing authorities and responsibilities and maintains the important distinction between domestic and foreign intelligence activities. Importantly, however, the IC must formalize capabilities and methodologies to "fuse" this data, while protecting Americans' PII, so that any insights that are relevant to the foreign intelligence mission are discovered and leveraged appropriately. Domestic data that does not include PII—such as economic data, climate data, generalized sociological statistics, and so on should generally be made available to the foreign-focused IC members either through purchase, information sharing agreements, legal mechanisms such as national security letters, or other routine channels.

But if the Leviathan is to be more heavily fed, its chains must also be reinforced. The American people can no longer accept the emaciated oversight and a near-total lack of transparency regarding the US intelligence enterprise—particularly regarding the realm of government data acquisition and use.

For starters, Congress must improve its intelligence and cybersecurity oversight. The House and Senate Select Committees on Intelligence should require an annual report from the US IC on what PMD it is accessing, how this PMD is being leveraged (with specific examples of positive and negative outcomes), how the nation's geopolitical rivals are using this information, and other relevant reporting. The Director of National Intelligence should also consider issuing an annual unclassified report cataloging the IC's PMD acquisitions and partnerships. Some will argue that in the name of protecting sources and methods this information cannot or should not be shared. On the other hand, the reality is that if the citizens of the nation do not trust the government with this data in the first place, there will be no sources and methods to protect.

Finally, Congress should adopt the Cyberspace Solarium Commission's recommendation that the House and Senate form permanent select committees on cybersecurity.[49] All cybersecurity-related budgetary and legislative jurisdiction should fall under these two committees and they would be responsible for overseeing the Executive's efforts to integrate cybersecurity strategy and policy within the government and between

government and industry. A key aspect of this role would include overseeing how government and the private sector secure the PMD they acquire and exploring new technologies and methodologies that enable PMD to be leveraged while also expanding individual anonymity (e.g., homomorphic encryption).

Many other changes are in order but cannot be exhaustively cataloged here. The fundamental point that must be reiterated, however, is that if the state requires access to Americans' PMD in order to secure the nation, the government must also be willing to constrain itself to more robust oversight and accountability. If the Leviathan cannot or will not submit, it cannot be allowed to run free. Americans decided long ago that they would rather endure threats from abroad than tyranny at home.

As the nation negotiates this new balance between security and liberty, there is one obvious action that must be taken no matter how these tensions are resolved.

Limiting Foreign Government Access to US PMD

Even if the reader is not persuaded that PMD is vital for national security, the governments of other nations certainly are. The present risks of our citizens' data being sold to foreign governments are grossly underappreciated. Although plugging this gaping hole in our data security touches on a range of hot-button issues, banning the sale of sensitive American data to adversarial governments should be an obvious priority for quick, decisive action.

Unsurprisingly, China already steals the type of bulk data sets on Americans that data brokers sell. In July of last year, FBI Director Christopher Wray noted, "If you are an American adult, it is more likely than not that China has stolen your personal data."[50] Indeed, one of the largest Chinese hacks of Americans' personal data was that of Equifax, a leading data broker, resulting in the People's Republic of China (PRC) gaining information on almost half of all Americans. The Chinese Communist Party theoretically could have legally purchased the same information, probably with greater ease. We also know from the director of the United States National Counterintelligence and Security Center that China is using both "illegal and legal means" to collect bulk personal data of the sort sold by data brokers.[51] Here is one example of how this data could be used against us.

Imagine that a hostile foreign nation is given access to huge stores of American social media data like photos, phone numbers, family members and contacts, locational data, online viewing and purchasing habits, political and social affiliations, "keyboard stroke patterns," and so on—all of which are routinely captured. Now imagine this government were to focus on the data generated around an important military installation like Fort Bragg, North Carolina—home to one of our nation's

elite special mission units. Using just this data, a sophisticated intelligence activity could begin to identify individual members of this unit and their families. They could use GPS locations (or their absence) and social media posts discussing "TDYs"[52] or "vacations" or "alone time" as a type of indication and warning notice for when members of this unit might be deploying. They could also follow the GPS locations of spouses to discover patterns of life or "inappropriate" relationships that could be leveraged for influence or blackmail. All of these, and much more nefarious deeds, are easily done with the information collected by virtually every application downloaded to a mobile phone.

