Information Society Project
Yale Law School

DIGITAL FUTURE WHITEPAPER SERIES

Nowhere to Hide: Data, Cyberspace, and the Dangers of the Digital World

Andrew Burt

December 2020

Contents

Nowhere to Hide: Data, Cyberspace and the Dangers of the Digital World

Privacy is dead. So is trust. And you're not who you think you are.

These are the three main arguments—each deeply related—that I aim to relay in this essay.[1] Each is based upon the profound impact of digital technologies on our lives. They are the new realities of our digital age, in which the line between cyber and physical is becoming blurred, if not meaningless.

What do I mean by digital? Anything that can be digitized. And what do I mean by digitized? Anything that is represented as computer code.

When I talk about the digital transformation of our everyday lives, I'm referring to how our images are captured in the course of our daily existence by increasingly ubiquitous surveillance cameras that, as of ten years ago, were recording 4 billion hours of footage a week in the U.S. alone.[2]

I'm referencing how almost every modern method of communication, from emails to texts to social media and word processing, now involves, at minimum, the use of computers and, at maximum, the use of the protocols the Internet is based upon.[3]

I'm alluding to the so-called Internet of things, which seeks to turn everyday objects into connected devices and digitize their functions. Examples of the ever-expanding Internet of things range from the practical to the ludicrous. There are, for instance, pacemakers that control when patients' hearts beat—by one account every new such device implanted in the U.S. is now connected to the Internet.[4] There are Internet-connected cars—an estimated 64 million of them on the roads last year.[5] And there are door knobs, toilets, lightbulbs, and practically anything you can imagine.

In fact, here's a challenge: Pick a noun—any noun—and conduct an online search with that word in addition to the phrase "smart device." If experience is any guide, you will find someone, somewhere, selling a variant of it. The lesson? We live in a world that's increasingly and irresistibly digital, and this digitization is having deep and profound impacts on our daily lives.

To start with is the impact we can feel: The world is becoming more convenient. In that sense, this trend is quite positive. It's easier to pay, to communicate, and to travel. The more connected we become, the more autonomous we feel. When living through something like a global pandemic, it turns out that it's quite handy to be able to conduct a huge amount of our daily activities remotely, through one screen or several. Digitization is, in this sense, a very good thing.

But there are also downsides, many of them unseen or hard to intuit: Software-based systems have allowed small groups of technologists to dictate the combined behavior of our connected devices and to determine the structure of our online lives. It is by reaping the fruits of this power that an estimated one out of every 11,600 people in Silicon Valley is now a billionaire, granting the home and birthplace of commercial computing the highest per capita concentration of wealth anywhere in the world.[6] The fact that we live more and more of our lives online also contributes to the tenuous state of our security and privacy, as is demonstrated with every new data breach. In 2019 alone, an estimated 15 billion records were exposed online.[7]

If we are honest with ourselves—as technologists, as consumers, as policymakers—both individuals and organizations alike are lost in a sea of connected devices. But rather than take note of this disorientation, or perhaps even reorient ourselves, our collective response has been to swim farther from shore, adopting more devices and more algorithms and connecting everything we can to the Internet.[8]

Collectively—and again, if we are being honest—we simply do not know what to focus on, what precise regulation is called for, or how to protect all the data we generate.

This is a conclusion that has long been apparent to many of us. While serving in government, I once heard a senior U.S. official state that never before has so much effort resulted in so little action as with cybersecurity. Which is to say that we have been lost for quite some time.

Indeed, for as long as software has been relied upon, officials and researchers alike have been sounding alarm bells—sometimes comically, but nonetheless gravely. Here, for example, is how one Congressional report described the issue of data security: "If architects built buildings the way programmers build programs, then the first woodpecker to appear would destroy civilization."[9] This was in 1989.

Here's how the head of the Central Intelligence Agency described a variation of the same problem: "We are staking our future on a resource that we have not yet learned to protect."[10] This was in 1998.

Every year since 2013 the U.S. intelligence community has ranked cybersecurity as the greatest threat facing the United States—more significant, even, than the threat of terrorism.[11] Examples of these types of warnings are not hard to find—not because such prognostications require such foresight, but because it is not all that hard to be right about the risks of digital technologies. Their dangers are plentiful, and we use them more and more.

Yet layered underneath all our privacy and security vulnerabilities, there are also three much less obvious effects of these trends—less apparent than what I've focused on until now. I will spend the following pages outlining three of these trends. They're what I started this essay with, and they are worth repeating: Privacy is dead. So is trust. And you're not who you think you are.

After I overview each, I will make a handful of concrete suggestions about what we can and should do to address each development—as lawyers, as policymakers, and as citizens around the world. The sky may seem like it is falling in cyberspace, I will argue, and with good reason, but it need not fall as fast or land as hard.

I. Privacy Is Dead

Let's start with the end of privacy. Dating back to the seminal 1890 law review article by future Supreme Court Justice Louis Brandeis and attorney Samuel Warren, the idea of privacy has been defined as something along the lines of the right "to be let alone."[12] The right to privacy actually has a much longer history, brilliantly illustrated by James Whitman in “The Two Western Cultures of Privacy: Dignity Versus Liberty.”[13] For anyone not familiar with this piece, I highly recommend it.

But Warren and Brandeis’ notion will do for us, and for now, because it captures the fundamental assumption underlying all notions of privacy: that there is a quantifiable difference between being observed and being identified. The public world of observation is not one that we can remove ourselves from. Individuals must, in the course of their everyday interactions, venture out into the public sphere, where strangers might observe how we look, who we’re with, what we’re wearing and more. Being identified, on the other hand, is an altogether different type of activity. Identification pierces the veil, so to speak, of our anonymity—preventing us from being left alone and potentially intruding upon our “penumbral rights of ‘privacy and repose,’” as famously described in *Griswold*.[14]

But if we are being honest with ourselves, this distinction—the foundation of our very privacy—is no longer. To put it as simply as possible: If you can be observed, you can be identified.

