Surveillance Capitalism and Government Access to Data

The pandemic makes it particularly easy to appreciate how far every aspect of human life is datafied. The alarms that wake us, the music we play, sometimes even the coffee maker and the fridge are connected to the web and record data about us. As we sleep, health monitors gather data about us, and as we move through our day, our fitness trackers and other medical trackers monitor us. The contrast between our lives and the lives of our parents when they were our age could not be more striking. Few things remain secret. Our commute to work, our weekend getaways, our favorite bars and book shops, the TV shows we binge-watch, the annotations on our e-books – this is all information collected by the companies Shoshanna Zuboff has called ‘surveillance capitalists’.

The datafication is almost comprehensive and is more intrusive than most people realize. How long we communicate with people, what we read and how long it holds our attention are rendered into datasets. Even our moods are gradually read and mapped from our behavior and the words we use. Although the internet has enabled easy and inexpensive communication, it has also meant that surveillance capitalist companies have access to almost all our communication. All our conversations have moved online. A man in Lagos texting his mother who lives across the street, and a woman in Italy calling her partner in Brazil might use the same company’s services. Those of us who have lived in cities all our lives might find ourselves turning to Google for directions, mapping every step we take to find a long-elusive bookshop or restaurant that is now datafied and available to us. As we do, we make our search and our trip visible to the company.

Thinking through the consequences of this transformative change in our society triggers a number of questions. I discuss three here. The first is about the recording of data, the second is about how it can be used, and the third is about the debates about whether governments have any rights or claims over data collected from their citizens.

Data is everywhere

We are living in an increasingly ‘datafied’ society, orienting around big datasets and innovation in the processing and analysis of this data. Big data has permeated most sectors and most people’s daily lives. This has visible and invisible consequences that we are coming to accept as normal. Our search and social media algorithms use data and predictions to decide what to display. We have credit scoring systems, and systems to flag who is high risk during bail and sentencing here. We even use algorithms and big data to hire people for jobs. Cities are using big data for policing, traffic management and welfare.

It is common for people to speak of data as if it is a natural resource, to be ‘extracted’ and used in ways that affect the political economy. However, extraction implies that data is everywhere in its raw form. The truth about data is actually that it is ‘rendered’ rather than extracted. This implies that a choice is made in how data is recorded and the dataset is the creation of human beings, embedded with our values. The choices made while recording data affect what we learn from the data. Take for example, the recording of deaths. If the data fields used require that every deceased person is listed as either male or female, the dataset will fail to identify people who do not identify as either. Such an exclusion might mean that if people of non-binary genders have a higher mortality rate, that trend will remain invisible.

Avenues for error and misdirection are plentiful. We can see this in the way data is recorded in the context of the pandemic resulting from the novel corona virus. When hospitals record COVID-
19 deaths of people with an underlying condition as being caused by the underlying condition, rather than by COVID-19, the records of pandemic-related deaths are skewed. Data-driven decisions necessitate the availability of reasonably accurate data. Where there are avenues for error, there are also avenues for manipulation. Human beings can choose not to record data, or to record it incorrectly. Certain kinds of crimes committed against persecuted minorities for example are under-reported or misreported so that they do not appear to be hate crimes.

Large datasets are now collected and held by private companies, where previously they were usually held by government in a period in which data collection necessitated different resources. Sometimes this privately held data is used to deliver public services, as is evident from the many private companies that are playing a role in tracking the pandemic.

The companies that hold the big datasets make architectural choices about how the data is stored and accessed. How this data can be accessed and used is affected by design choices made in the context of the dataset. For example, end-to-end encryption puts all communication data out of the reach of people who are not a part of the encrypted conversation. Alternatively, data stored in aggregate form – the number of people searching for ‘COVID-19 symptoms) per day in a city for example – may not always reveal the details of each unit, making it difficult to identify who these people are without additional data.

**Data is used in more ways than we can imagine**

Data and data-driven solutions are used by different actors for different purposes. To comprehend the interconnectedness of these systems and the many ways in which they are used, it may help to think of them as cyber-physical systems, consisting of “a collection of computing devices communicating with one another and interacting with the physical world via sensors and actuators in a feedback loop.”

Laura De Nardis points out that framing it this way helps us understand the variations in which these systems can exist, including:

1. the digitization of everyday objects, including consumer IoT devices;
2. the Internet of self, which applies to devices in close proximity to the body, such as wearables and medical devices
3. the industrial Internet of things which covers connected objects embedded in industrial sectors, such as agriculture or manufacturing
4. emergent cyber-physical systems, which refers to objects that are ‘born digital’ such as 3D printing and robotics.