Two main difficulties present themselves to redressing the issue. First, enforceability will be challenging. Data—even vast quantities of data—are notoriously "slippery," meaning it is difficult to track where it goes or what it is used for once it is transferred. While there is some ability to "hash" or "beacon" data so that it can be traced, these capabilities would be quickly overtaken by the scale of the data in question. An honest assessment must admit that, even if China is banned from purchasing American PMD, it is likely to acquire it through commercial cutouts and to continue to steal it. But imperfect security is not a justification for assuming unnecessary risk. To put it metaphorically, right now hostile regimes like those in Beijing and Moscow are making uncontested layups by purchasing US PMD. A ban on these purchases would at least push them back to the three-point line and put a hand in their face.

A second difficulty is the economic dimension. Given Chinese governments' unrestricted access to the data of companies operating in the PRC, regulations on data transfers could be disruptive and costly to a wide swath of businesses that work with companies in China.[53] Depending on the form of the restrictions, businesses from a host of other countries that deal heavily in data, like Ireland, could also suffer considerable losses along with their American counterparts.[54]

Any viable solution would have to carefully address both of these challenges, balancing business interests with enforceability and maintaining enough adaptability to account for rapidly evolving technologies and privacy concerns.

So far, a few options have emerged. A new bill would have the Secretary of Commerce identify categories of personal data that are important to protect and data-receiving countries of concern, in order to administer licenses for data export.[55] Others have suggested more intermediary measures, such as requiring data-selling companies to declare their foreign customers, or expanding the Committee on Foreign Investment in the United States process to restrict adversaries from buying their way into American data-brokering operations.[56]

Putting aside further questions of methods, however, the primary challenge to addressing the threat remains an insufficient sense of urgency. Corporate bulk data transfers don't quite trip the same alarms that hypersonic missiles do. But in a world in which data is the new oil, there is a very real sense in which these companies can sell off American security to our adversaries—with potentially devastating consequences.[57] Yet the national security dimension of data brokering is pretty straightforward: Selling Americans' sensitive data to unfriendly foreign governments is a pressing security threat that should not be permitted.

Conclusion

Data is becoming the most plenteous and valuable resource on the planet. In it, we find a seemingly inexhaustible source of insight about ourselves and the world in which we live. These insights enable amazing opportunities and advancements for human thriving. Even more, technologists are pioneering mind-boggling methods for collecting, collating, understanding, and using data—many of which would have been thought to be impossible only a decade ago.

Disruption has always been a natural part of innovation, and certainly this is the case today. Political leaders, particularly, are being forced to accept that intelligence, "knowledge and foreknowledge of the world around us—the prelude to decisions and action,"[58] is no longer the exclusive domain of governments but is, instead, a booming industry driven by private sector actors and capabilities. In recognition of this reality, the US intelligence community is turning to industry for help in fulfilling its constitutional mission to provide for the common defense. This provokes serious issues.

The IC's need for private market data (PMD) is clear. But the risks that come with government access to PMD are also clear. While the national security relevance of such access is increasingly compelling, it must be accompanied by corresponding constraints and accountability. A government unwilling to accept such restrictions and transparency inherently demonstrates that it cannot be trusted with such data.

Equally concerning is the PMD access currently enjoyed by hostile foreign governments like China. It is nothing short of madness for the US government to allow the sale of this data to entities we know are using it to imperil American people and interests. The idea that Beijing may have greater access to US PMD than the American government is obviously unacceptable and should be immediately addressed.