This is thanks to a confluence of factors: an abundance of data (in a digitized world, everything we do generates data), cheaper and faster computing power and storage (also called the “cloud”), and new techniques for identification brought about by machine learning (commonly referred to as AI). Taken together, the methods for collecting almost any form of data and rendering it identifiable keep growing.

Here are just a few examples of activities that are no longer private: walking, writing (text or computer code), and even the act of owning a cell phone that is powered on.

Let’s start with walking. Late in 2018, the *Associated Press* published a report on the Chinese government’s use of gait analysis to identify individuals based on the way their bodies move when they walk. According to the report, the techniques “can identify people from up to 50 meters ([or] 165 feet) away, even with their back turned or face covered. This can fill a gap in facial recognition, which needs close-up, high-resolution images of a person’s face to work.”[15] The very act of walking in public in China, and increasingly elsewhere, is no longer a private one because of these techniques. It is, of course, no surprise that this type of technology is being perfected in China, a state whose very legitimacy relies on its ability to monitor its citizens. Data is now the lifeblood of modern software systems, and software is the exercise of power over space and time.

Now on to writing. Machine learning researchers have demonstrated that, given a baseload of enough writing samples, the authorship of new text can be identified with frightening accuracy. Here is how one group of researchers sums up the problem: "Given the increasing availability of writing samples online, our result has serious implications for anonymity and free speech—an anonymous blogger or whistleblower may be unmasked unless they take steps to obfuscate their writing style."[16]

And then, of course, there are our cell phones, the modern individual's most prized possession. So vital is possessing and using a mobile phone that there is even a name for the psychological fear caused by going without such a device: nomophobia.[17] The very act of keeping a cell phone turned on creates a record of where your cell phone is over time in relation to cellular towers, which constantly ping each phone to understand their proximity to the closest tower, plotting the movements of each device neatly on a map. As far back as 2008, researchers were describing all the ways cellular data could identify unique individuals within hundreds of thousands of records.[18] This information can, of course, reveal incredibly sensitive details, including not just an individual's identity, but also intimate patterns of life, like whether or not a detective is investigating a crime scene late one night, as one analysis of supposedly anonymous cell phone location data in New York City divulged.[19]

This is not, of course, meant to be an exhaustive list. If everything we do generates data—and data, at large enough volumes, generates insights we cannot predict—the types of intimate insights that arise from our data will only grow. And these insights will continue to *surprise* us—that is, after all, the entire value of increasingly powerful techniques like machine learning. Human minds cannot predict, or sometimes even understand, these methods.[20] Surprise is, and will remain, a central feature in the insights that sophisticated algorithms deliver.

So what does all this mean for our privacy? Or, to more closely mirror the language used by the legal system in the United States, what do these trends mean for our "reasonable expectations" related to all our data? Again, if we are being honest, it means that our expectations of privacy *must* diminish as we generate more data, which we seem committed to doing over time.[21]

And so it is hard, perhaps even impossible, to conclude anything other than that privacy as we've known it—privacy as it's been defended in the courts, conceived of in law school classrooms, and thought about in the minds of consumers—is dead. We can and should expect less and less privacy as we generate more and more data. Instead, something else must rise to take the traditional notion of privacy's place. If not, we will be left with an irrelevant legal concept that fails to apply to an increasingly urgent set of rights.

I have a few suggestions for what, exactly, that framework should look like, which I will return to at the close of this essay.

II. Trust Is Dead

Now on to trust. One of my favorite books on technology in recent years—and perhaps the most overlooked—was published by a Norwegian academic named Olav Lysne and given the extremely non-user friendly name: "The Huawei and Snowden Questions: Can Electronic Equipment from Untrusted Vendors be Verified? Can an Untrusted Vendor Build Trust Into Electronic Equipment?" The book was aimed at a technical audience, and it's no wonder that it didn't make a huge splash. But its conclusions are deeply impactful for the few readers who have made it this far into this essay.

So here is Olav's story.

A few years ago, the Norwegian government realized that it bought almost all of its critical technology from outside of Norway. With a population of just over 5 million people, it would be ludicrous and expensive to try to make all their critical technology internally—think of the software that runs things like electric stations, water pumps, cellular towers or the switches and routers that connect the country to the Internet at large. And so the Norwegian government came up with a pressing national security question: How could it trust all the technology it was increasingly reliant upon? What could Norway actually *do* to verify that the software it was using was trustworthy?[22]

That's where Olav came in. Olav helped look into the question for the government, first by chairing a commission on the subject and later culminating in "The Huawei and Snowden Questions" book.

Olav's answer was that the government simply could not verify the software it used as trustworthy. In fact, no one can. The very nature of our current software systems—and the very complexity underlying them, their supply chains and more—makes it impossible to detect intentionally inserted vulnerabilities into software by software makers.

To be clear, the paradigm that Olav was looking at is a bit different than that of traditional data security. In traditional information security, we tend to think of the adversary and the vendor as being unrelated. Malicious actors are generally conceived as third parties, looking in from the outside after a transaction between the vendor and the buyer has taken place. Instead, the question Olav focused on was this: How do we know that the people we're buying our software from haven't sabotaged it somehow? How, for example, can we prove they're not trying to surveil us? Or that there isn't an intentionally implanted "kill switch" that would render the software unfunctional in the future? Or that the software is not defrauding us in some way?

Public examples of this type of attack are few and far between, and usually involve state actors compromising supply chains. In October of 2018, for example, the cover of *Bloomberg Businessweek* featured a controversial story that alleged factories in China implanted a tiny microchip, not much bigger than a grain of rice, into circuit boards designed for data centers. According to the article, "the chips allowed the attackers to create a stealth doorway into any network that included the altered machines."[23] This insertion was said to have impacted around 30 U.S. companies, including Amazon and Apple. That would be one example of the type of worry Olav was addressing.

