

Big Data is fuelling the revolution in artificial intelligence

But Big Data's complexities prevent society gaining the full benefit from information collected in cyberspace.

Strava is an app that styles itself as "the social network for those that strive".¹ Runners, cyclists and others use the app to track, share and compare their exertions and even determine a 'Suffer Score'.

The San Francisco-based company, for its part, can analyse the data generated by its undisclosed number of users, thought to number in the millions. For no particular reason, Strava recently hosted a 'heat map' of three trillion location points it had collected from one billion activities.² What could go wrong with featuring "the largest, richest and most beautiful dataset of its kind" when revamped algorithms ensured Strava's privacy rules were enforced? Just that it created one of the largest security breaches of recent years.

An Australian student in January used the data trails of people exercising in remote areas to deduce the location of "clearly identifiable and mappable" secret US airstrips, bases and outposts in Afghanistan and Syria.³ Others spotted Turkish and Russian military activity in Syria, French military bases in the Sahel region of Africa and secretive SAS bases in the UK. To widen the security breach, Strava users could see the profiles of soldiers and aid workers serving in these remote areas.

The error of Strava's military users was to fail to 'opt out' of the default setting that shared their data with the company and others. Thus, troops jogging around secret bases or on patrol revealed their locations to the world through their phones or GPS devices.

The Strava saga highlights some of the contentious issues surrounding 'Big Data', a term that describes the volunteered and observed information collected from internet users and connected objects that can be analysed to learn more about those users, things and their surroundings. The issues include privacy, ownership, the privileges and responsibilities of data-gatherers and the ignorance surrounding Big Data's characteristics. From an overarching perspective, the economic traits of Big Data, privacy issues and the lack of clear property rights around this core ingredient of the artificial-intelligence revolution prevent the growth of markets that would steer the data in an efficient and timely way to where it could generate the most benefit for the economy and society.

The challenge for policymakers is to define ownership and encourage data markets while minimising the risks that Big Data can pose,

which extend to security lapses and the concentration of power and rewards among relatively few companies. Policymakers, outside the US at least, are acting to tackle issues surrounding Big Data. Best they are resolved soon. Ever more data is being harvested. Society might as well put it to its best use.

Much data gathered is innocuous and worthless and thus uncontroversial, it must be said. The privately-owned data gatherers are conscious of their responsibilities to protect the data that is only valuable because these businesses found a way to commercialise it. The amount collected is so overwhelming it almost amounts to default privacy protection – a Facebook user, for instance, is one of two billion people so his or her information is unlikely to be identified.

The commercial imperative is such, however, that the more data is gathered the more incentives the accumulators have to profit from that information. While artificial intelligence embeds itself further into everyday living, demand for data will only grow its controversial aspects and market failures will demand resolution. But it won't be easy to make Big Data a 'new asset class' as the World Economic Forum envisaged successful data markets in 2011⁴ or minimise its controversies. Big Data is different and its future uses – and pitfalls – are unknowable.

Faulty comparison

The amount of stored data is exploding as the capture and use of data becomes embedded across everyday life such that some see that Big Data's use in essential services, emergencies, scientific research and boosting knowledge and decision making overall makes it a 'public good'.⁵ The US-based International Data Corp estimates the digital world will grow tenfold from 2016 to 2025.⁶ The value of Big Data is ballooning because, along with massive increases in computing power, it is powering the 'deep learning' models underlying the recent advances in artificial intelligence that are changing daily life for people, business and government.