The American people and their government leaders cannot avoid these realities. Instead, they must adapt to them by refining our institutions and the critical balance between liberty and security. These changes necessarily require uncomfortable choices

that bring with them no ironclad assurances of safety. But while an evolution as discussed in this paper does not guarantee success, a lack of adaptation will guarantee failure.

In the final analysis, one thing is clear: Going forward, we will all be "known." It is simply a matter of by whom and for what purpose.

NOTES

Thanks to Jack Goldsmith and Andrew Keane Woods for comments on a prior draft.

1 Chris Mills Rodrigo (@millsrodrigo), Twitter (Jan. 22, 2021, 1:28 PM), https://twitter.com/millsrodrigo/status/1352684462795067393.

2 Sara Morrison, *A Surprising Number of Government Agencies Buy Cellphone Location Data. Lawmakers Want to Know Why*, Vox (Dec. 2, 2020, 4:25 PM), https://www.vox.com/recode/22038383/dhs-cbp-investigation-cellphone-data-brokers-venntel.

3 *Id*.

4 *See* David Reinsel, John Gantz & John Rydning, Data Age 2025: The Evolution of Data to Life-Critical—Don't Focus on Big Data; Focus on the Data That's Big 7 (2017), https://www.import.io/wp-content/uploads/2017/04/Seagate-WP-DataAge2025-March-2017.pdf.

5 *See* Jacquelyn Bulao, *How Much Data Is Created Every Day in 2021?*, TechJury (Aug. 6, 2021), https://techjury.net/blog/how-much-data-is-created-every-day/.

6 *Data Never Sleeps 8.0*, Domo, https://www.domo.com/learn/infographic/data-never-sleeps-8 (last visited Aug. 14, 2021).

7 Lionel Sujay Vailshery, *IoT and Non-IoT Connections Worldwide 2010-2025*, Statista (Mar. 8, 2021), https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/.

8 *See What Are Data Brokers—and What Is Your Data Worth?* [Infographic], WebFX Blog (Mar. 16, 2020), https://www.webfx.com/blog/internet/what-are-data-brokers-and-what-is-your-data-worth-infographic/; Kevin B. Johnston, *Top 15 Broker-Dealer Firms for 2020*, Investopedia (Aug. 2, 2019), https://www.investopedia.com/investing/broker-dealer-firms/.

9 Fed. Trade Comm'n, Data Brokers: A Call for Transparency and Accountability IV (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

10 *Id*.

11 *Id*.

12 *Id*. at 8.

13 *What Are Data Brokers—and What Is Your Data Worth?* [Infographic], *supra* note 8.

14 Thomas H. Davenport, Abhijit Guha & Dhruv Grewal, *How to Design an AI Marketing Strategy: What the Technology Can Do Today—and What's Next*, Harv. Bus. Rev., July–Aug. 2021, https://store.hbr.org/product/how-to-design-an-ai-marketing-strategy/S21041.

15 *Id*.

16 *Id*.

17 *Id*.

18 *Id*.

19 *See* Statista Research Department, *Facebook's Advertising Revenue Worldwide from 2009 to 2020*, Statista (Feb. 5, 2021), https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/.

20 *See* Statista Research Department, *Global Programmatic Advertising Spending from 2017 to 2021*, Statista (Apr. 1, 2021), https://www.statista.com/statistics/275806/programmatic-spending-worldwide/.

21 Natasha Bertrand & Michael B. Kelley, *This Letter from George Washington Marks the Birth of American Espionage*, Bus. Insider (Feb. 25, 2015, 6:42 AM), https://www.businessinsider.com/this-letter-from-george-washington-is-the-birth-of-american-espionage-2015-2.

22 *George Washington, Spymaster*, George Washington's Mount Vernon, https://www.mountvernon.org/george-washington/the-revolutionary-war/spying-and-espionage/george-washington-spymaster/ (last visited Aug. 14, 2021).

