KHAZANAH
RESEARCH
INSTITUTE

Views 11/19

31 July 2019

# Personal Data Privacy and Surveillance Capitalism

Khazanah Research Institute

Views 11/19 31 July 2019

Personal Data Privacy and Surveillance Capitalism.
This view was prepared by Rachel Gong, a researcher, and Hui San Chiam, a research intern, from the Khazanah Research Institute (KRI). The authors are grateful for the valuable comments from Nazihah Muhamad Noor and John Loh. The views and opinions expressed in this article are those of the authors and may not reflect the official views of KRI.

Views

## Personal Data Privacy and Surveillance Capitalism

By and large, Malaysia has embraced digital connectivity. The Malaysian Communication and Multimedia Commission (MCMC) reports that as at the end of 2018 there were 1.3 mobile phones per citizen[1], four out of five Malaysian internet users have a social networking account such as Facebook, and one out of two internet users have bought something online[2]. As the country looks towards strengthening its digital economy with big data analysis and machine learning, there is a need for policymakers to give serious consideration to not only the economic but also the social implications of digitalisation. One of the many digital policy issues that merits attention is personal data privacy.

It is likely that even the most privacy-minded people who do not wish to be tracked by digital platforms will have come to the conclusion that there is no real option to opt out of being surveilled while remaining active members of society. Some may seek solace in the rationalisation that as long as they are not doing anything illegal, some loss of privacy is simply the trade-off the digitalised world makes for convenience and security.

We assert that this rationalisation is beside the point. The point is that, left unchecked, unregulated data collection by both governments and corporations can have adverse effects for individuals, vulnerable segments of society and society as a whole. To support our case, we lay out some examples of data misuse and abuse and call for policymakers and practitioners to take into account privacy considerations and potential unintended consequences when developing digital economy policies and implementing digital technologies.

With increased computing power, protecting privacy is no longer just a matter of limiting data sharing. In today's networked society, advanced data science makes it possible to profile a person by analysing their digital footprint, which includes seemingly public and innocuous data such as their search history, online interactions, or simply their location when they access the internet. We contend that striking the right balance between individual rights to privacy and collective wellbeing is an important policy consideration.

[1] (Malaysian Communications and Multimedia Commission 2019)
[2] (Malaysian Communications and Multimedia Commission 2018)

## Digital Platforms, Surveillance Capitalism and Data Privacy

Digital platforms, such as Facebook and Grab, establish a commercial network of interdependent end users and providers to enable the provision of products, which include both goods and services[3]. These platforms create value by providing online communication channels and marketplaces that allow for rapid and inexpensive scaling. Multisided platforms such as Amazon allow producers and consumers to interact directly to find and obtain a wider range of products at lower marginal costs[4]. Amazon expanded its range of products from retail goods to publishing to payments to online storage to the point that it now provides the backbone infrastructure for many other digital platforms[5].

Expansion of this magnitude depends on a strong understanding of what users want and what they are willing to give in order to get it. How much a user is willing to pay for a product is just one factor to be considered; another factor is how much personal data a user is willing to give up for convenience and personalisation. When we open the Google Assistant app, and it welcomes us with a greeting appropriate for the time of day, a localised weather and traffic report, and a reminder to buy a birthday present for our mothers, we take for granted the degree to which our lives are being monitored.

Constant tracking by digital platforms like Google and Facebook gives them access to more user data than ever, both in terms of numbers of users and the types of data they collect on each user. As a start, Facebook collects all the data of users who are logged in, including profile information, shared photos, interactions and activities, likes and follows, network data, and location data. How often users log in, how long we spend on the platform and what items we click on are all tracked[6]. Facebook also uses cookies, little pieces of data, to track what its users do online even when they are not logged in to Facebook[7]. There is justified concern among researchers that not only will user data that are knowingly shared be used without informed consent but also that user data that are unknowingly shared will be used without informed consent in ways unintended by the data provider[8].