These cyber-physical systems process, analyze, and use data in a range of ways depending on their goals. Consumer systems may be directed at developing better products, and predictive marketing. Amazon and Facebook’s targeted advertising are familiar to everyone. With the use of RFID chips and the sale of data sets, information collected in one context might be used for different purposes in a different context. Sometimes this leads to high profile crises like the Cambridge Analytica scandal. There are worries that user-profiling through big datasets collected across a range of our activities may affect hiring and eligibility for loans or credit cards.

States are also beginning to use data driven systems to perform their core functions. Data-driven algorithmic sentencing and policing systems are among examples of this. While states also hold large public datasets, it is common for informational capitalist companies to aggressively collect data and to retain exclusive use of it. These companies possess the processing power to combine
machine learning and these datasets to create new products, strategies and conclusions. They are increasingly participating in the delivery of public services, where governments want to use cyber-physical systems but cannot muster the computational power or data for them. This is evident in government responses to the novel corona virus, where many states are partnering with technology companies to monitor movement in public spaces, to monitor quarantine and for contact-tracing and other such activities that have to take place at scale.

The privately held data raises many troubling questions. On one hand are the many ways in which companies end up harm individuals either directly, or through a chain of events that may lead to a large-scale impact on democracy like electoral misinformation. On the other hand, problems arise when governments wish to access the data and are not able to do so in the manner of their choosing. These problems can get complex when either law or technology makes it difficult for companies to share data with law enforcement agencies.

Governments and Access to Data

Who can forget the headline-grabbing standoffs between surveillance capitalist companies and law enforcement? Underlying the standoffs are typically either technological architecture or laws that make it impossible for the companies to share data with law enforcement agencies on request. A classic example of architecture that makes data sharing impossible might be the Apple versus FBI case over the San Bernardino attack which was about whether Apple should make it possible to access information on iPhones without the password. The ways in which laws and architecture can both act as barriers to data sharing can be seen in the trouble WhatsApp has run into in Brazil for refusing to share encrypted messages relating to a drug-trafficking investigation. This is a result of technology or architecture—end-to-end encryption—that makes access to this data difficult, as well as US law that does not permit WhatsApp to share data stored in the United States of America. The second order effects of this US law have led to stand-offs between American surveillance capitalist companies and law enforcement around the world.

These cases are about targeted surveillance, and they have no bearing on law enforcement’s access to meta data like contacts: the fact that two parties communicated or how long a call between them lasted. It is also important to note that even end-to-end encryption can be subverted. State actors are able to purchase services from cyber-intelligence companies like the NSO group, that manage to secretly embed malware on targets’ phones and access the encrypted communication by accessing the devices at the ends of the communication. The use of these services is alleged to have resulted in the murder of journalist Jamal Khashoggi as it later emerged that his friend’s phone was illegally accessed, offering access to details that Khashoggi had shared with him through encrypted messages. Although governments are publicly engaged in what looks like a battle for backdoors, access to encrypted communication potentially affects senior state functionaries like diplomats who may be engaged in sensitive communication that other states wish to access.

However the problem is real, particularly in the context of communication data that a law enforcement agency should legitimately be able to access. Where previously it was possible to access private communication following executive or judicial orders, now much communication is on US-based communication platforms. Even if the local laws of the country in question compel the sharing of data, the communication platforms cannot share it. It does not follow that the data is inaccessible. The foreign state has to use the time-consuming MLAT system to gain access to the data. Unsurprisingly, this has led to discontent. It is especially irksome to law enforcement to know that two people in the same neighborhood in their country can communicate with each other on an American information platform such that the conversation will take months, possibly
longer to access. In response to these concerns the USA has developed a way to work around its older law with the CLOUD Act, which sets up a new system for data sharing.

Although this system is still unfolding, with only one data-sharing agreement signed with the United Kingdom so far, it is worth watching closely. As these agreements are signed, investigative agencies in non-US countries will suddenly gain access to communication. These agencies engage not just in targeted surveillance, but in mass surveillance. This sudden change, giving them access to an unprecedented amount of personal communication data is almost certain to have an impact, although it is hard to say what form that will take.

Conclusion

It will be interesting to see whether this change will result in more public-private co-operation over predictive information, such as who is most susceptible to secessionist propaganda, or to the lures of a drug trafficking ring. Over time, it is likely the states will deploy cyber-physical systems and their increasing capacities for surveillance. China has already done so in a manner that is remarkably effective. The implications of the surveillance net widening and closing over people across the world will be enormous and are worth thinking through in detail.

Recommended reading
Laura De Nardis, *The Internet in Everything* (2020)