23 *See* Press Release, Office of the Director of National Intelligence, DNI Releases Appropriated Budget Figure for 2020 National Intelligence Program (Oct. 21, 2020), https://www.dni.gov/index.php/newsroom/press-releases/item/2161-dni-releases-appropriated-budget-figure-for-2020-national-intelligence-program; Web Politics Editor, *'Top Secret America'—Yahoo! News on the 'Top 10 Blockbuster Revelations,'* Wash. Post (July 21, 2010), http://voices.washingtonpost.com/top-secret-america/2010/07/top_secret_america_-_yahoo_new.html.

24 *See* Off. Dir. Nat'l Intel., National Intelligence Strategy of the United States of America 2019, at 4, https://www.dni.gov/files/ODNI/documents/National_Intelligence_Strategy_2019.pdf.

25 Josh Kerbel, *The US Intelligence Community's Kodak Moment*, Nat'l Int. (May 15, 2014), https://nationalinterest.org/feature/the-us-intelligence-communitys-kodak-moment-10463.

26 Cent. Intel. Agency, A Consumer's Guide to Intelligence, at vii (1999).

27 *See id*.

28 Off. Dir. Nat'l Intel., The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines, at III (2019), https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf (emphasis added).

29 *Id*. at IV (quoting Principal Deputy Director of National Intelligence Sue Gordon).

30 *Id*. at V.

31 *See* Robert Cardillo, Director, Nat'l Geospatial-Intel. Agency, Remarks at the 31st Annual Small Satellites – Big Data Conference (Aug. 7, 2017), https://www.nga.mil/news/Small_Satellites_-_Big_Data.html.

32 *Data Never Sleeps 8.0*, *supra* note 6.

33 Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, at VII (2019).

34 Brooke Auxier et al., *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*, Pew Research Center (Nov. 15, 2019), https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

35 *Id*.

36 *Id*.

37 *Id*.

38 S. O'Dea, *Smartphone Ownership in the US 2011–2021*, Statista (May 12, 2021), https://www.statista.com/statistics/219865/percentage-of-us-adults-who-own-a-smartphone/.

39 Brooke Auxier & Monica Anderson, *Social Media Use in 2021*, Pew Research Center (Apr. 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

40 Press Release, Facebook, Facebook Reports First Quarter 2021 Results (Apr. 28, 2021), https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-First-Quarter-2021-Results/default.aspx.

41 Julia Angwin, Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance 6 (2014).

42 Scott D. Sagan & Benjamin A. Valentino, *Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants*, 42 Int'l Sec. 41, 41–46 (2017).

43 Clifton B. Parker, *Public Opinion Unlikely to Curb a US President's Use of Nuclear Weapons in War, Stanford Scholar Finds*, Stan. Univ. News Serv. (Aug. 8, 2017), https://news.stanford.edu/2017/08/08/americans-weigh-nuclear-war/.

44 *Id*.

45 *Amid G7 Summit, Most Americans Confident in Biden's Handling of World Affairs*, Ipsos (June 13, 2021), https://www.ipsos.com/en-us/news-polls/g7-kicks-off-most-americans-confident-bidens-handling-world-affairs.

46 Mohamed Younis, *New High in Perceptions of China as US's Greatest Enemy*, Gallup (Mar. 16, 2021), https://news.gallup.com/poll/337457/new-high-perceptions-china-greatest-enemy.aspx.

47 *Id*.

48 *US Adults' Knowledge about the World*, Council on Foreign Rels. (Dec. 2019), https://www.cfr.org/report/us-adults-knowledge-about-world.

49 US Cyberspace Solarium Comm'n, 116th Cong., Final Report 2 (2020), https://www.solarium.gov/report.

50 Christopher Wray, Director, Fed. Bureau Investigation, The Threat Posed by the Chinese Government and the Chinese Communist Party to the Economic and National Security of the United States, Remarks at the Hudson Institute Video Event: China's Attempt to Influence US Institutions (July 7, 2020), https://www.fbi.gov/news/speeches/the-threat-posed-by-the-chinese-government-and-the-chinese-communist-party-to-the-economic-and-national-security-of-the-united-states.