In her book "*The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*", Shoshana Zuboff[9] explains how digital platforms engage in surveillance capitalism. Platforms like Google collect and analyse users' "behavioural surplus" for purposes besides service improvement, especially for predicting and modifying users' behaviour in order to generate more behavioural surplus. Behavioural surplus, also known as data exhaust, is secondary data generated during a transaction. For example, an internet user searches for "flights from Kuala Lumpur to Bali". The search term constitutes primary behavioural data; behavioural surplus includes data on

[3] (Rossotto et al. 2018)
[4] (Asadullah, Faik, and Kankanhalli 2018)
[5] (Khan 2017)
[6] (Facebook 2018)
[7] (Pierson and Heyman 2011)
[8] (Whitley 2009)
[9] (Zuboff 2019)

what time of year (or day) the user conducted the search, how often the search was repeated, on what sort of device the search was conducted, and so on. Big data analysis of the primary data can be useful in identifying popular travel destinations, but analysis of the secondary data can reveal a lot more about the people interested in travel.

Because users are generally unaware of the amount of behavioural surplus they produce and because analysis of such data had previously been extremely costly and time-consuming, surveillance capitalism has gone unchecked until recently. Digital platforms were able to experiment with big data analysis and expand their collection of behavioural surplus without any regulations. But as governments, corporations and society are becoming more aware of how powerful and pervasive big data analytics can be, the real issue is not simply digital platforms hoarding user data. The real question is how we will choose to employ these technological tools. How will we manage the use of data to help and not harm?

There is evidence that, without proper foresight and management, the misuse of personal data can have adverse effects on multiple levels. The first level is at an individual level when users' personal data reveal private information that they had not intended to share. The second level affects vulnerable segments of society whose shared characteristics make them susceptible to bias and discrimination. The third level of societal adverse effects occurs when aggregated personal data are used to modify collective behaviour without users' knowledge. The following sections explore some examples of these adverse effects.

## Individual Privacy Violations

Article 12 of the 1948 Universal Declaration of Human Rights[10] reads: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks." Warren and Brandeis[11] long ago advocated the "right to be let alone", noting that everyone needs private spaces—whether their bodies, their family lives and social relationships, their property, or information about them—to remain free from external interference[12]. Essential to what it means to develop a rounded sense of self is to have some sort of space where one can be free from the fear of judgement[13]. According to privacy advocate Greenwald, without privacy, the fear of being watched creates a prison in the mind and causes self-censorship as it subdues free expression and free thought[14].

Location data may seem innocuous, but in fact this information reveals a lot about a person. Regular travel patterns can be used to infer a person's home address, place of work, leisure preferences, or private vices. Ride-sharing apps employ geolocation surveillance intended to protect drivers and passengers, but, in the case of Uber, location tracking has led to the exposure of extramarital affairs[15] and patterns of one-night stands[16]. In 2016, Uber faced a lawsuit when a whistle-blower exposed how employees used location data to track politicians, celebrities and even exes[17]. All these constitute violations of individual privacy rights.

Although Uber has ceased operations in Malaysia, local counterparts such as Grab, MyCar, and EzCab operate in much the same way. Per Grab's privacy policies, the ride-hailing app collects similar data, such as precise and approximate location information as well as trip updates[18][19]. Therefore, it is reasonable to assume that many ride-hailing apps have similar capabilities for tracking and analysing its users' data. Without regulations, we are susceptible to violations of individual privacy rights, not only by foreign digital service providers, but also local companies.

[10] (United Nations 2015)
[11] (Warren and Brandeis 1890)
[12] (Buttarelli 2017)
[13] (Introna 1997)
[14] (Greenwald 2016)
[15] (Trigg 2017)
[16] (Uber 2012)
[17] (Chiang 2017)
[18] (Grab 2019)
[19] (Uber 2018)

## Bias and Discrimination

Digital platforms collect vast quantities of our regular micro behaviours that can then be analysed to reveal unexpected correlations to meso-level trends, such as susceptibility to illness, risk of traffic accidents[20], or tendency towards criminal activity. Using machine learning for complex big data analyses can be very useful, but it can also put vulnerable segments of society at risk of discriminatory effects.

Researchers have been exploring the use of Facebook user data to predict medical conditions[21]. For example, identifying emotion-associated language markers allow inferences to be made about a person's susceptibility to clinical depression. These language markers are consistent with some characteristics of depression: loss of drive and interest, reduced engagement and interactions, and social withdrawal that eventually leads to social isolation[22]. Health insurers use predictive algorithms like this to identify hidden health issues that require medical services, but these predictive algorithms also lead to higher medical insurance premium rates for customers with supposedly higher health risks even though they may not actually be at risk[23]. Furthermore, the ease with which public social media data can be obtained means that the data are susceptible to misinterpretation or misuse.