For people, the Big Data gathered can lead to tailored internet services, being better informed, improved health diagnosis and treatments, even access to finance unavailable to people who fail to surmount traditional borrowing hurdles. Government and institutions use Big Data to improve governance and inform their decisions, or even spy on their citizens, as China's Communist regime does by

¹ Quotes on Strava's homepage on 30 January 2018. strava.com

² Strava release. "The global heatmap, now 6x hotter." 1 November 2017. medium.com/strava-engineering/the-global-heatmap-now-6x-hotter-23fc01d301de

³ Twitter account of Nathan Ruser, who describes himself as 'founding member of @IUCAnalysts. Monitoring a number of conflicts, especially Syria and Iraq'. twitter.com/Nrg8000/status/957318498102865920

⁴ World Economic Forum. 'Personal data: The emergence of a new asset class.' 17 February 2011. weforum.org/reports/personal-data-emergence-new-asset-class

⁵ See Linnet Taylor. 'The ethics of Big Data as a public good: which public? Whose good?' 28 December 2016. Published on the website of the National Center for Biotechnology Information. ncbi.nlm.nih.gov/pmc/articles/PMC5124068/

⁶ International Data Corp. 'Data Age 2025: The evolution of data to life critical.' April 2017. seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf

mounting algorithmic surveillance of people to determine a 'social credit' or 'sincerity' score for each.⁷

In the business world, companies such as Google and even Strava owe their existence to Big Data (and mobile phones). For driverless cars, Big Data is as essential as the wheels on these vehicles. Big Data can boost productivity by automating previously manual tasks, allows farms, mines, plants and factories to run more efficiently, and is the means by which internet companies identify advertising niches among their users. Business demand for datasets will only rise because companies are seeking to compile or access Big Data to innovate and to defend their market shares.

The data-generated wealth of companies such as Alphabet (owner of Google), Amazon and Facebook has prompted many people to describe Big Data as the 'new oil'. That comparison doesn't withstand scrutiny, however. The different traits of Big Data and oil highlight why it won't be easy to create flourishing data markets akin to oil's anytime soon.⁸

Oil, like any commodity, is indistinguishable across producers. Oil's supply is finite and it is single use. Oil's worth comes from the difficulty and expense of finding and extracting the substance from the earth. Ownership is usually clear-cut. These features mean that oil is easy to price and trade.

Data, on the other hand, is not a commodity because each dataset is unique. Its supply is infinite. Data is easy to gather, simple to copy and can be reused and multipurposed. Data is only as good as the algorithms that run over it for insights. The ownership of data is contentious. Privacy concerns restrict data's change of ownership and possible uses, and install security obligations on holders. Buyers are uncertain as to the worth of any dataset because they are yet to know what insights algorithms can glean from it. These features make data difficult to price and trade.

While some data is traded, usually for advertising purposes, much data sits in isolated collections (at Google and Facebook) rather than flows via markets to where it might be more useful for the economy and society. Data-hungry businesses, universities and medical researchers will mostly be forced to gather their own data, an expensive and timely process, if even possible. Thus, Big Data might not be the boost to productivity it could be.

Disputed possession

Capitalism's great success at raising living standards rests on the rigorous enforcement of property rights. A core complexity with Big Data is that its ownership is contested. The clash is politically charged because the question sits as the heart of today's concerns that Big Data concentrating power in the hands of a few giant global companies.

The prevailing world of Big Data is one where the companies that gather the data control the information by default. Concerns are mounting, however, that Big Tech is generating massive gains from Big Data and not providing enough in return to users or society. German Chancellor Angela Merkel told the World Economic Forum in January that: "The answer to the question of who owns this data will ultimately decide whether democracy, participation, sovereignty in the digital age and economic success can go together".⁹

Europe's response is to make laws that keep the ownership and control of Big Data with the subjects. Under the General Data Protection Regulation¹⁰ that is effective from May, businesses must release people's data on request from the data subject. Under the regulation's 'Article 20 – Right to data portability',¹¹ people can essentially sell their data to a rival business (possibly prompted by incentives). Under 'Article 17 – Right to erasure ('right to be forgotten')',¹² people can ask businesses to erase data about them, similar to their right to ask Google to remove links to articles about them from search results.

Australia announced in November an intention to pass a similar law to Europe's, to enable customers to switch between companies to gain better deals. The upcoming 'Consumer Data Right' bill will first apply to customers of banks, utilities and telecoms before extending to other industries.¹³

Consumer-slanted data-ownership laws could help people place a value on their data while easing concerns about privacy violations. Such laws, however, usually prove problematic. The laws only cover a portion of the data collected. Compliance can be costly. Companies often hinder the spirit of the law. Enforcement powers are feeble and the punishments weak. People are often too complacent to reclaim their data ownership. And even with portable data, entrants to markets will find it hard to dislodge incumbent companies cemented in stranglehold positions by network effects.