Another example comes from Volkswagen's so-called "dieselgate" scandal, in which the automaker reportedly inserted software into its diesel cars that intentionally defrauded emissions testers. VW's engine software then looked to see if laboratory emissions testing was being conducted, in which case it would activate its advertised emissions controls;

otherwise, the cars would emit up to 40 times more emissions in real–world driving.[24] In this way, Volkswagen fooled regulators and consumers, all of whom thought they were interacting with relatively environmentally friendly cars.

Olav examined nearly every component of a software system—from the actual software, to the code that compiles programming languages into an executable (that is, turning programming language into 0s and 1s), to the ways that software is updated and managed over its lifetime and much more. In nearly every dimension he found that a software vendor can hide malicious code from consumers with relative ease.

A few examples: Integrated circuits are the chips that make up circuit boards, and are composed of tens of thousands gates on semiconducting materials like silicon. As Olav states, "full backdoors into a system can be created with a microscopic number of additional gates on a chip."[25] In fact, adding as few as 1,341 gates to just one chip can create a backdoor into an *entire* system.

Another example: One of the most subtle methods of inserting vulnerabilities into software, according to Olav, involves making the malicious behavior dependent upon an external stimulus so the malicious code is triggered by an external event and is therefore harder to detect. Here's how he describes the possibility of detecting this type of vulnerability:

> Having this external stimulus encoded in only 512 bits would yield 13.4×10^{153} combinations. For comparison, the universe has existed for approximately 4×10^{17} seconds. If the strongest computer on Earth had started computing at the beginning of the universe, it would still not have made any visible progress on the problem of testing these combinations.[26]

If, then, our activities are all increasingly reliant upon software systems, and if software systems cannot be provably free of malicious vulnerabilities, what does that mean for our use of digital technologies?

It means that trust—the probability that a third party's actions will align with our own desires—is both the most central aspect of any transaction *and* also the least quantifiable. Trust is and will become as important as the product it is that we're actually buying.

Some companies, like Apple, have begun to adapt to this trend and are publicly making trust a core component of their brand.[27]

But because trust is not quantifiable it also means that, at least in one sense, it cannot exist. Trust will be a feature that will live equally in the world of branding and imaging as anywhere else. Trust will be illusory. Trust will be nothing more than a marketing campaign.

More interestingly, at least from the legal point of view, our ability to contract will also change because of these same factors. When we purchase a software product, the transaction does not occur at one single point in time—there is no "meeting of the minds" between a software vendor and a user who aims to use that software system, especially as these systems evolve in complexity. The transaction takes place in ways that are in some sense new. Indeed, a group of legal researchers coined a term for precisely this relationship, calling it "a tethered economy," explaining that, "[a]s sellers blend hardware and software—as well as product and service—tethers yoke the consumer to a continuous post-transaction relationship with the seller."[28]

Now add in the fact that trust is the key, but elusive, factor in this continuous relationship, and you have a very real problem. Here's Olav again: "When we ask ourselves if a vendor can be trusted, we have to ask ourselves if we believe the vendor will remain trustworthy for the entire lifetime of the product we're buying."[29] The reverse is also true—we must also ask if we trust the entire past life of the product up until the present. Every moment in time for a software system is one in which a vulnerability can be introduced or created.

And so trust, like privacy, is dead in the sense that traditional frameworks no longer apply to our current reality. We may trust, but we cannot verify.

III. You're Not Who You Think You Are

Lastly is who we think we are—or our identities. Each of our identities is a composite built from our history, our shared experience, our preferences, our self-conceptions and more. The existentialist philosophers of the 20th century famously put a huge amount of effort into explaining how our identities were, ultimately, meaningless constructs. But even to them, identities were fictions that *could* be believed in, in that they were or could be useful. Today, however, this fiction is evaporating in both profound and subtle ways.

Because nearly every act we undertake creates data, our very existence creates a record of our activities that defines us better than any fiction ever could. And yet, because we do not, and cannot, expect to see all this data, the very act of generating data gives organizations the ability to draw insights into our individual lives that we cannot possess on our own.

We don't see, for example, the minute-by-minute records of our cell phones pinging nearby towers, mapping our lives in practically real-time. Cell phone companies and Internet service providers see that data.

We don't see our shopping histories or the intimate insights they yield, like when Target famously predicted almost a decade ago that a teenager was pregnant before her family knew, based on her consumption patterns.[30] Retailers compile all that data.

We don't see the hundreds of thousands of rows in massive databases compiled by Google, Facebook and other technology giants that detail, point by point, nearly every activity we undertake online. Indeed, save for a handful of data scientists working at some of the largest companies in the world, most consumers, even the most technologically literate ones, are unaware of the profoundly personal insights that all this data can yield.

A few years ago, a data consultant wrote an article for the *Guardian* about being granted access to all the information Google had about him. Here's what he found:

> The photos you've taken on your phone, the businesses you've bought from, the products you've bought through Google . . . data from your calendar, your Google hangout sessions, your location history, the music you listen to, the Google books you've purchased, the Google groups you're in, the websites you've created, the phones you've owned, the pages you've shared, [down to] how many steps you walk in a day.[31]

Any single company possessing this level of detail about its users is both beyond comprehension—what other group has been so thoroughly surveilled in human history?—and understandably unsettling, as the author described. At the same time, however, many of us also give all this data willingly to companies like Google when we use them. And we use them more and more. In fact, from Amazon's Alexa to Apple's Siri, we are only entrusting these companies with more of our data, and more personal data at that.