Malaysian law enforcement appear to be keen to adopt new digital technologies to improve public safety. Earlier this year, the country's first artificial intelligence-based facial recognition camera system was introduced in Penang to identify criminals on the street[24]. The auxiliary police force is also integrating on-body cameras with high-end facial recognition feature in criminal identification[25]. Artificial intelligence software such as CloudWalk Technology use predictive algorithms that draw from criminals' facial characteristics, body language and gait patterns across photos and videos to identify suspicious behaviour[26]. However, these tools are unreliable because inherent biases in algorithms and the training data they rely on[27] could potentially lead to misidentification, racial profiling and discrimination in criminal sentencing[28]. The city of San Francisco, concerned over potential misuse of this technology, banned the use of facial recognition by all city agencies, including law enforcement[29].

## Societal Effects

Platform monopolies like Facebook and Amazon have so much individual data in their hands that their big data analyses can have far-reaching societal effects, as was the case when Cambridge Analytica got hold of the personal data of millions of Facebook users[30].

[20] (Kita and Kidziński 2019)
[21] (Eichstaedt et al. 2018; Merchant et al. 2019)
[22] (Kupferberg, Bicks, and Hasler 2016)
[23] (Allen 2018)
[24] (The Star 2019a)
[25] (Tao 2018)
[26] (Mozur 2019)
[27] (Buolamwini 2017)
[28] (Angwin et al. 2016)
[29] (Conger, Fausset, and Kovaleski 2019)
[30] (Meredith 2018)

In 2014 a researcher named Aleksandr Kogan developed an app for Facebook. Several hundred thousand Facebook users who took a personality quiz using that app consented to share their personal data with the app. Unbeknown to them, the app also collected the personal data of all their friends, resulting in a pool of data for over eighty-seven million Facebook users. All this data was available to Kogan, who sold it to Cambridge Analytica. Cambridge Analytica then used this data to develop psychographic profiles of voters, which allowed it to micro-target political ads in the 2016 US Presidential election, potentially giving the Trump campaign, with whom it was affiliated, an unfair advantage. Individual privacy rights were violated and this affected the political landscape of an entire nation. When this scandal became public, it rightly raised many questions around data privacy and Facebook's apparent lack of respect for its users' personal data rights. Big data-driven methods of political persuasion like this are increasing globally[31].

Political actors have long used data to target messaging and influence behaviour, but never on this scale and with this degree of micro-targeted advertising. In the 2016 EU referendum, exploitation of Facebook data and misleading targeted ads supporting the Leave campaign led to the unexpected referendum result among British voters[32]. Investigative journalist Carole Cadwalladr explained how residents of the traditionally left-wing town of Ebbw Vale were targeted with fake information about Turkish immigration on Facebook, resulting in a spike of "Leave" voters[33]. Not only are messages customised to the concerns of specific profile groups, but such groups are also often unaware that they have been targeted.

[31] (Tactical Tech's Data and Politics team 2019)

[32] (Cadwalladr 2018)

[33] (Cadwalladr 2019)

## Discussion and Conclusion

We recognise that the collection and use of anonymised personal data can be useful in many fields, including policy research. For example, location data can reveal travel patterns indicating where traffic flow and public transportation should be improved. Accurate and verifiable data, and lots of it, should inform policy decisions, but policymakers must draw the fine line between individual rights and collective benefits that governments and corporations must walk. Traditionally, data privacy concerns have revolved around personal data falling into the wrong hands. Because of the power of networked computing analysis, these security concerns are just one piece of the privacy puzzle. Privacy advocates are becoming increasingly concerned about data privacy as it relates to surveillance. Contemporary privacy concerns include not just the personal data, such as medical records, but also additional information that can be inferred from personal data, such as increased health risks.

Malaysia's Personal Data Protection Act (PDPA) 2010 addresses data privacy vis-à-vis security. It applies to personal data exchanged in commercial transactions, such as income declarations when applying for a bank loan. The PDPA stipulates that personal data collected for commercial purposes cannot be shared with third parties without consent but it is rarely enforced[34] and does not regulate what types of data are collected or how those data are used. The Communications and Multimedia Ministry announced in March that it is in the process of reviewing the PDPA to update it in line with the European Union's General Data Protection Regulation (GDPR)[35]. The GDPR, in addition to data security requirements already in the PDPA, also limits personal data collection to only what is necessary, allows companies to keep data only for current requirements, and prohibits them from saving data for potential future use[36].