51 Zach Dorfman, *China Used Stolen Data to Expose CIA Operatives in Africa and Europe*, Foreign Pol'y (Dec. 21, 2020, 6:00 AM), https://foreignpolicy.com/2020/12/21/china-stolen-us-data-exposed-cia-operatives-spy-networks/.

52 A military acronym for "temporary duty assignment," or travel.

53 *See* Klon Kitchen, *Why America Needs a Clear Policy to Deal with Chinese Cyber Security Concerns*, Nat'l Int. (Feb. 18, 2021), https://nationalinterest.org/blog/buzz/why-america-needs-clear-policy-deal-chinese-cyber-security-concerns-178434.

54 *See New US Senate Bill May Stop Ireland Processing US Data, Unless Ireland Acts on GDPR Enforcement*, Irish Council for C.L. (Apr. 15, 2021), https://www.iccl.ie/news/new-us-senate-bill-may-stop-ireland-processing-us-data-unless-ireland-acts-on-gdpr-enforcement/.

55 *See* Drew Harwell, *Wyden Urges Ban on Sale of Americans' Personal Data to 'Unfriendly' Foreign Governments*, Wash. Post (Apr. 15, 2021, 7:00 AM), https://www.washingtonpost.com/technology/2021/04/15/personal-data-foreign-government-ban/.

56 Michael Kans, *Data Brokers and National Security*, Lawfare (Apr. 29, 2021, 8:01 AM), https://www.lawfareblog.com/data-brokers-and-national-security.

57 *See The World's Most Valuable Resource Is No Longer Oil, but Data*, Economist (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

58 *See* Cent. Intel. Agency, *supra* note 26 at vii.

hoover.org

27 26 25 24 23 22 21 7 6 5 4 3 2 1

The preferred citation for this publication is Klon Kitchen, *The Business of Knowing: Private Market Data and Contemporary Intelligence*, Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 2110 (November 30, 2021), available at https://www.lawfareblog.com/business-knowing-private-market-data-and-contemporary-intelligence.

About the Author

KLON KITCHEN

Klon Kitchen is a senior fellow at the American Enterprise Institute. Previously, Kitchen was the director of the Heritage Foundation's Center for Technology Policy and the national security advisor to Senator Ben Sasse. Kitchen has served in several positions within the intelligence community, including at the National Counterterrorism Center and the Office of the Director of National Intelligence.

The Jean Perkins Foundation Working Group on National Security, Technology, and Law

The Jean Perkins Foundation Working Group on National Security, Technology, and Law brings together national and international specialists with broad interdisciplinary expertise to analyze how technology affects national security and national security law and how governments can use that technology to defend themselves, consistent with constitutional values and the rule of law.

The group focuses on a broad range of interests, from surveillance to counterterrorism to the dramatic impact that rapid technological change—digitalization, computerization, miniaturization, and automaticity—are having on national security and national security law. Topics include cybersecurity, the rise of drones and autonomous weapons systems, and the need for—and dangers of—state surveillance. The group's output will also be published on the *Lawfare* blog, which covers the merits of the underlying legal and policy debates of actions taken or contemplated to protect the nation and the nation's laws and legal institutions.

Jack Goldsmith is the chair of the National Security, Technology, and Law Working Group.

For more information about this Hoover Institution Working Group, visit us online at http://www.hoover.org/research-teams /national-security-technology-law-working-group.

Hoover Institution, Stanford University
434 Galvez Mall
Stanford, CA 94305-6003
650-723-1754

Hoover Institution in Washington
1399 New York Avenue NW, Suite 500
Washington, DC 20005
202-760-3200

HOOVER INSTITUTION