So the question is this: Why is the scope of this information about us so disquieting? And the answer, in my view, is that this level of data collection represents a loss of agency.

It is now entirely possible for these companies, based on all the data we generate, to know more about ourselves than we do. As we come to terms with the fact that other organizations can and will profile us better than we can understand ourselves—making predictions about our preferences and our past and future activities better than we can on our own—we will watch some semblance of our autonomy, of our ability to assert control over our identities, recede. Want to know who you are? Ask Google to tell you. Or Facebook.

Some, like Harvard Business School's Shoshana Zuboff, have taken to calling this new paradigm "surveillance capitalism," where we as consumers are mined for our data, much like natural resources such as mineral deposits are stripped of their raw materials.[32] But I think this framing misses the bigger point. This is not simply about the harvesting and selling of our data; it's about the growing power of our data to tell us meaningful information about ourselves that we are otherwise blind to. It's about the ability of our data, at scale, to predictively and meaningfully draw insights that we cannot intuit on our own.

Just over a decade ago, *Wired's* editor in chief wrote an article entitled, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete."[33] His point was that as techniques like machine learning proliferate, we end up prioritizing correlations over causation—the goal ends up being collecting massive volumes of data, rather than divining causal relationships within that data. And that is, of course, the power and the opportunity of all the data we create. We can learn a tremendous amount about our world without having to understand *why*. As legal scholar Paul Ohm describes it:

> We are embarking on the age of the impossible-to-understand reason, when marketers will know which style of shoe to advertise to us online based on the type of fruit we most often eat for breakfast, or when the police know which group in a public park is most likely to do mischief based on the way they do their hair or how far from one another they walk.[34]

To be clear, this is not all bad news. Although I may have shaped it that way by focusing on the downsides, I don't actually see all these developments as entirely negative. In the healthcare space, for example, many of these trends will certainly lead to new insights, perhaps better diagnoses, more easily accessible medical services and even saved lives.[35] There is a world of undiscovered correlations hidden within all our data that I am genuinely excited about.

But for here, in this essay, I take the widespread collection of our data, and the invasive insights that data can yield, as but one illustration of a deep and profound change in how we understand ourselves and our world.

IV. What Can We Do?

What should we do about all these changes? I will make three specific recommendations in the pages that follow.

i. Slow Adoption Rates, Increase Understanding

My first proposition is this: Slow down.[36]

At their core, our most pressing digital problems all derive from the unprecedented rate at which we've embraced networked devices.[37] For perspective, the Internet is amongst the most rapidly adopted technologies in human history.[38] It took over 100 years after the invention of the telephone before its near universal adoption in the U.S. Similar adoption time frames followed for electricity and automobiles. Yet it's taken only one decade for nearly 4 in every 5 Americans to own a smartphone.[39] Some estimates now predict a global jump from 31 billion connected devices in 2020 to 75 billion networked devices by 2025.[40]

If all our major problems are caused by the frantic pace at which we're adopting networked technology, and if we are generating ever higher volumes of data in the process, then slowing this adoption—and consequently, increasing the chances that we might actually understand the world we're adopting—is our first and best hope. I am not, of course, advocating we adopt a luddite approach towards technology. This is not, as William Buckley described of conservatism, an argument for standing "athwart history, yelling, 'Stop!'"

What might slowing down actually entail? For starters, new laws should be introduced that mandate that any system with an IP address—a prerequisite for connecting to the Internet—either have a finite lifetime or accept updates. The number of devices that can't be updated when flaws are found is long—and continues to grow. Amongst devices that used Google's Android operating system in 2016, for example, a reported 29 percent could not be patched.[41] Once connected, faulty devices can easily be taken over by malicious actors and used to cause harm.[42]

Clearer liability for cybersecurity flaws will also help us to more responsibly adopt digital technologies writ large. Software makers whose code causes harms should be held to account, as is the case with other consumer or industrial products. Currently, penalties for cybersecurity defects tend to arise from failures in reporting *after* breaches or from misrepresentations in a product's terms of service. But neither contributes to safer code. Instead, we must work to clarify what constitutes minimum cybersecurity requirements and standardize liability when these benchmarks are broken. This will help remove vulnerable software from the market.[43]

Clearer liability standards will also, in turn, reduce the pace of adoption—causing vendors to slow down a process that, up until now, prioritizes speed to market above all else. That is, after all, what counts the most in the technology industry—how quickly companies can place a minimum viable product in the marketplace. "Sell first and secure later" might as well be the mantra for most purveyors of modern technology.

There's one final way we should seek to slow down: We must also preserve analog alternatives to digital capabilities. This is a moral, economic, and national security imperative all wrapped into one.[44]

The drive to digitize is, in many senses, a reflection of market pressures that over-incentivize efficiency in the short term. In the long term, however, efficiencies can cause deep vulnerabilities that are easily exploited. As security researcher Bruce Schneier describes it, the "drive for efficiency leads to brittle systems that function properly when everything is normal but break under stress."[45] When these systems do inevitably break, our safety will lie in the systems we preserved.

I can't say at this moment *every* single analog system that must be preserved. But as we embrace new technologies, we cannot—or at least, we should not—blindly replace *all* the systems they're meant to improve. Some redundancy is called for, and we must think closely about what we're losing as new technologies supplants old ones.

ii. We Can't Consent To What We Don't Understand

Almost all current privacy frameworks are based upon user consent in one way or another—that's why users are so frequently forced to agree to complex terms and conditions when browsing online. The idea is that users can, at the point of collection or at the point that their data is generated, meaningfully understand what it is they are trading it in for. Indeed, this is the central idea behind how users engage with companies like Google, Facebook, Twitter and others whose services are the result of something akin to the following barter: "You get our technology, we get your data." The assumption is that both parties can understand the value of what they're trading.