Both the PDPA and the GDPR include data security requirements that are staples of information networks but they are not by themselves enough to combat surveillance capitalism and to avoid the consolidation of a wealth of personal data in the hands of a few—at best unwitting and at worst unscrupulous—powerful players. To that end, more research and, ironically, more data, are needed to evaluate how limits can be placed on data collection and distribution while allowing innovation and improving efficiency.

Digital technologies evolve faster than policy and legislation, but that is no reason for policymakers not to seriously consider privacy concerns when developing digital policy. To paraphrase Surina Shukri, CEO of Malaysia Digital Economy Corporation (MDEC), "The data journey…starts with 'what'. Once you have the what, it then becomes the 'so what'. From there, it becomes 'now what'."[37] At the very least, users should be explicitly informed as to what personal data, including behavioural surplus, are being collected by digital platforms and to what ends their data are being

[34] As at the end of 2018, there have been only five reported cases of PDPA breaches. http://www.pdp.gov.my/index.php/en/pusat-media/berita/989-pengguna-data-yang-telah-dikenakan-tindakan-di-bawah-akta-perlindungan-data-peribadi-2010-akta-709

[35] (The Star 2019b)

[36] (European Parliament 2016)

[37] https://twitter.com/mymdec/status/1123802562677891073

used. Wherever possible, users should be able to opt out of being tracked and having their personal data collected, for example by internet cookies, without losing access or functionality.

## References

Allen, Marshall. 2018. "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates." ProPublica. July 17, 2018. https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias." *ProPublica*, May. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Asadullah, Ahmad, Isam Faik, and Atreyi Kankanhalli. 2018. "Digital Platforms: A Review and Future Directions," September.

Buolamwini, Joy Adowaa. 2017. "Gender Shades : Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers." Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/114068.

Buttarelli, Giovanni. 2017. "Privacy Matters: Updating Human Rights for the Digital Society." *Health and Technology* 7 (4): 325–28. https://doi.org/10.1007/s12553-017-0198-y.

Cadwalladr, Carole. 2018. "AggregateIQ: The Obscure Canadian Tech Firm and the Brexit Data Riddle." *The Guardian*, March 31, 2018, sec. UK news: The Cambridge Analytica Files. https://www.theguardian.com/uk-news/2018/mar/31/aggregateiq-canadian-tech-brexit-data-riddle-cambridge-analytica.

———. 2019. *Facebook's Role in Brexit -- and the Threat to Democracy*. TED Talk. https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy.

Chiang, Angel. 2017. "Former Uber Employee Files Suit for Retaliation in Reporting Insecure Data Privacy Practices." *American Bar Association*, September 1, 2017. https://www.americanbar.org/groups/litigation/committees/privacy-data-security/practice/2017/former-uber-employee-files-suit-for-retaliation-insecure-data-privacy-practices/.

Conger, Kate, Richard Fausset, and Serge F. Kovaleski. 2019. "San Francisco Bans Facial Recognition Technology." *The New York Times*, May 14, 2019, sec. U.S. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

Eichstaedt, Johannes C., Robert J. Smith, Raina M. Merchant, Lyle H. Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A. Asch, and H. Andrew Schwartz. 2018. "Facebook Language Predicts Depression in Medical Records." *Proceedings of the National Academy of Sciences* 115 (44): 11203–8. https://doi.org/10.1073/pnas.1802331115.

European Parliament. 2016. *General Data Protection Regulation (GDPR)*. *OJ L*. Vol. 119. http://data.europa.eu/eli/reg/2016/679/oj/eng.

Facebook. 2018. "Facebook Data Policy." Facebook. April 19, 2018. https://en-gb.facebook.com/about/privacy/update.

Grab. 2019. "Grab Privacy Policy." Grab MY. March 14, 2019. https://www.grab.com/my/privacy/.

Greenwald, Glenn. 2016. "New Study Shows Mass Surveillance Breeds Meekness, Fear and Self-Censorship." *The Intercept* (blog). April 28, 2016. https://theintercept.com/2016/04/28/new-study-shows-mass-surveillance-breeds-meekness-fear-and-self-censorship/.

Introna, Lucas D. 1997. "Privacy and the Computer: Why We Need Privacy in the Information Society." *Metaphilosophy* 28 (3): 259–75. https://doi.org/10.1111/1467-9973.00055.

Khan, Lina M. 2017. "Amazon's Antitrust Paradox." *The Yale Law Journal* 126 (3): 564–907.