To overcome this complacency and help build data markets, five US economists (including one from Microsoft) have suggested that people think of their time on Facebook and other platforms not as leisure but as work – after all, people are turning their demographic profiles, social class, interests, political leanings and other personal information into advertising revenue for these companies. The 'Data as Labour' model suggested by the economists would see people paid for their data. To bargain against the companies that are thriving under the "Data as Capital" model of doing business, the economists suggest internet users could organise themselves into a "data labour union" that "could credibly call a powerful strike". Absent such militancy (and user unions and strikes against Facebook sound far-fetched), they say that governments will need to intervene to address the perceived imbalance in rewards between the users

⁷ See The Atlantic. 'China's surveillance state should scare everyone.' 2 February 2018. theatlantic.com/international/archive/2018/02/china-surveillance/552203/. A translation of the State Council's "Planning outline for the construction of a social credit system (2014-2020)" can be found at: chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/

⁸ See World Economic Forum paper by Adam Schlosser, project leader, digital trade and data flows, World Economic Forum. 'You may have heard data is the new oil. It's not.' 10 January 2018. weforum.org/agenda/2018/01/data-is-not-the-new-oil/

⁹ German Chancellor Angela Merkel. 'Speech by Federal Chancellor Angela Merkel at the World Economic Forum annual meeting in Davos on 24 January 2018.' 24 January 2018.

bundesregierung.de/Content/EN/Reden/2018/2018-01-24-bk-merkel-davos_en.html

¹⁰ General Data Protection Regulation. gdpr-info.eu/

¹¹ Article 20 GDPR. 'Right to data portability.' gdpr-info.eu/art-20-gdpr/

¹² Article 17, GDPR. 'Right to erasure ('right to be forgotten')'. gdpr-info.eu/art-17-gdpr/

¹³ Australian government Department of Prime Minister and Cabinet. "Australians to own their own banking, energy, phone and internet data." Media release. 26 November 2017. ministers.pmc.gov.au/taylor/2017/australians-own-their-own-banking-energy-phone-and-internet-data

and tech platforms.¹⁴ Others have suggested laws that force data gatherers to share their data, but this raises privacy issues.