But that core assumption is mistaken. The value of machine learning, as applied against our data, lies in its unpredictability—in human minds not being able to find the patterns themselves. As a result, data's true value can only be recognized at scale and, significantly, at a later point in time.

So what does this mean? It means that we will never fully be able to understand what it is we're consenting to with our data. To reassert control over our data, which I'd contend is *the* central aim of privacy, we need to think about this framework differently—about what it means to trade our data for any object of value. And it means that we must stop thinking about consent as central to protecting our privacy—it is not, and it is becoming less meaningful every day.[46]

Central to a more robust construct are purpose-based restrictions on data—setting what uses certain data can and cannot be applied to outside of any particular transaction. The EU does a very good job with this in the General Data Protection Regulation, the main, quite stringent data privacy legislation in Europe which began to be enforced in 2018. The GDPR sets forth six—and only six—legal bases for processing data.[47] These bases are something that the government authorizes; they constitute a positive right that derives from legislation.[48]

What I'm suggesting is that the U.S. and other jurisdictions learn from this approach and make these types of overarching restrictions—which can and should be divorced from the user's immediate understanding—central to protecting our privacy.[49] If an organization doesn't know what the data will be used for, and cannot state it clearly at the point of collection and each subsequent stage of use, that organization should not be able to collect that data to begin with.

iii. Data, Data Everywhere and Not a Drop to Drink

Last is our need for more data, despite the fact that we as individuals cannot stop generating it. If we cannot understand our digital environment, which is one of my central theses in this essay, we must do better to situate ourselves within it. We must share our collective data better across organizations, and we must make that data more accessible, in as close to real time as possible.

There's often a default assumption in the world of data science that more data leads to better insights, which is not exactly what I'm advocating. What I'm arguing is that, in the present case, no data leads to no insights, which is where we all too frequently find ourselves.

We must, as a result, do a better job of facilitating access to data about our environment across organizations. There have been a handful of attempts to facilitate data sharing in the area of cyber threat indicators, for example, but these attempts have not fared well, largely because they have failed to provide concrete incentives that would enable the type of widespread data sharing that they sought to achieve.[50] The same goes with healthcare data and data interoperability more broadly, which has long been the subject of stalled research efforts and boastful PR campaigns, but with few tangible results to show.[51]

One of the futures I'm worried about—and indeed, a future that we are already entering—is one in which only the largest companies, with the largest amounts of data, can take advantage of the most powerful technologies because they're among the few organizations that can understand the digital world clearly. This is a future in which the Amazons, the Googles, and the Facebooks preclude almost every other organization, or even individual, from operating meaningfully in cyberspace. It's a world where only a few companies can utilize all the benefits of techniques like machine learning because only they can access the requisite data. And so while data must be better protected, it also must be easier to access across organizations.

Is this a realistic recommendation? I believe so, thanks largely to a handful of technical solutions that enable faster, more secure, and privacy-preserving usage of data, commonly referred to as "privacy enhancing technologies," or PETs. These technologies include techniques like differential privacy, federated learning, synthetic data, and a few others.[52] By incentivizing their use from a liability standpoint and by making them less exotic and easier to use for front-line developers, I believe that PETs will be key to a future where data is put to use more intelligently and securely.

Practically speaking, the use of PETs also aligns with the type of use-based restrictions on data I argue for above, in that they can limit the scope and scale of the data being shared while also encouraging its use. Data collection therefore can (and should) be highly restricted at the moment it is gathered, but it should also allow for secondary uses if and when that data is sufficiently de-identified. Indeed, the idea of loosening restrictions on anonymized data is already embedded into many existing data protection frameworks. The Health Insurance Portability and Accountability Act in the U.S., to cite just one example, allows for secondary use of medical data if that data has been sufficiently anonymized in exactly this way.[53]

What this means in practice is that policymakers should begin to formally encourage the use of PETs by reducing legal liability when these techniques are properly implemented, which will in turn result in lower insurance premiums for organizations worried about safeguarding their data, further incentivizing their use. There is more research to be done in making PETs more readily available and understandable to end users, but I believe they hold the key to increasing protections on data while also making that data easier to access.

V. Easy Choices...That We Refuse to Make

We are, in short, undergoing profound shifts in our lives thanks to the rate at which we're adopting networked technology—shifts related to our privacy, our security, our transactional relationships, even our identities.

But these same shifts—the way they occur, the pace at which they occur—are not simply *fated* to happen. We can and we must be proactive about these changes. They are occurring because we are making a collective choice, every day, to adopt these technologies.

And that means there is nothing inevitable in the story I have outlined in these pages, about the specific technologies we use or even the rate at which we use them.

We can slow down.

We can place intelligent, carefully written laws on the books to help us make sense of our digital environment and to help us assert control over our data.

We can preserve what we risk losing.

The question is: what's stopping us?[54]

Notes & Acknowledgements

I wear many hats—as the head of a law firm, an officer in a technology company, and here at the ISP as a visiting fellow. As a result, I believe it's important to disclose how those roles may have affected my thinking, including some of the arguments I make in this paper.

To start with is the fact that this whitepaper series is sponsored by Immuta, a software company focused on data governance which I have helped to build and in which I hold financial interests (I am chief legal officer there). Because I'm an editor for this whitepaper series and also the author of its first whitepaper, I have relied on Nabiha and Nik, the other editors of this series, to ensure that all editorial decisions are made impartially and don't inappropriately advance any one company or financial interest.

Regarding some of the arguments I make above, it's also worth noting that Immuta has, over the last few years, invested significant resources in building technologies like differential privacy, which I enthusiastically cite in this paper. I am a believer in the potential of privacy enhancing technologies like differential privacy in no small part because of my exposure to them through my work at Immuta.