Kita, Kinga, and Łukasz Kidziński. 2019. "Google Street View Image of a House Predicts Car Accident Risk of Its Resident." *ArXiv:1904.05270 [Stat]*, April. http://arxiv.org/abs/1904.05270.

Kupferberg, Aleksandra, Lucy Bicks, and Gregor Hasler. 2016. "Social Functioning in Major Depressive Disorder." *Neuroscience & Biobehavioral Reviews* 69 (October): 313–32. https://doi.org/10.1016/j.neubiorev.2016.07.002.

Malaysian Communications and Multimedia Commission. 2018. "Internet Users Survey 2018." 1823–2523. Malaysian Communications and Multimedia Commission. https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/Internet-Users-Survey-2018.pdf.

———. 2019. "MCMC: Facts and Figures, 4Q 2018." Malaysian Communications and Multimedia Commission. https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/4Q18.pdf.

Merchant, Raina M., David A. Asch, Patrick Crutchley, Lyle H. Ungar, Sharath C. Guntuku, Johannes C. Eichstaedt, Shawndra Hill, Kevin Padrez, Robert J. Smith, and H. Andrew Schwartz. 2019. "Evaluating the Predictability of Medical Conditions from Social Media Posts." *PLOS ONE* 14 (6): e0215476. https://doi.org/10.1371/journal.pone.0215476.

Meredith, Sam. 2018. "Facebook-Cambridge Analytica: A Timeline of the Data Hijacking Scandal." *CNBC*, April 10, 2018. https://www.cnbc.com/2018/04/10/facebook-cambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html.

Mozur, Paul. 2019. "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority." *The New York Times*, April 14, 2019, sec. Technology. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

Pierson, Jo, and Rob Heyman. 2011. "Social Media and Cookies: Challenges for Online Privacy." *Info* 13 (6): 30–42. https://doi.org/10.1108/14636691111174243.

Rossotto, Carlo Maria, Prasanna Lal Das, Elena Gasol Ramos, Eva Clemente Miranda, Mona Farid Badran, Martha Martinez Licetti, and Graciela Miralles Murciego. 2018. "Digital Platforms: A Literature Review and Policy Implications for Development." *Competition and Regulation in Network Industries* 19 (1–2). https://doi.org/10.1177/1783591718809485.

Tactical Tech's Data and Politics team. 2019. "Personal Data: Political Persuasion. Inside the Influence Industry. How It Works."

Tao, Li. 2018. "Malaysian Police Wear Chinese Start-up's AI Camera to Identify Suspected Criminals." *South China Morning Post*, April 20, 2018. https://www.scmp.com/tech/social-gadgets/article/2142497/malaysian-police-wear-chinese-start-ups-ai-camera-identify.

The Star. 2019a. "Use of Biome-tric Facial Recognition Must Be Regulated." *The Star Online*, January 3, 2019, Nation edition. https://www.thestar.com.my/news/nation/2019/01/03/use-of-biometric-facial-recognition-must-be-regulated/.

———. 2019b. "Gobind: Personal Data Protection Law Being Reviewed." *The Star Online*, March 18, 2019, Nation edition. https://www.thestar.com.my/news/nation/2019/03/18/gobind-personal-data-protection-law-being-reviewed/.

Trigg, Rose. 2017. "Cheating Frenchman Sues Uber for €45m Blaming Glitch in App for His Divorce." *The Local*, February 8, 2017. https://www.thelocal.fr/20170208/frenchman-sues-uber-for-45-million-after-glitch-lets-his-wife-track-him.

Uber. 2012. "Rides of Glory." *Uber's Blog* (blog). March 26, 2012. https://web.archive.org/web/20141118192805/http:/blog.uber.com/ridesofglory.

———. 2018. "Uber Privacy Policy." May 25, 2018. https://privacy.uber.com/policy/.

United Nations. 2015. "Universal Declaration of Human Rights 1948." United Nations. https://www.un.org/en/udhrbook/pdf/udhr_booklet_en_web.pdf.

Warren, Samuel D., and Louis D. Brandeis. 1890. "The Right to Privacy." *Harvard Law Review* 4 (5): 193–220. https://doi.org/10.2307/1321160.

Whitley, Edgar A. 2009. "Informational Privacy, Consent and the 'Control' of Personal Data." *Information Security Technical Report* 14 (3): 154–59. https://doi.org/10.1016/j.istr.2009.10.001.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.