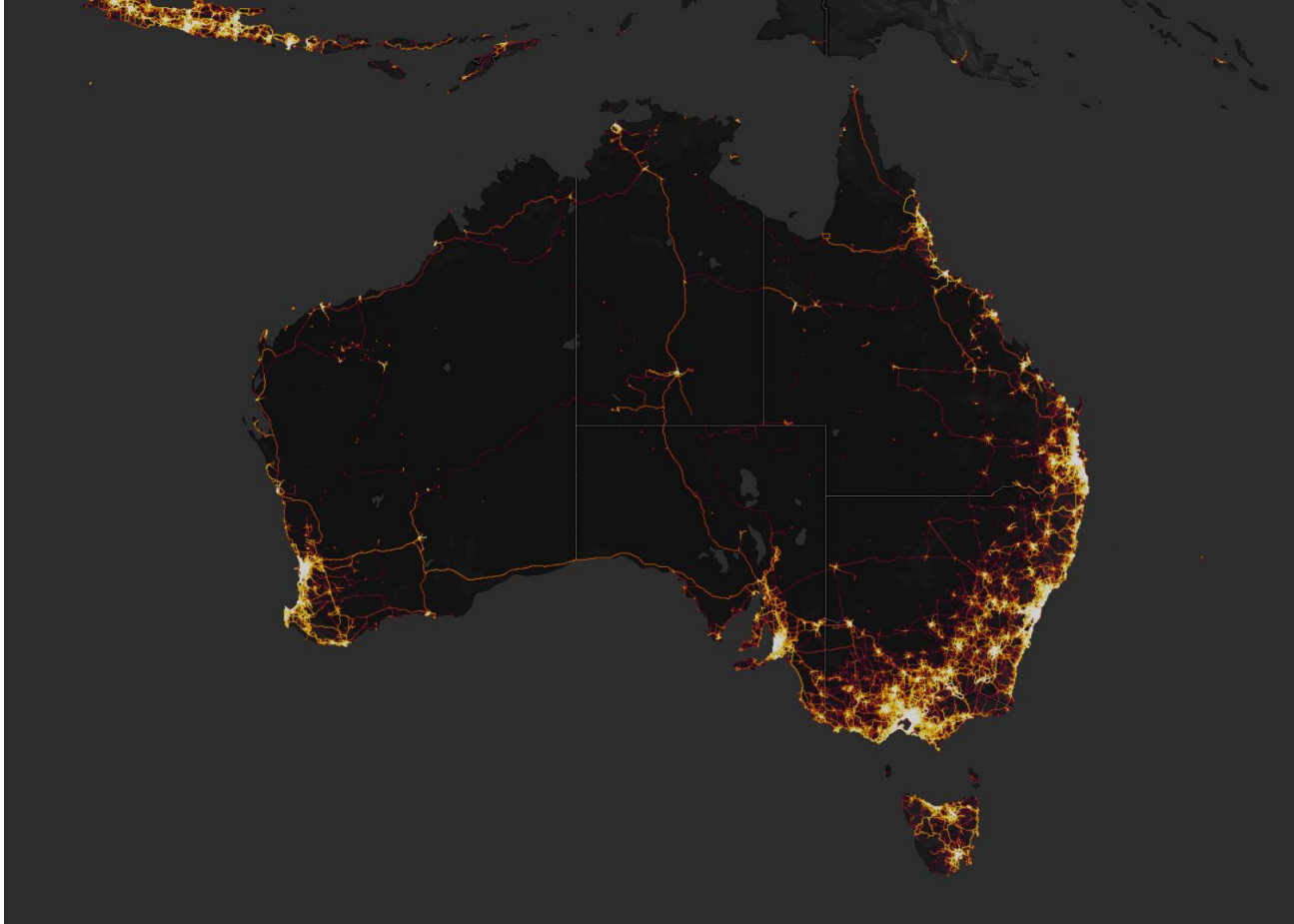
As the nature of Big Data embeds it with complexities surrounding privacy and ownership among other issues, society will take a while to work out how to gain the most out of Big Data and minimise Strava-style hiccups.

By Michael Collins, Investment Specialist

Strava heat map of Australia

Strava website. 13 March 2018.

[strava.com/heatmap/#4.57/136.01525/-27.81602/hot/all](https://www.strava.com/heatmap/#4.57/136.01525/-27.81602/hot/all)



Important Information:

This material has been prepared for general information purposes and must not be construed as investment advice. This material does not constitute an offer or inducement to engage in an investment activity nor does it form part of any offer or invitation to purchase, sell or subscribe for in interests in any type of investment product or service. This material does not take into account your investment objectives, financial situation or particular needs. You should read and consider any relevant offer documentation applicable to any investment product or service and consider obtaining professional investment advice tailored to your specific circumstances before making any investment decision. This material and the information contained within it may not be reproduced or disclosed, in whole or in part, without the prior written consent of Magellan Asset Management Limited. Any trademarks, logos, and service marks contained herein may be the registered and unregistered trademarks of their respective owners. Nothing contained herein should be construed as granting by implication, or otherwise, any licence or right to use any trademark displayed without the written permission of the owner. No part of this material may be reproduced or disclosed, in whole or in part, without the prior written consent of Magellan Asset Management Limited.

¹⁴ Imanol Arrieta Ibarra, Leonard Goff, Diego Jiménez Hernández, Jaron Lanier and E. Glen Weyl. 'Should be treat data as labor? Moving beyond "free"'. (Lanier is from Microsoft while Weyl is connected with Microsoft and Yale. Ibarra and Hernández are from Stanford. Goff is from Columbia.)