In my work at the law firm, I also have clients who are at varying stages in the process of adopting some of the technologies described above, which has exposed me to many of the risks I also cite first-hand. While few would accuse my views above as being optimistic, to the extent that I am a promoter of new digital technologies, I have many of these clients and engagements to thank.

On the topic of thanks, I would also like to thank a host of friends and colleagues for their feedback and support in making this whitepaper and the larger series possible. This includes Dennis Hirsch, whose invitation to speak at the Ohio State's Moritz School of Law Program on Data and Governance in 2019 gave rise to many of the ideas expressed in this whitepaper. Dan Geer, who I continue to believe is one of the leading thinkers in the world of information security, helped me to develop many of these ideas in the course of our writing and ongoing discussions, and in particular in co-authoring "Flat Light: Data Protection for the Disoriented, from Policy to Practice" in 2018. I'd also like to express my gratitude to Jack Balkin, Nikolas Guggenberger, Nabiha Syed, Ben Lorica and Matthew Carroll for all their help and guidance, and to Spurthi Jonnalagadda for her editing skills. And, of course, to Claudia Naím-Burt, who makes everything possible.

References

[1] The text of this essay is loosely adapted from a lecture, “Flat Light In A Digital World,” delivered at The Ohio State Moritz College of Law Data Points Lecture Series February 6, 2019.

[2] “There are an estimated 30 million surveillance cameras now deployed in the United States shooting 4 billion hours of footage a week.” James Vlahos, *Surveillance Society: New High-Tech Cameras Are Watching You*, POPULAR MECHANICS (Oct. 1, 2009), https://www.popularmechanics.com/military/a2398/4236865/. “In 2015, the global video surveillance industry was valued at about $20 billion, and is expected to grow to $63.2 billion by 2022.” Jordan G. Teicher, *Gazing Back at the Surveillance Cameras That Watch Us*, N.Y. TIMES (Aug. 13, 2018), https://www.nytimes.com/2018/08/13/lens/surveillance-camera-photography.html.

[3] Indeed, the digital world has become so ubiquitous that couples in the United States are now more likely to meet online than in the physical world. *See* Alex Shashkevich, *Meeting Online Has Become the Most Popular Way U.S. Couples Connect*, Stanford Sociologist Finds, STAN. NEWS (Aug. 21, 2019), https://news.stanford.edu/2019/08/21/online-dating-popular-way-u-s-couples-meet/.

[4] “[E]very new pacemaker implanted in the United States is cloud-connected.” Neta Alexander, *My Pacemaker is Tracking Me From Inside My Body*, ATLANTIC (Jan. 27, 218), https://www.theatlantic.com/technology/archive/2018/01/my-pacemaker-is-tracking-me-from-inside-my-body/551681/.

[5] I. Wagner, *Connected Cars Worldwide - Statistics & Facts*, STATISTA (Sept. 15, 2020), https://www.statista.com/topics/1918/connected-cars/.

[6] The Wealth-X Billionaire Census 2019, WEALTH-X (May 9, 2019), https://www.wealthx.com/report/the-wealth-x-billionaire-census-2019/.

[7] *Number of Records Exposed in 2019 Hits 15.1 Billion*, RISKBASED SECURITY (Feb. 10, 2020), https://www.riskbasedsecurity.com/2020/02/10/number-of-records-exposed-in-2019-hits-15-1-billion/.

[8] Andrew Burt & Daniel E. Geer, *Flat Light*, HOOVER INST. (Nov. 20, 2018), https://www.hoover.org/research/flat-light;*see also* Daniel E. Geer, *A Rubicon*, HOOVER INST. (Feb. 2, 2018), https://www.hoover.org/research/rubicon.

[9] United States. Congress. H. Comm. on Sci., Space, and Tech. Subcomm. on Investigations and Oversight, *Bugs in the Program: Problems in Federal Government Computer Software Development and Regulation: Staff Study*, Volume 4, U.S. Government Printing Office, 1989.

[10] Richard Danzig, *Surviving on a Diet of Poisoned Fruit: Reducing the National Security Risks of America’s Cyber Dependencies*, CNAS (July 21, 2014), https://www.cnas.org/publications/reports/surviving-on-a-diet-of-poisoned-fruit-reducing-the-national-security-risks-of-americas-cyber-dependencies.

[11] *Worldwide Threat Assessment of the U.S. Intelligence Community: Hearing Before S. Select Comm. on Intelligence*, 113th Cong. (2013) (statement of James R. Clapper, Director of National Intelligence); *Worldwide Threat Assessment of the U.S. Intelligence Community: Hearing Before S. Select Comm. on Intelligence*, 113th Cong. (2014) (statement of James R. Clapper, Director of National Intelligence); *Worldwide Threat Assessment of the U.S. Intelligence Community: Hearing Before S. Armed Services Comm.*, 114th (2015) (statement of James R. Clapper, Director of National Intelligence); *Worldwide Threat Assessment of the U.S. Intelligence Community: Hearing Before S. Armed Services Comm.*, 114th (2016) (statement of James R. Clapper, Director of National Intelligence); *Worldwide Threat Assessment of the U.S. Intelligence Community: Hearing Before S. Select Comm. on Intelligence*, 115th Cong. (2017) (statement of Daniel R. Coats, Director of National Intelligence); *Worldwide Threat Assessment of the U.S. Intelligence Community: Hearing Before S. Select Comm. on Intelligence*, 115th Cong. (2018) (statement of Daniel R. Coats, Director of National Intelligence); *Worldwide Threat Assessment of the U.S. Intelligence Community: Hearing Before S. Select Comm. on Intelligence*, 115th Cong. (2019) (statement of Daniel R. Coats, Director of National Intelligence),

[12] Samuel Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, (1890).

[13] James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151 (2003).

[14] Griswold v. Connecticut, 381 U.S. 479 (1965).

[15] Dake Kang, *Chinese 'Gait Recognition' Tech IDs People by how They Walk*, AP NEWS (Nov. 6, 2018), https://apnews.com/bf75dd1c26c947b7826d270a16e2658a. Note that gait analysis is becoming more prominently used across the world, not simply to identify individuals but also for sentiment analysis as well. *See* Matt Simon, *This Robot Can Guess How You're Feeling by the Way You Walk*, WIRED (May 18, 2020) https://www.wired.com/story/proxemo-robot-guesses-emotion-from-walking/.

[16] Arvind Narayanan, et al., *On the Feasibility of Internet-Scale Author Identification*, 2012 IEEE, https://ieeexplore.ieee.org/document/6234420; *see also* Aylin Caliskan, et al., *When Coding Style Survives Compilation: De-anonymizing Programmers from Executable Binaries,* Network and Distributed Sys. Security Symp. (Dec. 18, 2017), https://arxiv.org/abs/1512.08546.

[17] A conjunction of the phrase "no mobile phone phobia."

[18] Gonzalez, M. Hidalgo, C. & Barabasi A., *Nature*, 2008.

[19] Fengli Xu, *Trajectory Recovery From Ash: User Privacy is Not Preserved in Aggregated Mobility Data*, WWW '17: Proceedings of the 26th International Conference on World Wide Web 1241 (2017), https://arxiv.org/pdf/1702.06270.pdf; Jennifer Valentine-DeVries et al, *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. TIMES (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.

[20] Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. R. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/. For additional ways that we can be identified, see bioacoustic signatures, Michelle Hampson, *The Bioacoustic Signatures of Our Bodies Can Reveal Our Identities*, SPECTRUM (Nov. 4, 2019), https://spectrum.ieee.org/the-human-os/telecom/security/the-bioacoustic-signatures-of-our-bodies-can-reveal-our-identities, and browsing histories, Sarah Bird, Ilana Segall & Martin Lopatka, *Replication: Why We Still Can't Browse in Peace:*

On the Uniqueness and Reidentifiability of Web Browsing Histories, 16 SYMP. ON USABLE PRIVACY AND SECURITY 489, (Aug. 10–11, 2020).

[21] We generate, for example, an estimated 2.5 quintillion bytes of data every day. To people who are unfamiliar with that metric, one quintillion is a thousand raised to the power of six, or 10^{18}. Bernard Marr, *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*, FORBES (May 21, 2018), https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#64123e1e60ba.

[22] With "trust" being defined, loosely, as the probability that a third party's actions would align with the first party's own expectations and desires.

[23] Jordan Robertson & Michael Riley, *The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies*, BLOOMBERG BUSINESSWEEK (Oct. 4, 2018), https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies.

[24] Russell Hotten, *Volkswagen: The Scandal Explained*, BBC NEWS (Dec. 10, 2015), https://www.bbc.com/news/business-34324772.

[25] Olav Lysne, THE HUAWEI AND SNOWDEN QUESTIONS 34 (Aslak Tveito et al. eds, 2018), at 36.

[26] *Id.* at 72–73. Specifically, the discussion relates to dynamic malware detection.

[27] Throughout 2019, for example, the company ran ads that stated, "What happens on your iPhone, stays on your iPhone." Chance Miller, *Ahead of CES, Apple Touts 'What Happens on Your iPhone, Stays on Your iPhone' With Privacy Billboard in Las Vegas*, 9TO5MAC (Jan. 5, 2019) https://9to5mac.com/2019/01/05/apple-privacy-billboard-vegas-ces/. Apple continues to build on this brand association, announcing at the end of 2020 that it would require developers to publish so-called "privacy nutrition labels" associated with apps in their app store. Ian Carlos Campbell, *Apple Will Require Apps to Add Privacy 'Nutrition Labels' Starting December 8th*, VERGE (Nov. 5, 2020) https://www.theverge.com/2020/11/5/21551926/apple-privacy-developers-nutrition-labels-app-store-ios-14.

[28] Chris Hoofnagle, Aniket Kesari & Aaron Perzonawski, *The Tethered Economy*, 87(4) GEO. WASH. L. REV. 783, (2019).

[29] Lysne, *supra* note 25, at 6.

[30] Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp.

[31] Dylan Curran, *Are You Ready? Here is All the Data Facebook and Google Have on You*, GUARDIAN (Mar. 30, 2018), https://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy.

[32] Shoshana Zuboff, THE AGE OF SURVEILLANCE CAPITALISM (2019).

[33] Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (June 23, 2000), https://www.wired.com/2008/06/pb-theory/.

[34] Paul Ohm, *The Fourth Amendment in a World Without Privacy*, 81 MISS. L.J. 1318, (2012).

[35] Lily Peng & Varun Gulshan, *Deep Learning for Detection of Diabetic Eye Disease*, GOOGLE AI BLOG (Nov. 29, 2016) https://ai.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html.

[36] As also described with Dan Geer in Andrew Burt & Dan Geer, *Improving Cybersecurity Means Taking More Care With What We Digitize*, HARV. BUS. REV. (Feb. 1, 2019), https://hbr.org/2019/02/improving-cybersecurity-means-taking-more-care-with-what-we-digitize.

[37] *Id.*

[38] Derek Thompson, *The 100-Year March of Technology in 1 Graph*, ATLANTIC (Apr. 7, 2012), https://www.theatlantic.com/technology/archive/2012/04/the-100-year-march-of-technology-in-1-graph/255573/.

[39] *Mobile Fact Sheet*, PEW RESEARCH CENTER (June 12, 2019), http://www.pewInternet.org/fact-sheet/mobile/.

[40] Lauren Horwitz, *The Future of IoT Miniguide: The Burgeoning IoT Market Continues*, CISCO (July 19, 2019), https://www.cisco.com/c/en/us/solutions/internet-of-things/future-of-iot.html.

[41] Lisa Vaas, *29% of Android Devices Can't be Patched by Google*, NAKED SEC. (Apr. 21, 2016), https://nakedsecurity.sophos.com/2016/04/21/29-of-android-devices-cant-be-patched-by-google/.

[42] Garrett M. Graff, *How a Dorm Room Minecraft Scam Brought Down the Internet*, WIRED (Dec. 13, 2017), https://www.wired.com/story/mirai-botnet-minecraft-scam-brought-down-the-Internet/.

[43] Much like California did in 2018, Adi Robertson, *California Just Became the First State With an Internet of Things Cybersecurity Law*, VERGE (Sept. 28, 2018), https://www.theverge.com/2018/9/28/17874768/california-iot-smart-device-cybersecurity-bill-sb-327-signed-law.

[44] Dan Geer expresses this same point succinctly: "[L]uckily, the non-digitalized space is still reasonably pervasive and salvageable, but not for long. To let it slip away is to write off from society's ledger the centuries of investment that have brought us to where we are." Daniel E. Geer, *A Rubicon*, HOOVER INST. (Feb. 2, 2018), https://www.hoover.org/research/rubicon.

[45] *The Security Value of Inefficiency*, SCHNEIER ON SECURITY (July 2, 2020), https://www.schneier.com/blog/archives/2020/07/the_security_va.html.

[46] Poll after poll, survey after survey, for example, indicates that American consumers understand the depth of their loss of control and their inability to comprehend all the ways their data is put to use. Brooke Auxier, *How Americans See Digital Privacy Issues Amid the COVID-19 Outbreak*, PEW RESEARCH CTR. (May 4, 2020) https://www.pewresearch.org/fact-tank/2020/05/04/how-americans-see-digital-privacy-issues-amid-the-covid-19-outbreak/; *see also* Elettra Bietti, *Consent as a Free Pass: Platform Power and the Limits of the Informational Turn*, 40 PACE L. REV. 307, (2020).

[47] *See* Art. 6, GDPR.

[48] Other jurisdictions, like Brazil, have imitated this approach.

[49] There are signs that this approach is already catching on in the U.S., with California passing the California Privacy Rights Act via referendum this year. The CPRA, which is clearly modelled off of the GDPR, would make purpose limitations a central mechanism for limiting the use of consumer data once collected.

[50] *See, e.g.*, Cyber Information Sharing Act, S.B. 754, 114 Cong. (2015).

[51] *See, e.g.*, *HHS Finalizes Historic Rules to Provide Patients More Control of Their Health Data*, HHS (Mar. 9, 2020) https://www.hhs.gov/about/news/2020/03/09/hhs-finalizes-historic-rules-to-provide-patients-more-control-of-their-health-data.html. Some of the biggest tech giants even introduced an open-source effort called "The Data Transfer Project," designed to make it easy to move data between companies. There appears to have been no work on the project since 2019. *See* https://github.com/google/data-transfer-project/projects; Craig Shank, *Microsoft, Facebook, Google and Twitter Introduce the Data Transfer Project: An Open Source Initiative for Consumer Data Portability*, MICROSOFT BLOG (July 20, 2018), https://blogs.microsoft.com/eupolicy/2018/07/20/microsoft-facebook-google-and-twitter-introduce-the-data-transfer-project-an-open-source-initiative-for-consumer-data-portability/.

[52] The UN Handbook on Privacy-Preserving Computation Techniques outlines PETs well: Mark Craddock, *UN Handbook on Privacy-Preserving Computation Techniques*, http://publications.officialstatistics.org/handbooks/privacy-preserving-techniques-handbook/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf.

[53] *See, e.g.*, the HIPAA Privacy Rule, which sets forth two methods of anonymization.

[54] Technological fatalism. Short sightedness. Bias in favor of new things. Market forces. Distraction. Nothing, in short, that *need* stop us. Just forces that *can*.

Published by the Information Society Project at Yale Law School, Baker Hall,
100 Tower Parkway Room 412, New Haven, CT 06520 United States of America.

The Digital Future Whitepaper Series is made possible thanks to the support of Immuta.

The ideas and opinions expressed in this whitepaper are those of the authors and do not reflect the views of Yale Law School or any other organizations including sponsors.

About the Author

Andrew Burt is managing partner at bnh.ai, a boutique law firm focused on AI and analytics, and chief legal officer at Immuta. He is also a visiting fellow at Yale Law School's Information Society Project.

Previously, Andrew was Special Advisor for Policy to the head of the FBI Cyber Division, where he served as lead author on the FBI's after action report on the 2014 Sony data breach.

A frequent speaker and writer, Andrew has published articles on law and technology for the New York Times, the Financial Times and Harvard Business Review, where he is a regular contributor. He holds a JD from Yale Law School.

Digital Future Whitepaper Series

The Digital Future Whitepaper Series, launched in 2020, is a venue for leading global thinkers to question the impact of digital technologies on law and society. The series aims to provide academics, researchers and practitioners a forum to describe novel challenges of data and regulation, to confront core assumptions about law and technology, and to propose new ways to align legal and ethical frameworks to the problems of the digital world.

The Digital Future Whitepaper Series is edited by ISP fellows Andrew Burt, Nikolas Guggenberger, and Nabiha Syed. Spurthi Jonnalagadda (Yale Law School '22) served as the research assistant for this whitepaper.

Information Society Project

The Information Society Project (ISP) is an intellectual center at Yale Law School, founded in 1997 by Professor Jack Balkin. Over the past twenty years, the ISP has grown from a handful of people gathering to discuss Internet governance into an international community working to illuminate the complex relationships between law, technology, and society.

Information Society Project
Yale Law School

127 Wall Street
New Haven, CT 06511

203.432.4992
www.law.yale.edu/isp